

Volume 67 • Number 2 • April 2019

Acta Geophysica

PAN

POLISH ACADEMY OF SCIENCES



Institute of Geophysics
Polish Academy of Sciences



Springer



Estimated casualties in possible future earthquakes south and west of the M7.8 Gorkha earthquake of 2015

Max Wyss¹ · Deepak Chamlagain²

Received: 30 November 2018 / Accepted: 5 February 2019 / Published online: 11 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019, corrected publication 2019

Abstract

The 2015 M7.8 Gorkha earthquake has moved the upper, unbroken, part of the Main Himalayan Thrust (MHT) and the neighboring sections of this fault closer to failure. Using the program and data set of QLARM, which has been correct in fatality estimates of past Himalayan earthquakes, we estimate quantitatively the numbers of fatalities, injured and strongly affected people when assumed ruptures along these two sections will happen. In the Kathmandu up-dip scenario with M8.1, we estimate that more than 100,000 people may perish, about half a million may be injured, and 19 million are likely to be affected strongly, if we assume the high virtual attenuation observed for the 2015 Gorkha earthquake exists here also. Likewise, if the 100 km underthrusting segment west of Gorkha ruptures, we quantitatively estimate that 12,000–62,000 people may perish and 4 million to 8 million will be strongly affected, in a down-dip (lower half of the thrust plane) and an up-dip rupture (upper half) scenario, respectively. If the up-dip part of the MHT cannot rupture by itself, and greater earthquakes are required to generate the several meters of displacement observed in trenches across the MHT, then our estimates are minima.

Keywords Himalayan earthquakes · Earthquake risk · Earthquake losses in Nepal

Introduction

The 2015 Gorkha M7.8 earthquake and its aftershock of M7.3 devastated large parts of Nepal and killed 9870 people (Ministry of Home Affairs, Nepal 2016). The main shock occurred on a Saturday, when children were not in school. Had they been in the collapsing schools, the death toll could have been double. Here, we estimate the numbers of casualties (fatalities and injured), when parts of the Main Himalayan Thrust (MHT) neighboring the 2015 ruptures will also rupture.

Undoubtedly, all parts of the MHT will generate earthquakes when the stress accumulating due to the Indian plate plowing into Asia is sufficient for failure (e.g., Bilham et al. 2001), but the time when this will happen for the various segments is not known. Small and medium size earthquakes occur frequently, sometimes generating local disasters. Here,

we estimate losses in great earthquakes of $M > 7.5$ only, so that government and first responders have a quantitative assessment of the earthquake disasters looming in Nepal and neighboring parts of India.

The MHT often ruptures in a down-dip and a separate up-dip earthquake, each covering about half of the thrust plane of the same segment (e.g., Avouac et al. 2015; Wesnousky et al. 2017). When the down-dip earthquake happens, as in the case of the M7.8 Gorkha earthquake of 2015, the stress in the up-dip part, as well as in the adjoining western and eastern segments, is increased, bringing these neighboring parts of the MHT closer to failure (e.g., Bilham et al. 2017).

The questions we are asking here are: How many fatalities and injured are to be expected if (1) the up-dip earthquake complementary to the 2015 Gorkha earthquake occurs, and (2) if the western adjoining 100 km segment of the MHT ruptures in its down-dip and up-dip parts separately, or in its entire width of 90 km, at once.

A pairing of earthquakes, as we propose here for our scenarios, has not been observed historically, yet several meters of displacement have been documented along many parts of the MHT (e.g., Bilham et al. 2001; Kumar et al. 2006; Lave et al. 2005; Pathier et al. 2006; Kaneda et al. 2008; Sapkota et al. 2012; Rajendran et al. 2015; Bungum et al. 2017;

✉ Max Wyss
max@maxwyss.ch

¹ International Centre for Earth Simulation Foundation, Geneva, Switzerland

² Department of Geology, Trichandra M. Campus, Tribhuvan University, Kathmandu, Nepal

Wesnousky et al. 2017). It follows that large to great earthquakes must be expected in the segments of the MHT we address here. Some interpretations of Himalayan tectonics (e.g., Bilham et al. 2017) envision much larger earthquakes than those proposed here to cause the displacements along the outcrops of the MHT, in which case our loss estimates are minima.

Source models for hypothetical future earthquakes

The up-dip Kathmandu scenario source parameters

We assume that a rupture complementary to and up-dip from the 2015 Gorkha earthquake has a length, $L(\text{up-dip}) = 182$ km, equal to the maximum length of the aftershock zone of 2015 (e.g., Hayes et al. 2015). The width we take as the distance between the southern end of the 2015 aftershocks and the outcropping fault: $W(\text{up-dip}) = 63$ km. From the resulting rupture area $A = L * W$, we estimate the magnitude of the complementary Kathmandu up-dip earthquake as M8.1 (Wyss 1979; Wells and Coppersmith 1994) (Table 1). This magnitude is somewhat larger than the one in 2015 because the remaining width of the failure plane is larger than the one that broke in 2015.

Earthquake sources are modeled in the program QLARM (Trendafiloski et al. 2011) simplified as lines drawn along the part of the rupture plane, which radiates energy most strongly. The position of this line within the assumed rupture plane is important because it controls the proximity to areas of dense population.

Validation tests for QLARM results have been carried out for Mexico (Wyss and Zuniga 2016), India and Pakistan (Wyss et al. 2018), and China (Li et al. 2018). In general, these validations proceeded as follows. The earthquake source parameters were given by publications including magnitude, hypocenter, and isoseismal maps. When available, inversion models for the rupture plane were also considered. In addition, the numbers of casualties were known from official counts. QLARM estimates of fatalities were

validated, when the observed numbers were within the calculated range for events with many casualties, and within factors of three for small events, where the sum of fatalities was approximately equal to those caused by the collapse of a single large building.

Verifying QLARM for loss calculations in the Himalaya, it was found that this line should not be placed along the surface break, but parallel to it and within about the center of the rupture area (Wyss et al. 2018). This is especially true for the M7.6 Kashmir earthquake of 2005 (e.g., Pathier et al. 2006), where observed intensities (Zare et al. 2009) can only be matched by calculated ones with the line source placed at the center of the known ruptured area, not along the surface rupture. Applying this procedure to the up-dip Kathmandu scenario, we propose to use the line source for which the end points are given in Table 1.

The west-of-Gorkha scenarios source parameters

West of the Gorkha 2015 rupture, there is a section of the MHT with low seismicity during the instrumental observation period. We assume that this part may rupture in an earthquake of the same M7.8, as the 2015 Gorkha event. Its length is assumed to be L (west of Gorkha) = 100 km. The width of the seismically active underthrusting plane is about 90 km.

We further assume in two scenarios that this section may fail in separate complimentary down-dip and up-dip earthquakes, like other segments of the MHT. Each of these two ruptures are assigned half the entire width, that is, 45 km (Table 1). The magnitudes corresponding to each of these areas are calculated to be 7.8. Alternatively, the entire width of the thrust plane may rupture in a single earthquake from the deepest parts to the fault outcrop. In that case, the magnitude calculates as M8.1 (Table 1).

The down-dip epicenter and rupture line for the west-of-Gorkha hypothetical earthquake are at the same distance from the surface trace of the MHT as the Gorkha epicenter. The up-dip rupture line is assumed to be at the center of the upper half of the thrust plane. The all-dip rupture line is placed at the middle of the other two hypothetical lines.

Table 1 Source parameters of the hypothetical earthquakes that may follow the M7.8 Gorkha earthquake of 2015

Name	Epi Lat (deg)	Epi Lon (deg)	Depth (km)	Length (km)	Width (km)	M	Y_1 (deg)	X_1 (deg)	Y_2 (deg)	X_2 (deg)
Kathmandu up-dip	27.44	85.19	15	182	63	8.1	27.75	84.30	27.13	86.07
West of Gorkha down-dip	28.51	84.03	20	100	45	7.8	28.71	83.58	28.31	84.48
West of Gorkha up-dip	28.01	83.78	15	100	45	7.8	28.13	83.27	27.88	84.29
West of Gorkha all-dip	28.28	83.88	15	100	90	8.1	28.41	83.40	28.15	84.35

X_i and Y_i are the assumed end points of the respective line rupture. M is calculated from the rupture area. The epicenter is placed at the center of the line rupture

All lines are parallel to the seismicity trend and the surface rupture of the MHT. The endpoints of these hypothetical rupture lines are given in Table 1.

Additional parameters and properties of the scenarios

The assumed attenuation of seismic waves influences the estimated strong ground motions and therefore the losses. In the case of the 2015 Gorkha earthquake, the intensities observed were surprisingly low (e.g., Martin et al. 2015). This can be explained by strong attenuation, or by weak high-frequency radiation from the source. In QLARM, we cannot model the frequency content of the source; thus, we fold the two effects into a single parameter, the “virtual attenuation.”

Matching the Intensities observed in 2015, it was found (Wyss 2017) that the parameter $C3 = 2.3$ in the attenuation equation of Shebalin (1985) must be applied. We use Shebalin’s equation because its parameters easily move the attenuation curve as a function of distance from the source into the correct position (Wyss 2017). Based on this observation, we propose $C3 = 2.3$, a strong virtual attenuation, for the most likely models of earthquakes in the Kathmandu area. This leads to a minimum estimate of all the estimated losses.

The hypothetical depth of the source also strongly influences the loss estimates. With shallow earthquakes, the uncertainties of their depths tend to be equal in size to the depth itself. We assume 15 km for the up-dip Kathmandu scenario in comparison with other depth estimates in the Himalaya. The bulk of the energy radiated generally does not come from the shallowest part, but from deeper within the crust, as seen in countless cross sections of earthquake source inversions. Therefore, a hypothetical hypocentral depth of 15 km is appropriate. For the northernmost rupture model, a depth of 20 km is selected (Table 1).

Soil conditions affect the intensities locally. Because these are not known for the couple of thousand settlements affected by intensities V+ in envisioned future earthquakes, we trust only in the estimate of the sum of the casualties, not in the values returned by the computer for individual settlements.

The nature of the building stock is modeled in a very simple way in QLARM because detailed information is not available for the small settlements in the study area. Nevertheless, the fatalities calculated for the 2015 Gorkha earthquake, using the known source properties, agree with the observed ones (Wyss 1979). Therefore, QLARM with its data set for population and building stock is verified for Nepal and we use it.

The time of day when an earthquake strikes affects the losses. At 3AM the occupancy rate of building is the highest,

thus casualties are most numerous. We selected 6AM as the time for scenario earthquakes because this means we assume average occupancy.

Loss estimate for the Kathmandu up-dip scenario

The expected intensities for the Kathmandu up-dip scenario at the settlements in the data base of QLARM are shown in Fig. 1 (left), using the source parameters given in Table 1. The resulting mean damage in each settlement is shown on the right in Fig. 1. Using the influence of damaged and collapsing buildings on occupants and the ratio of the distribution into fatalities and injured (Wyss and Trendafiloski 2011), the casualty numbers are calculated (Trendafiloski et al. 2011).

The ranges of fatalities and injured for the Kathmandu up-dip earthquake scenario using the unusually strong attenuation ($C3 = 2.3$) are listed in Table 2 using the formal probability that 98% of possible results are captured, given the known uncertainties, but not those of the source and transmission properties. Alternatively, if we assume a virtual attenuation close to the world average, the estimated losses would be about three times larger than those given in Table 2.

The number of “strongly affected” people is defined here as those within the area of intensity VI+ (last column Table 2). Injuries and even fatalities are possible within the minimum intensity used for this estimate. People in areas with shaking less than Intensity VI are also “affected,” but less severely. For the purpose of estimating the extent of the disaster, our quantitative definition of “strongly affected” is proposed as a useful parameter.

The numbers of people in the areas shaken by intensities (VI plus VII) and those in areas with VII+ are further separated (Table 2) because there is a significant difference in the devastation in these two areas.

Loss estimate for the west-of-Gorkha scenarios

For the breaking of the west-of-Gorkha underthrusting segment, we model two separate earthquakes, a down-dip half and an up-dip half, as well as a single all-dip rupture of the entire width. We do this because this separation exists in historic earthquakes, including the Kathmandu segment as demonstrated by the earthquake in 2015. This bimodal rupturing of the Himalayan megathrust has been discussed recently by Dal Zilio et al. (2019).

For the west-of-Gorkha segment, we also calculate casualties using the low virtual attenuation ($C3 = 2.3$) needed for

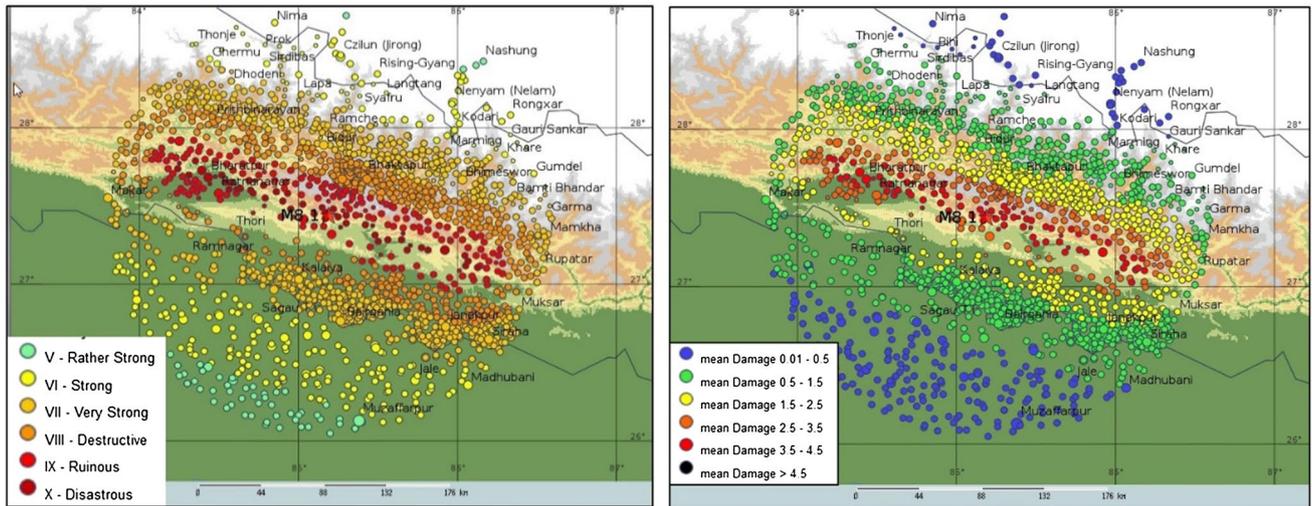


Fig. 1 Intensities (left) and mean damage (right) in case of an up-dip earthquake complimentary to the Gorkha earthquake of 2015, but with M8.1. Each dot represents a settlement in the QLARM data base. The assumed source parameters for the scenario rupture

are given in Table 1, and the estimated results are given in Table 2. The calculation is limited to $R=200$ km. The virtual attenuation of $C3=2.3$ is used. Epicenter: 28.23°N/84.73°E

Table 2 Ranges of fatalities and injured (formal minima and maxima) for the hypothetical scenario for the Kathmandu up-dip and the three west-of-Gorkha earthquake scenarios with source parameters in Table 1

Name	M	Imax	Fatalities (calc)		Injured (calc)		Pop(VI+ VII)	Pop(VIII+)	Pop(VI+)
			Min	Max	Min	Max			
Kathmandu up-dip	8.1	9.0	47,000	269,000	177,000	911,000	14,023,000	5,091,000	19,114,000
West of Gorkha down-dip	7.8	8.0	2,000	21,000	8,000	80,000	3,711,000	248,000	3,958,000
West of Gorkha up-dip	7.8	9.0	17,000	106,000	66,000	357,000	5,898,000	2,056,000	7,954,000
West of Gorkha all-dip	8.1	9.0	27,000	116,000	96,000	374,000	5,754,000	2,144,000	7,899,000

The estimated population located in regions with three intensity ranges given by Roman numerals are listed in the last three columns. The models are calculated for an unusually strong attenuation $C3=2.3$. The population assumed present is extrapolated from 2013 (QLARM data) to 2017, assuming a growth of 5% for the 4 year period (http://cbs.gov.np/sectoral_statistics/population). All results are rounded to the nearest thousand

matching the 2015 intensities generated by the Gorkha M7.8 earthquake. By “virtual attenuation,” we mean the combined effect of the seismic spectrum radiated by the source and the absorption along the path, factors that cannot be separated, unless PGA values are recorded on hard rock at more than one station along the wave paths. The casualties and the numbers of strongly affected people for the three models for the rupture of the west-of-Gorkha segment of the plate boundary are listed in Table 2. Alternatively, if we assumed a virtual attenuation close to world average ($C3=3.5$), the estimated losses would be about three times larger than those given in Table 2.

The maps of intensities calculated and the theoretically derived mean damage for the down-dip rupture of the west Gorkha seismic gap in an earthquake of M7.8, using $C3=2.3$, are shown in Fig. 2. The corresponding maps for a

complementary earthquake rupturing the up-dip half of the west-of-Gorkha seismic gap are shown in Fig. 3.

The scenario for a possible all-dip rupture of the 100 km segment west of Gorka yields similar results with moderately higher casualties as the up-dip scenario (Table 2). Separate maps are not shown for the all-dip scenario because they are similar to those shown in Fig. 3 for the up-dip scenario.

The flip-effect along the MHT fault trace

An augmentation of the surface displacements along outcropping thrust earthquakes has been proposed by Brune (1996) and confirmed by Gabuchian et al. (2017). The excess amplitudes are estimated to amount to 25–50%.

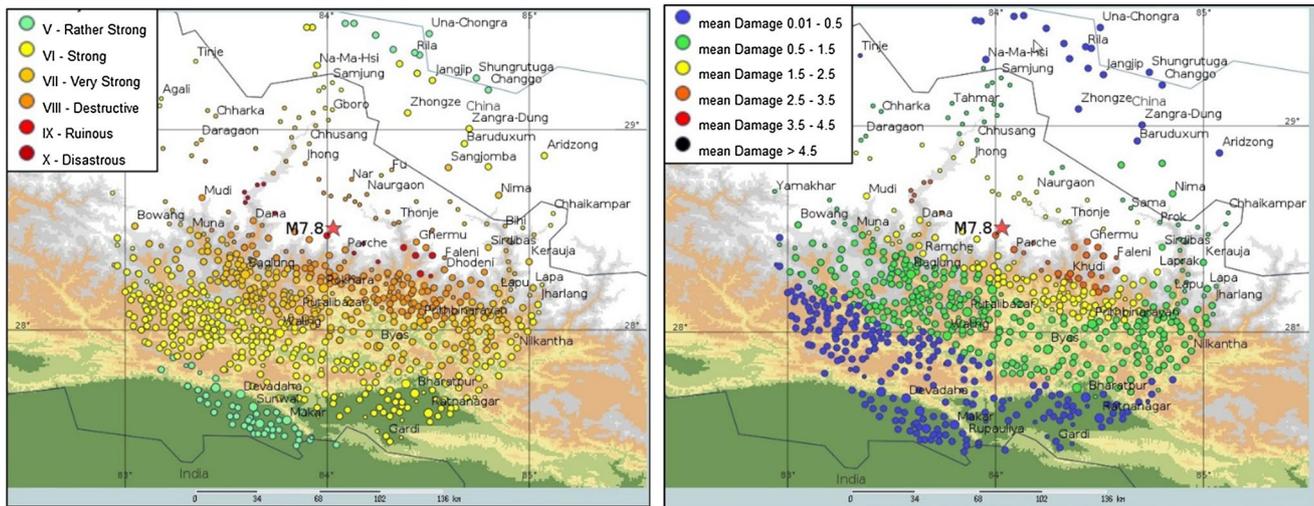


Fig. 2 Intensities (left) and mean damage (right) in case of a down-dip earthquake of M7.8 west of and adjoining Gorkha. Each dot represents a settlement in the QLARM data base. The assumed source

parameters for the scenario rupture are given in Table 1, and the estimated results are given in Table 2. The calculation is limited to $R=200$ km. The virtual attenuation resulting from $C3=2.3$ is used

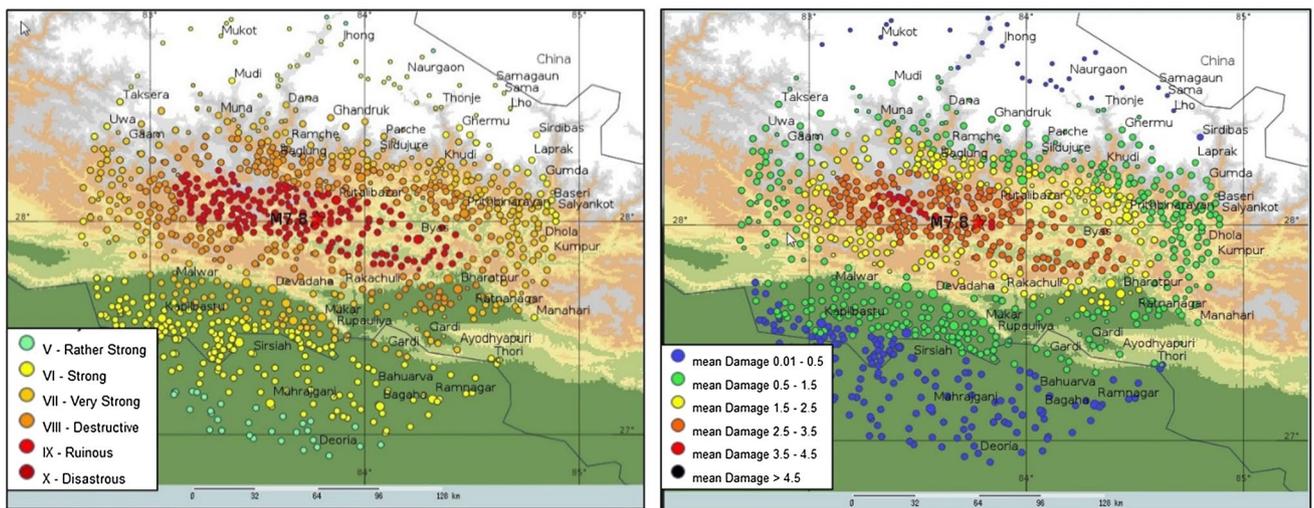


Fig. 3 Intensities (left) and mean damage (right) in case of an up-dip earthquake of M7.8 west of and adjoining Gorkha. Each dot represents a settlement in the QLARM data base. The assumed source

parameters for the scenario rupture are given in Table 1 and the estimated results are given in Table 2. The calculation is limited to $R=200$ km. A virtual attenuation resulting from $C3=2.3$ is used

On the basis of locations of shattered rocks, Brune (2001) estimated that this effect occurs within about 7 km of the surface trace.

The additional casualties due to the flip-effect have not been added because they are much smaller than the uncertainties in our loss estimates. Nevertheless, the several thousand additional casualties close to and on the up-thrust side of the surface trace will be major tragedies for the affected population. Therefore, mitigating measures should concentrate on settlements located in the 7 km wide zone north of the trace of the MHT.

Discussion and conclusions

We emphasize that the loss estimates in the scenarios presented here contain important uncertainties because they are affected by many parameters not precisely known. They are order of magnitude estimates. Because of these uncertainties the numbers of casualties and affected people in Table 2 are rounded.

The influence of the virtual attenuation is an example of how a single parameter can influence the loss estimates.

We present all four scenarios with a strong virtual attenuation (Table 2) because this attenuation factor was required to match the observed intensities due to the 2015 Gorkha earthquake (Wyss 2017). However, the losses would be three times larger if a virtual attenuation closer to the world average were to be used for these estimates.

The losses expected for the up-dip Kathmandu scenario (Table 2) are significantly larger than those recorded after the Gorkha 2015 earthquake. The epicenters of the M7.8 and M7.3 earthquakes in 2015 were in mountainous, thinly populated areas. The expected up-dip Kathmandu earthquake will rupture in a much more populated area, affecting parts of India as well as Nepal; therefore, higher casualties will result.

The calculated mean fatalities in the Kathmandu up-dip earthquake are 158,000, the mean number of injured is about half a million. With an additional 19 million people strongly affected, this will be a truly great disaster, according to our estimate.

The losses expected for the up-dip west-of-Gorkha scenario are significantly larger than those expected for the down-dip break of the west-of-Gorkha seismic gap (Table 2). The difference between the two earthquakes with the same assumed magnitude is that they rupture in areas with different population densities (Table 2, Figs. 2 and 3).

The calculated mean fatalities in the down-dip west-of-Gorkha earthquake are 12,000, the mean number of injured are 44,000, with an additional 4 million people strongly affected. These results assume very strong virtual attenuation and M7.8.

The calculated mean fatalities in the up-dip west Gorkha gap earthquake are 62,000, the mean number of injured are 212,000, with an additional 8 million people strongly affected. This is again a minimum estimate of losses due to a great earthquake west of Gorkha because the attenuation could be less and the rupture length could be larger than assumed.

If the rupture should include the entire underthrusting width of this segment, the all-dip scenario, the numbers of mean fatalities and injured are estimated as 62,000 and 209,000. The population strongly affected would exceed 8 million.

Crude estimates of casualties in eight Himalayan scenarios were published in March 2005 (Wyss 2005). The then named scenarios Kashmir and Kathmandu, both with assumed M8.1, have actually happened in earthquakes killing numbers of people within a factor of 2.5 of the published estimates (Wyss 2006, 2017).

In addition, loss estimates using QLARM within 30 min of large Asian earthquakes are mostly correct within factors of about three for large events and a few hundred fatalities for medium sized earthquakes (Wyss 2014). These real-time alerts are calculated based only on rapidly calculated

epicenters, magnitudes and assumed depths. Given the good success rate with few errors, QLARM estimates can be trusted, understanding that significant uncertainties exist.

Therefore, the current estimates, performed with an improved program and dataset, should be taken seriously. Preparations for the predicted extent of the disasters should be made.

References

- Avouac J-P, Meng L, Wei S, Wang T, Ampuero J-P (2015) Lower edge of locked Main Himalayan Thrust unzipped by the 2015 Gorkha earthquake. *Nat Geosci*. <https://doi.org/10.1038/ngeo2518>
- Bilham R, Gaur VK, Molnar P (2001) Himalayan seismic hazard. *Science* 293:1442–1444
- Bilham R, Mencin D, Bendick R, Bürgmann R (2017) Implications for elastic energy storage in the Himalaya from the Gorkha 2015 earthquake and other incomplete ruptures of the Main Himalayan Thrust. *Q Int XXX*:1–19
- Brune JN (1996) Particle motions in a physical model of shallow angle thrust faulting. *Proc Indian Acad Sci* 105:L197–L206
- Brune JN (2001) Shattered rock and precarious rock evidence for strong asymmetry in ground motions during thrust faulting. *Bull Seismol Soc Am* 91:441–447. <https://doi.org/10.1785/0120000118>
- Bungum H, Lindholm CD, Mahajan AK (2017) Earthquake recurrence in NW and central Himalaya. *J Asian Earth Sci* 138:25–37
- Dal Zilio L, van Dinther Y, Gerya T, Avouac J-P (2019) Bimodal seismicity in the Himalaya controlled by fault friction and geometry. *Nat Commun* 10(48):1–11. <https://doi.org/10.1038/s41467-018-07874-8>
- Gabuchian V, Rosakis AJ, Bhat HS, Madariaga R, Kanamori H (2017) Experimental evidence that thrust earthquake ruptures might open faults. *Nature* 545(7654):336–339
- Hayes GP et al (2015) Rapid characterization of the 2015 Mw 7.8 Gorkha, Nepal, earthquake sequence and its seismotectonic context. *Seismol Res Lett* 86(6):1557–1567
- Kaneda H et al (2008) Surface rupture of the 2005 Kashmir, Pakistan, earthquake and its active tectonic implications. *Bull Seismol Soc Am* 98(2):521–557
- Kumar S, Wesnousky SG, Rockwell TK, Briggs RW, Thakur VC (2006) Paleoseismic evidence of great surface rupture earthquakes along the Indian Himalaya. *J Geophys Res* 111(B3):B03304. <https://doi.org/10.1029/2004jb003309>
- Lave J, Yule D, Sapkota S, Basant K, Madden C, Attal M, Pandey R (2005) Evidence for a great medieval earthquake (~1100 A.D.) in the central Himalayas, Nepal. *Science* 307:1302–1305
- Li JW, Wyss M, Wu ZL, Zhou SY (2018) Estimated casualty risk for six scenario great earthquakes in Sichuan-Yunnan region, China, for disaster preparation. In: International conference on the decade memory of the Wenchuan earthquake, edited, Ministry of Science and Technology of the People's Republic of China, Chengdu, China, p 109
- Martin SS, Hough SE, Hung C (2015) Ground motions from the 2015 M7.8 Gorkha, Nepal, earthquake constrained by a detailed assessment of macroseismic data. *Seism Res Lett* 86(6):1524–1532
- Ministry of Home Affairs (2016) The 2015 Gorkha earthquake: experience and learning. Government of Nepal, p 268
- Pathier E, Fielding EJ, Wright TJ, Walker R, Parsons BE, Hensley S (2006) Displacement field and slip distribution of the 2005 Kashmir earthquake from SAR imagery. *Geophys Res Lett* 33(20):L20310. <https://doi.org/10.1029/2006gl027193>

- Rajendran CP, John B, Rajendran K (2015) Medieval pulse of great earthquakes in the central Himalaya: viewing past activities on the frontal thrust. *J Geophys Res* 120:1623–1641. <https://doi.org/10.1002/2014jb011015>
- Sapkota SN, Bollinger L, Klinger Y, Tapponnier P, Gaudemer Y, Tiwari D (2012) Primary surface ruptures of the great Himalayan earthquakes in 1934 and 1255. *Nat Geosci*. <https://doi.org/10.1038/NGEO1669>
- Shebalin NV (1985) Regularities of the natural disasters. *Nauki o zemle, Znanie* 11:48 (in Russian)
- Trendafiloski G, Wyss M, Rosset P (2011) Loss estimation module in the second generation software QLARM. In: Spence R, So E, Scawthorn C (eds) *Human casualties in earthquakes: progress in modeling and mitigation*. Springer, Berlin, pp 381–391
- Wells DL, Coppersmith KJ (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area and surface displacement. *Bull Seismol Soc Am* 84(4):974–1002
- Wesnousky SG, Kumahara Y, Chamlagain D, Pierce IK, Karki A, Gautam D (2017) Geological observations on large earthquakes along the Himalayan frontal fault near Kathmandu, Nepal. *Earth Planet Sci Lett* 457:366–375
- Wyss M (1979) Estimating expectable maximum magnitude of earthquakes from fault dimensions. *Geology* 7:336–340
- Wyss M (2005) Human losses expected in Himalayan earthquakes. *Nat Hazards* 34:305–314
- Wyss M (2006) The Kashmir M7.6 shock of 8 October 2005 calibrates estimates of losses in future Himalayan earthquakes. Paper presented at proceedings of the conference of the international community on information systems for crisis response and management, Newark
- Wyss M (2014) Ten years of real-time earthquake loss alerts. In: Wyss M (ed) *Earthquake hazard, risk, and disasters*. Elsevier, Waltham, pp 143–165
- Wyss M (2017) Four loss estimates for the Gorkha M7.8 earthquake, 25 April 2015, before and after it occurred. *Nat Hazards* 86:141–150. <https://doi.org/10.1007/s11069-016-2648-7>
- Wyss M, Trendafiloski G (2011) Trends in the casualty ratio of injured to fatalities in earthquakes. In: Spence R, So E, Scawthorn C (eds) *Human casualties in natural disasters: progress in modeling and mitigation*. Springer, London, pp 267–274
- Wyss M, Zuniga R (2016) Estimated casualties in a possible great earthquake along the Pacific coast of Mexico. *Bull Seismol Soc Am* 106(4):1867–1874. <https://doi.org/10.1785/0120160013>
- Wyss M, Gupta S, Rosset P (2018) Casualty estimates in repeat Himalayan earthquakes in India. *Bull Seismol Soc Am* 108(5A):2877–2893. <https://doi.org/10.1785/0120170323>
- Zare M, Karimi-Paridari S, MonaLisa (2009) An investigation on Balakot, Muzaffarabad (Pakistan) earthquake, 8 Oct. 2005, Mw 7.6; geological aspects and intensity distribution. *J Seismol* 13:327–337



The use of earthquake closets in developing countries when large earthquakes strike

Max Wyss¹

Received: 3 December 2018 / Accepted: 19 February 2019 / Published online: 4 March 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

Given that more than 90% of earthquake fatalities occur in rural environments, one has to rethink how to protect the population. With most rural buildings being one- and two-story dwellings, the earthquake closet (EC) offers an affordable solution as a protection unit. Two recent earthquakes with nearly 100,000 fatalities each are used for estimating the lives that could be saved and how much this would cost. The cost of constructing an EC in a single-family home is taken to be \$500–\$600 in developing countries. For the “statistical value of life” \$1 million is used, a minimum amount given in the literature. The cost of hospitalization is derived from reports after an earthquake. The number of fatalities and injured avoided is estimated for an example earthquake each in China and Pakistan. The estimated dollar savings resulting in large earthquakes reach \$18.3 billion at a cost of \$1.3 billion, and \$10 billion at a cost of \$0.5 billion, respectively, in the two examples.

Keywords Earthquake losses · Earthquake protection · Rural population

Properties, accessibility, and usefulness of an earthquake closet

An EC is assumed to be an earthquake protection unit installed in a single-family home as described by Wyss (2012). It is envisioned as a strong structure, with dimensions of about 2 m by 2 m, in the structurally strongest corner of a home. This EC is similar to a tornado shelter: a small cubical in which a family could fit and be protected from heavy falling objects, such as a ceiling. Engineering advice would be desirable for building an EC, but a head of family could construct it from heavy beams and thus increase the chances of survival of a family by a factor of at least 1000 (Wyss 2012). The material used and the details of construction may vary, depending on available materials and advice by professionals. Neither of these details is important for the argument in this paper, which is that small, affordable ECs represent a feasible and affordable way to protect occupants of single-family dwellings from heavy earthquake impacts. The average cost of an EC is assumed to be no more than \$500–\$600 in developing countries with

earthquake problems. A tornado shelter in the USA costs at a minimum \$1000.

This cost of ECs is understood to be significantly lower than retrofitting an entire building. Estimates of costs of retrofitting vary. For some weak structures, it may be as high as 75% of the value of the dwelling (Yoshimura and Meguro 2004). Other proposals exist that envision lower costs. In this paper, no judgment is passed on methods of retrofitting; the focus is solely on the cost/benefit ratio of ECs in two examples of earthquakes that have killed nearly 100,000 people. However, the proposed method can be used to estimate cost/benefit ratio for other earthquake protection devices.

Assumptions and Facts.

1. The warning for inhabitants that a possibly dangerous earthquake is happening is assumed to be provided by the P-wave. When the ground shakes strongly, the trained inhabitants recognize the earthquake in progress, and have several seconds to dash into the EC before the larger and more destructive S-wave arrives. During the shaking by the P-wave, walking itself is not a problem because the frequencies are high and the amplitudes are low. False alarms are few or zero because strong shaking is clearly distinguishable from that in minor earthquakes. For example, during the rich aftershock sequence of the M7.8 Gorkha earthquake of 2015 only the Ms7.3 earth-

✉ Max Wyss
max@maxwyss.ch

¹ International Centre for Earth Simulation Foundation,
Geneva, Switzerland

quake located at a distance of 70 km from Kathmandu required action in that location.

- The present thought experiment assumes that in affected villages all households have installed an EC, and all dwellings have one to two stories. In towns, taller buildings exist, where ECs are not effective, so only a fraction of families is assumed to have installed them.
- Further, it is assumed that only a fraction of those with an EC is at home and able to get into it, when shaking starts. What these fractions are can be debated and the reader is invited to assume different values if those proposed here seem not satisfactory.
- It is assumed that two examples of population distribution and building properties around a recent great earthquake ruptures are useful for estimating the numbers of lives saved.
- The occupancy rate of dwellings with ECs is assumed to be 60% at the time of the earthquake. This is an average between the maximum and minimum occupancy rates at night and in the middle of the day. Schools, hospitals, and offices have different occupancy rates, but are not considered here.
- Values for loss of life given in the literature vary greatly. Here, values that seem reasonable are assumed, but other values can be substituted. However, we must not lose track of the most important fact, not assumption, in this thought experiment: For a head of household, the benefit of saving the life of a family member is infinite. This is also the attitude of first responders who mobilize and rush across continents to risk their own lives in saving injured earthquake victims.

The distribution of the affected population and assumed earthquake scenarios

The preoccupation with the population of large cities being the chief victims in great earthquakes is shown to be wrong by the facts (Wyss 2018). It is understandable that the imagination of most is caught by accounts of tens of thousands people perishing in earthquakes for example in Lisboa (1755, M8), Messina (1908, M7.1), Bam (2003, Mw6.6), and Kobe (1995, M6.9). In these cases, major earthquakes happened near large cities, but in most great earthquakes the affected population in villages is about 90%. In a rupture several 100 km long, villages are numerous and large cities are few, or do not exist, as clearly seen in the example of the Wenchuan earthquake (Mw8 2008) (Fig. 1). The percentages of people who died in villages and towns were 98% and 99% in the two examples that are analyzed here, namely in the Wenchuan (M8, 2008) and Kashmir (M7.6, 2005) earthquakes, respectively. For this reason, the EC proposed here as a tool for protection against earthquakes is far more

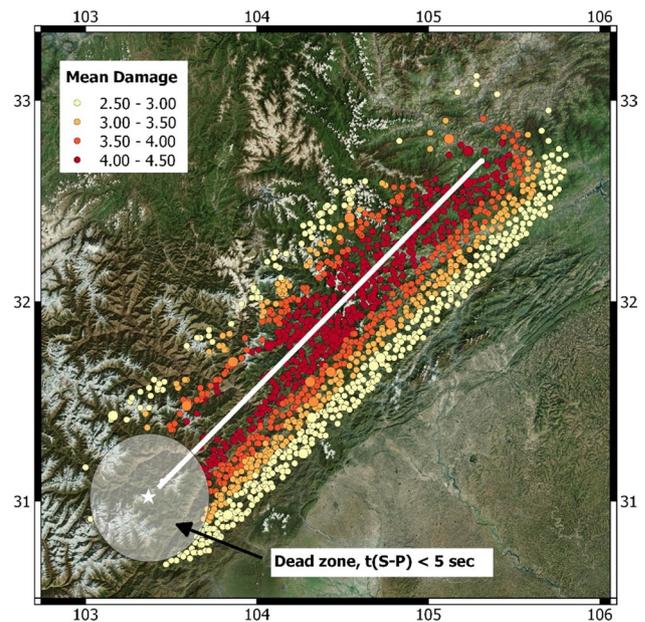


Fig. 1 Map of the mean damage (calculated by QLARM, Li et al. 2018) of the Wenchuan M8 earthquake with the model for the rupture shown as a white line. The star marks the epicenter (<https://earthquake.usgs.gov/earthquakes/eventpage/usp000g650/executive>). Each dot is a settlement within 46 km of the rupture, experiencing theoretically calculated mean damage (scale from 0 to 5), according to the color shown in the legend. Settlements outside this zone are assumed not to need an EC. Cities larger than 60 K inhabitants are not plotted because they are assumed not to be amenable to constructions of simple ECs. The transparent circle marks the “dead zone” where the travel time ($S-P$) is too short for owners of ECs to reach safety. The two size classes recognized (villages and towns) are marked by different size dots

applicable than one thinks at first blush. This study therefore addresses the chances of survival of the vast majority of the population in most large earthquakes.

For both of the examples used here, the following parameters are known: magnitude, length and width of the rupture plane, the precise location, the strike of the rupture, and the name of the generating fault, as well as the intensity distribution. These parameters are based on source inversions, aftershock distributions, geodetic measurements, and field surveys (e.g., Zhang et al. 2008; Chen and Booth 2011; Pathier et al. 2006; Thakur et al. 2006). (B) The number of fatalities and models for the distribution of buildings in vulnerability classes (Li et al. 2018; Wyss et al. 2018 and references therein) has been applied to test cases successfully.

The two examples of specific earthquakes are: the earthquakes at Wenchuan with M8 of May 12, 2008, and the one in Kashmir with M7.6 of October 8, 2005. Both represent cases in densely populated areas. The rupture length of the Wenchuan earthquake was about 300 km (Fig. 1).

The effectiveness of an EC is highest at a range of distances determined by time separation of the two body waves

$dt = t_S - t_P$ (t_S and t_P are the travel times from the epicenter to the EC, for the S- and P-waves, respectively). The time delay by which the S-wave follows the P-wave depends on the distance of travel, because the slower wave falls more and more behind the faster one with distance from the epicenter. In all strong earthquakes that cause serious damage, there are three groups of people with respect to earthquake danger. (A) Those who live so close to the earthquake initiation that they have no time to get to safety. These people cannot be helped by the availability of an EC. (B) That part of the population that lives far enough from the earthquake source that they do not need an EC because the likely damage to their dwellings is not life threatening. (C) The segment of the population that can be helped by having an EC lives in a range of dt between groups (A) and (B). In this case, the shaking is strong enough to be life threatening, and dt is long enough so inhabitants may move to safety.

The minimum time needed to get into an EC in a single-family home is assumed to be 5 s. This corresponds to a distance of 37.5 km from the epicenter (Fig. 1), based on velocities of the P- and S-waves of 5 km/s and 3 km/s, respectively. Thus, settlements where an EC helps, have to be farther from the epicenter than 37.5 km. Settlements that are beyond serious damage do not need an EC. In the case of the Wenchuan M8 earthquake, a cutoff distance of 46 km from the nearest point of the rupture line is used, corresponding to the extent of intensity $I=8$, which corresponds to a mean damage in a settlement of 2.5 on a scale of 0–5 (1 denotes negligible to slight damage, 5 means destruction; Gruenthal 1998). At this intensity of shaking the damage is life threatening, beyond less so. The same distance is also applied to the M7.6 Kashmir earthquake where it corresponds to $I=7.4$ and a mean damage to 2.25. This is so because the magnitude in Kashmir was lower than in China, but the strength of dwellings in the area of the Kashmir earthquake is inferior to those in the Wenchuan area (Table 1). These damage levels, and therefore, the probability of loss of life is comparable in the two examples; thus, the same distance from the rupture is assumed as a cutoff. For different cases of dwelling strengths and magnitudes, other levels of mean damage, and hence cutoff distances, can be assumed.

The population for whom an EC is useful follows from the aforementioned assumptions, as all the people living

within 46 km from the rupture line, minus those living within a radius of 37.5 km from the epicenter (Fig. 1). In both earthquake scenarios, small settlements are separately analyzed from medium sized ones. They will henceforth be called “villages” and “towns,” respectively.

Occupancy rate and total cost for ECs

The occupancy rate of dwellings in most countries is a maximum of approximately 95% at night and a minimum of 25% during the day. For the current exercise, it is assumed that the occupancy rate is the average, namely 60%, but any other number may be used to calculate different results.

The type of dwelling in which an EC is effective is a one- to two-story single-family home. It is assumed that these types of homes constitute 100% in villages. For the next larger class of settlements, the towns, it is assumed that only 40% of dwellings are amenable to an EC because some are multistory constructions. For cities larger than those listed in Table 1, it is assumed that no buildings exist in which a simple EC can be useful; thus, it is assumed that in larger cities no fatalities can be saved by ECs. However, in more detailed analyses the percentage of dwellings in large cities in which ECs are effective could be added and their contribution to saving lives may not be zero as assumed here.

The number of settlements within the radius of the S–P time of 5 s is listed in Table 2, along with the number in the useful range. ECs constructed in both of these areas are contributing to the cost, although those close to the epicenter turn out not to be useful. The size of an average family in China is three (<https://www.statista.com/statistics/278697/average-size-of-households-in-china/>) and in Pakistan it is eight (Nayab and Arif 2012) (Table 2). The number of families that are assumed to live separately is derived by dividing the sum of the population in all settlements considered by three and eight, respectively, for China and Pakistan. The cost to install an EC is assumed to be \$500 in China, half the least expensive tornado shelter in the USA. For Pakistan, the cost of an EC is assumed to be \$600 because it has to accommodate larger families.

Table 1 Model for the distribution of the building stock in villages of China and Pakistan/India using the vulnerability classes of the EMS-98 scale, with class A being the least resistant to shaking. This model

Country	Villages			Towns			
	A	B	C	A	B	C	D
China	50	35	15	5	10	70	15
Pakistan/India	71	27	2	42	45	13	0

is approximate, but has been used successfully in validation tests for both of these areas (Li et al. 2018; Wyss et al. 2018)

Table 2 Estimated total cost of ECs constructed within the area of distance 46 km from the rupture, including settlements within 37.5 km of the epicenter. It is assumed that each family builds an EC, and that there are three and eight people per family in China and Pakistan/India, respectively

Settlement size population thousands	Number of settlements in dead zone (S–P) < 5 s	Number of settlements in useful range	Number of families useful range	Cost per closet US \$	Total cost ECs millions US \$
<i>Wenchuan earthquake M8</i>					
< 6	146	1244	2,073,333	500	1158
6–40	1	18	240,000	500	121
<i>Kashmir earthquake M7.6</i>					
< 10	1657	3075	544,275	600	503
10–20	108	123	152,739	600	183

Estimates of fatalities and injured in the area of ECs

The likely numbers of fatalities and injured is a function of the people present and of the construction quality of the houses (Tables 1, 2). These two parameters are contained in QLARM, the computer code and data set used for calculating the damage field and casualties (Trendafiloski et al. 2011; Wyss et al. 2018). For China and India, QLARM has been verified to calculate correct numbers of casualties (fatalities plus injured) within a factor of about 2.5 in general (Li et al. 2018; Wyss et al. 2018). In the case of the Wenchuan and the Kashmir earthquakes, the average numbers of fatalities calculated were within factors of 1.1 and 1.2 of the reported values, respectively (Li et al. 2018; Wyss et al. 2018). Therefore, the casualties calculated by QLARM for the population segment of interest can be accepted as close to reality. The theoretical values for fatalities and injured in the regions of EC usefulness in the two test earthquakes are given in Table 3, separately for villages and towns.

The numbers of casualties saved, by people seeking refuge in ECs, must be assumed to be a fraction of those

living in dwellings with ECs. Some are likely not at home at the time of the earthquakes, and additional people are assumed to have been unable to reach the EC for some reason. Therefore, it is proposed that in this thought experiment only 25% of EC owners reached safety in their ECs in villages (Table 3) and have survived the earthquake. For inhabitants of towns in China, it is assumed that only 40% of the buildings are amenable for ECs; thus, only 10% of the population would reach safety (Table 3). In Pakistan/India where less advanced buildings prevail (Table 1), it is likely that a higher fraction of buildings in towns can benefit from an EC; thus, we assume that 15% of the population can reach safety in that case (Table 3).

Estimates of the value of life and cost of injured

The standard estimation of the value attached to life is called SVL (e.g., Environmental Protection Agency, definition in <https://www.epa.gov/environmental-economics/mortality-risk-valuation>). Different agencies and

Table 3 Potential numbers of fatalities and injured avoided by the use of ECs, and cost/benefit ratios. Numbers of fatalities calculated by the verified tool and data set QLARM (Wyss et al. 2018; Li et al. 2018)

Settlement size in population	Fatalities calculated	Injured calculated		SVL million US \$	Benefit (fat) million US \$	Benefit (inj) million US \$	Cost/benefit
<i>Wenchuan earthquake M8</i>							
< 6 K	72,500	185,475	25% Fat	25% Inj	1	18,125	185
			18,125	46,369			
6–40 K	1721	4518	10% Fat	10% Inj	1	172	2
			172	452			
<i>Kashmir earthquake M7.6</i>							
< 10 K	39,964	106,593	25% Fat	25% Inj	1	9991	107
			9991	26,648			
10–20 K	4453	12,999	15% Fat	15% Inj	1	668	8
			688	1950			

The percentages of fatalities and injured that are assumed to be saved are listed in the bold font

authors use various data sources and methods to arrive at SVLs in a number of contexts. Comparisons including SVLs in several countries have been published (Miller 2000; Viscusi and Aldy 2003). Based on data from 1995, Miller gives as a maximum for both India and China SVL (1995) = \$ 0.7 million. Viscusi states that SVL (India, year 2000) ranges from \$1.0 to \$4.1 million. In developed countries, the SVL ranges currently from \$2 to \$9.4 million (e.g., https://www.transportation.gov/sites/dot.gov/files/docs/VSL_Guidance_2014.pdf, <https://www.transport.govt.nz/assets/Uploads/Research/Documents/Social-cost-of-road-crashes-and-injuries-2016-update-final.pdf>, https://www.pmc.gov.au/sites/default/files/publications/Value_of_Statistical_Life_guidance_note.pdf).

In specific cases, governments may decide on recompense to families that lost a member in an earthquake. These payments are not based on SVL estimates. Such decisions may be based in part on the argument that the government has not committed any negligence and is not at fault for earthquake fatalities. Thus, to pay any sum to help surviving family members may be viewed as a voluntary act by a government and is not to be taken as the value of a life, instead it is intended simply to get the survivors back on their feet. For example the Chinese government paid the equivalent of \$720 per person killed by the Wenchuan earthquake (http://epaper.xxcb.cn/xxcba/html/2013-04/28/content_2703783.htm). Last access: June 1, 2018; <http://mzzt.mca.gov.cn/article/yqwlr/m/xgzcwj/200807/20080720018613.shtml>. Last access: June 1, 2018).

In this study, a generic value of SVL (generic) = \$1 million is used, with the intention that the present results can be applied to earthquakes in any country, by varying the SVL appropriately. In the two countries in which the example earthquakes took place, SVL(generic) is on the high side, but for developed countries, it is low.

A person is counted as injured if a hospital stay is required. The cost of such can vary from moderate to extremely high, including rescue and transport charges to the hospital. An average of \$4000 per injured person is assumed for the test cases. This value is at the upper end of numbers reported by several hospitals in Sichuan province where the Wenchuan earthquake occurred (Li et al. 2008; Zhang et al. 2008).

Cost/benefit estimate

From the point of view of a head of family, the benefit of a single saved life of the family is infinite. Also, a head of family thinks not only of the present, but also of future generations, caring about the fate of great grandchildren. In this view, a large earthquake on an active fault nearby that has

not ruptured for a couple of 100 years is highly likely to occur during the integrated time of the family living there. Therefore, lives are almost certainly saved in the family and their descendants. Thus, the benefit is infinite at a cost of \$500 and the cost/benefit ratio is 0 from this point of view.

For a government, the cost/benefit ratio is tangible where rescue, transport, and hospital costs are incurred for the injured. If it is assumed that hospitalization of an injured person costs on average \$4000, then the benefits for injured alone amounts to between \$100 million and \$200 million in the scenario earthquakes assumed (Table 3). Combining this with the benefits of saved lives, assuming an SLV of \$1 million, the cost/benefit ratio is estimated as 0.063 and 0.05, for villages in the two test earthquakes, respectively (Table 3). For towns, where more resistant structures exist (Table 1), the cost/benefit ratios are 0.696 and 0.271, respectively, in the two test cases (Table 3).

Discussion and conclusions

For a head of family living near a large active fault, this whole discussion is unnecessary. He/she simply spends the few hundred dollars it costs to increase the survival probability of family members by more than 1000 (Wyss 2012) for generations to come. For governments and organizations attempting to reduce earthquake risk, the two examples outlined here may be useful. The various parameters are specified and can be changed by interested parties to model different scenarios, resulting in different cost/benefit ratios that may apply elsewhere and for different earthquake magnitudes.

The two examples of large, devastating earthquakes simplify the problem of earthquake risk reduction to the retrospective analysis of ruptures that have happened and caused known numbers of fatalities. The present analysis does not offer a solution to what might be the ultimate question facing a government: “What should it do to protect all its citizens from earthquakes of all sizes possible on all active faults with different probabilities to occur?” Instead, I propose that protection from earthquakes should start with the simplest step: identify the most well-defined, active faults, define the region with high probabilities of substantial numbers of fatalities, and install earthquake closets in that area. In such a first step to prepare for future devastating earthquakes, all input parameters have to be estimated, based on models.

In the test cases, the input parameters for the earthquake sources, the population present, and the approximate building stock model are known relatively well. The casualty calculations are reliable because the method and data set to calculate them has been found to be correct in both areas (Li et al. 2018; Wyss et al. 2018). In addition, the total numbers of fatalities and injured are known in both

examples. The occupancy rates and the prices of unit costs have been assumed and may be changed if information exists that suggests different values. The present thought experiment is intended to stimulate similar applications in different circumstances and with different assumptions, but it is proposed to focus on well-defined faults, capable of large ruptures.

Earthquake protection units can have different designs and different costs. Various proposals address needs of different parts of society. An important situation that is not addressed in the present discussion is children in school. At the time when school is in, specially constructed heavy combined desks may help in reducing casualties. A very strong construction of desks is a necessary condition for the slogan “drop-cover-and-hold” to work. In the countries under discussion here, stone buildings collapse. At home, no one has a table strong enough to protect themselves from a falling ceiling. In this case, the slogan “drop-cover-and-hold” is misleading in the worst way. Children do benefit from ECs for about half of their weekday lives, and for full days during weekends and holidays, namely when they are in their homes. Likewise, ECs do not benefit factory and office workers during their time at work. However, the small settlements where most of the benefits calculated result in the scenarios presented here have neither factories nor office buildings.

The cost/benefit values calculated in the two scenarios presented are subject to several assumed values. These are clearly defined, and the reader can choose different values to experiment with. With the assumptions used however, the order of magnitude estimate shows that about \$15–\$20 are saved for every \$1 spent (Table 3). Such amounts are calculated for developing countries with buildings vulnerable to shaking and an SVL = \$1 million. The results are solid because they are based on specific, tested cases.

The method of protection against earthquakes proposed here would have been applicable to 98% and 99% of the population that died in the Wenchuan and Kashmir earthquakes, respectively. Although great cities located near active faults, like Istanbul and Tehran, are definitely of great concern with respect to earthquake losses, 90% of fatalities in most great earthquakes are found among the rural population (Wyss 2018). The focus on villages proposed here is truly what matters in large earthquakes.

Retrofitting existing dwellings in earthquake zones would be desirable, but most sources view this as prohibitively expensive. It is therefore strongly suggested that home owners, governments, and non-profit helpers consider installing ECs within 46 km of active faults, where large earthquakes have occurred in the relatively distant past, but will strike again at some unpredicted time.

Acknowledgements I thank J. W. Li for supplying references for costs for earthquake victims in China and Luis Triveno for references on earthquake protection costs. I also thank R. Console, R. Bishop, and S. Gupta for comments on the manuscript.

References

- Chen Y, Booth DC (2011) The Wenchuan earthquake of 2008: anatomy of a disaster. Science Press cooperating with Springer, Beijing
- Gruenthal G (1998) European Macroseismic Scale 1998, Cahiers du Centre Européen de Géodynamique et de Séismologie. Conseil de l'Europe, Luxembourg
- Li X, Chen M, Yin DH, Shao Y, Zhou Z, Chen Q (2008) Treatment analysis of 563 wounded patients in the Third People's Hospital of Chengdu after the Wenchuan earthquake. *Chin J Evid Based Med* 8:597–601
- Li JW, Wyss M, Wu ZL, Zhou SY (2018) Estimated casualty risk for six scenario great earthquakes in Sichuan-Yunnan region, China, for disaster preparation. In: International conference on the decade memory of the Wenchuan earthquake, edited, p 109, Ministry of Science and Technology of the People's Republic of China, Chengdu, China
- Nayab D, Arif GM (2012) Pakistan panel household survey sample site, attrition and socio-demographic dynamics. In: Poverty and social dynamics paper series, 1, Pakistan Institute of Development Economics, Islamabad, 28pp
- Pathier E, Fielding EJ, Wright TJ, Walker R, Parsons BE, Hensley S (2006) Displacement field and slip distribution of the 2005 Kashmir earthquake from SAR imagery. *Geophys Res Lett* 33:20. <https://doi.org/10.1029/2006GL027193>
- Thakur V, Jayagondaperumal R, Champatiray PK, Bhat MI, Malik M (2006) 8 October, 2005 Muzaffarabad earthquake and seismic hazard assessment of Kashmir gap in Northwestern Himalaya. *J Geol Soc India* 68:187–200
- Trendafiloski G, Wyss M, Rosset P (2011) Loss estimation module in the second generation software QLARM. In: Spence R, So E, Scawthorn C (eds) Human casualties in earthquakes: progress in modeling and mitigation. Springer, Berlin, pp 381–391
- Viscusi WK, Aldy JE (2003) The value of statistical life: a critical review of market estimates throughout the world, NBER working paper series. Working paper 9487, National Bureau of Economic Research, Cambridge, MA, 126pp
- Wyss M (2012) The earthquake closet: rendering early-warning useful. *Nat Hazard* 62:3927–3935. <https://doi.org/10.1007/s11069-012-0177-6>
- Wyss M (2018) Rural populations suffer most in great earthquakes. *Seismol Res Lett* 89(6):1991–1997. <https://doi.org/10.1785/0220180236>
- Wyss M, Gupta S, Rosset P (2018) Casualty estimates in repeat Himalayan earthquakes in India. *Bull Seismol Soc Am* 108(5A):2877–2893. <https://doi.org/10.1785/0120170323>
- Yoshimura M, Meguro K (2004) Proposal for retrofitting promotion system for low earthquake resistant structures in earthquake prone countries. Paper presented at 13th world conference on earthquake engineering, Vancouver, BC, Canada, August 1–6
- Zhang J, Chen W, Meng J, Du L, Wen J, Liu GJ, Li YP (2008) Research on structure and influencing factors of overall medical treatment expenses of 1635 inpatients of Wenchuan earthquake in West China Hospital. *Chin J Evid Based Med* 8:703–709



Teleseismic moment tensors of the 5 April 2017, Mw6.1, Fariman, northeast Iran, earthquake

Mohammad Ashtari Jafari¹

Received: 26 June 2018 / Accepted: 28 January 2019 / Published online: 5 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

On 5 April 2017, an Mw6.1 earthquake occurred about 50 km NE of the city of Fariman, northeast Iran. Several hundreds of aftershocks including two $M > 5$ events followed the main shock. The quake struck numerous towns and villages across the region, killed one person, and injured tens of people. Many schools and universities were evacuated around the epicentral area, and a lot of people left their residences for a few days. The northeastward motion of the central Iran toward Eurasia influences the epicentral region. Regional movements occur by shortening on the northwest-trending reverse faults. We studied teleseismic source parameters of this earthquake by applying different moment tensor decomposition methods including grid search for the nodal planes of the best double couple; linear inversion for a deviatoric moment tensor; grid search for the best double-couple moment tensor; grid search for the best deviatoric moment tensor; and grid search for the best full moment. Based on the moment tensors, the event occurred on a reverse fault following the regional compressional motion. The results of this study will provide useful information for future regional seismotectonic investigations and are of significant use for applications such as regional seismic hazard evaluations.

Keywords Nodal planes · Moment tensor inversion · Double couple · Deviatoric moment tensor · Full moment tensor

Introduction

On 5 April 2017, an Mw6.1 earthquake happened about 50 km northeast of Fariman in northeastern Iran (Table 1) that was felt through many surrounding districts. As reported by local media, the event left one dead and tens of injuries. There are reports of landslides, rock slides, fractures on the Earth's surface, and collapse of roofs and fractures in the walls of the human-made structures. The constructions in this area may be divided into two different groups: (a) built by brick, mud, and stone and (b) constructed considering engineering regulations. Buildings from both groups may damage during the earthquakes indicating that low adherence to seismic engineering construction regulations can

increase damages caused by medium-sized earthquakes in Iran (Ashtari Jafari 2008, 2016).

In this paper, we have studied teleseismic source parameters of the Fariman earthquake by employing different methods that take advantage of grid search and linear inversion methods. We applied grid search method to estimate the best (i) nodal planes, (ii) double-couple moment tensor, (iii) deviatoric moment tensor, (iv) full moment tensor, and (v) linear inversion for deriving the parameters of the deviatoric moment tensor. The moment tensor solutions provide important information, such as source process, source type (faulting type), seismic moment and moment magnitude of an earthquake. As the area under consideration suffers lack of instrumental source studies, the results of this study can not only help to better understand the regional sources (so the regional seismotectonics) but also be useful for other applications such as regional seismic hazard evaluations.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11600-019-00256-8>) contains supplementary material, which is available to authorized users.

✉ Mohammad Ashtari Jafari
muhammad@ut.ac.ir

¹ Institute of Geophysics, University of Tehran, P.O. Box 16765-533, Tehran 16, Iran

Table 1 Fariman earthquake location parameters based on local and international Agencies

Agency	Origin time—UTC	Latitude	Longitude	Magnitude
IRSC	2017-04-05-06:09:08	35.85°N	60.34°E	6.0 MN
GCMT	2017-04-05-06:09:12.55	35.81°N	60.37°E	6.0 Mw
USGS	2017-04-05 06:09:12.20	35.78°N	60.44°E	6.1 Mww

IRSC Iranian seismological center, *GCMT* global centroid moment tensor, *USGS* US geological survey

Seismicity and seismotectonics

Tectonic deformations in Iran are the consequence of the northward motion between the Persia and Eurasia that is represented by crustal shortening and strike-slip faulting. Based on Shabanian et al. (2009), the northward motion is accommodated in northeast Iran at a rate ranging from 4 to 11 mm per year. Around the Binalud Mountains, there is 2–4 mm per year of the right-lateral strike-slip displacement on both sides of the mountains. They suggested that northward movement of the central Iran and the Lut Block with respect to Afghanistan has to be taken up between the Doruneh fault and the Kopeh Dagh ranges with trivial absorption across the northern part of central Iran. Some piece of this rearrangement must be transferred between the Kopeh Dagh and the Binalud ranges in the form of strike-slip motion along localized major fault systems or as distributed deformation (Shabanian et al. 2009, and references therein). On the other hand, the northeast Iran seismotectonic regime is locally under the influence of the Kopeh Dagh Mountains that spread for several hundreds of kilometers. This range is separated in the south from the eastern Alborz Mountains by various reverse faults. Its west-central part extends to the Caspian Sea by the Atrak River, and its southeastern segment drains to the Kashaf-Rud depression (Berberian et al. 2000). The regional tectonics confirms the existence of compressional motion represented by the reverse faults such as the Torbat-e-Jam, Kashaf-Rud and North Neyshabur faults. Based on the non-instrumental data, the region has experienced three historical events occurred in 840 A.D., 1673 A.D. and 1678 A.D. The 1673 and 1678 events have caused severe destructions in Mashhad and Neyshabour and their surroundings (Berberian et al. 2000). According to instrumental data (1925–2005), the epicentral area has not suffered medium–strong quakes during this period (www.isc.ac.uk, last accessed 1 June, 2017). The instrumental data after 2006 from Iranian Seismological Center (IRSC) show diffused seismic activity across the same region (irsc.ut.ac.ir, last accessed 1 June, 2017). After the seismicity parameter maps of Iran compiled by Ashtari Jafari (2013), the

regional “a” and “b” values have been estimated to be around 4–4.5 and below one, respectively. The existing CMT solutions (www.globalcmt.org, last accessed 1 June, 2017) confirm the compressional motion along the reverse faults (Fig. 1a).

Data and methods

The seismic moment tensor produces an appropriate presentation for any arbitrarily oriented earthquake source by involving isotropic part (ISO), a double couple (DC), and compensated linear vector dipole (CLVD) components. The moment tensor may be decomposed in several ways, e.g., ISO plus DC and CLVD parts or ISO plus major and minor parts. For a review of the theory and applications of moment tensors, the reader can refer to Jost and Herrmann (1989) or Julian et al. (1998).

Grid search for double-couple source parameters

In this section, we estimated the depth, moment magnitude, strike, dip, and rake angles assuming a double-couple focal mechanism. For this purpose, we have replaced a direct moment tensor inversion by estimating the best earthquake mechanism parameters by grid search over the DC nodal plane orientations with depth. For each depth, observed traces are associated with pre-prepared Green’s functions (based on the AK135 velocity model) for that depth and for the source–receiver distances. The method consistently examines the entire span of the source variables (strike angles [0, 355], dip angles [0, 90], and rake angles [−90, 90]) by fitting the P, SV, and SH synthetic and recorded seismograms at each station. At every grid point, the approximated synthetic velocity ground motions were checked with the recorded data by assessing a suitable norm (L_2 norm) of the misfit between them. For any considered source depth, the change of the strike, dip, and rake was considered to be in ten degree steps while the source depth was changed in 1 km increments to get an idea of the resolution. Then, the error function was evaluated for the recorded and estimated seismograms (P, SV, and SH) for each estimate of the strike, dip, and rake. After computing error for all the considered values of the strike, dip, and rake and deriving the first best model, we made a new search within a new interval $\pm 20^\circ$ around the first best model but in five-degree increments (Herrmann et al. 2011). The final evaluated minimum misfit is the best DC with depth.

Linear deviatoric moment tensor inversion

If all the moment tensor elements have the same source time histories, then the seismogram at location “x” and time “t” can be written as:

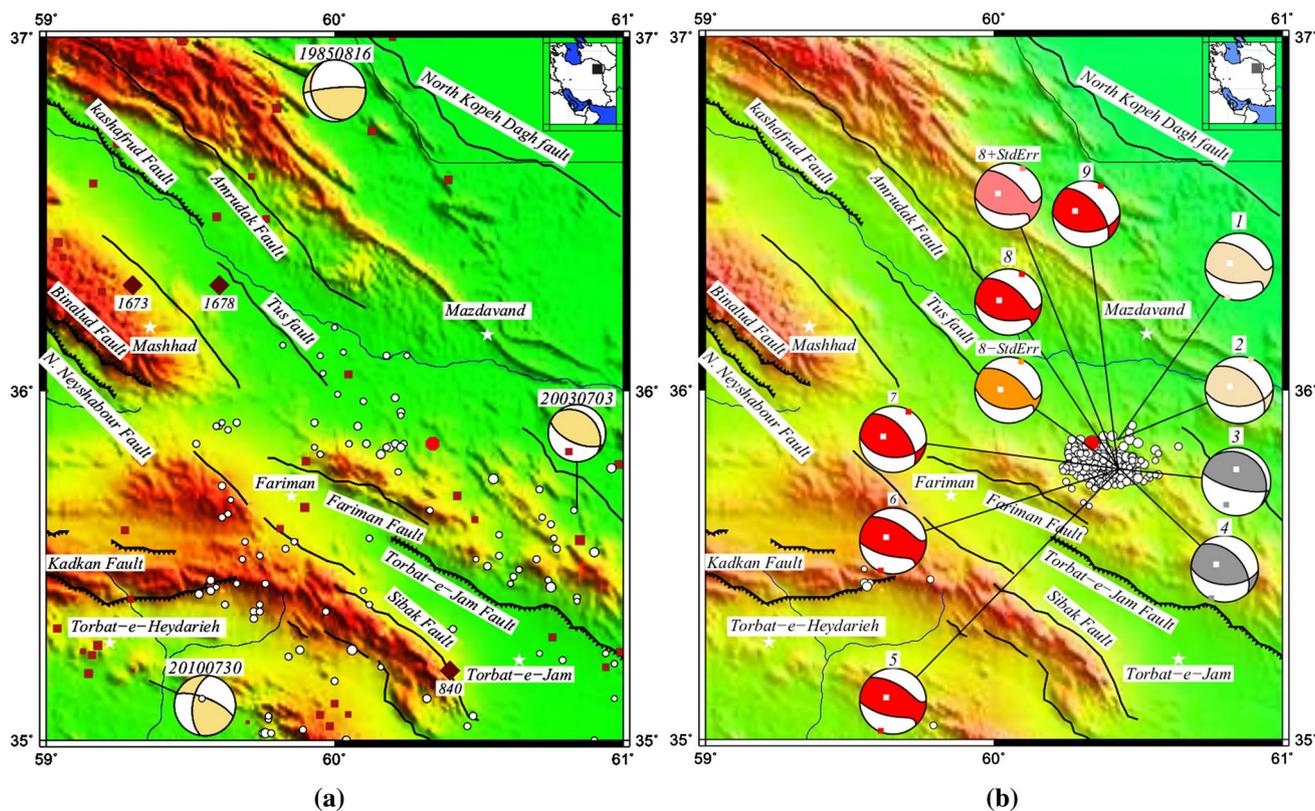


Fig. 1 **a** Regional seismicity and seismotectonics. Diamonds: historical seismicity; squares: instrumental seismicity (ISC: 1925–2005); circles: instrumental seismicity (IRSC: 2006–2017), stars: cities, beach balls: GCMT mechanisms. **b** Comparing the results of this study with GCMT and USGS. (1) GCMT moment tensor, (2) GCMT

nodal planes, (3) USGS W-phase nodal planes, (4) USGS body wave nodal planes, (5) best full-MT, (6) best DEV-MT, (7) best DC-MT, (8) best linear DEV-MT, *SE* standard error, and (9) best grid-searched nodal planes

$$S_i(\mathbf{x}_r, t) = M_{jk} [G_{ij,k}(\mathbf{x}_r, \mathbf{x}_s, t) * H(t)] \tag{1}$$

where S_i is the i th component of the seismogram; M_{jk} are the elements of the moment tensor; $G_{ij,k}$ are the derivatives of the Green’s functions; $H(t)$ is the time history of the source; \mathbf{x}_r is the position vector of the receivers; and \mathbf{x}_s is the position vector of the source. It is possible to rewrite the observed seismograms in the linear matrix form as:

$$\mathbf{S} = \mathbf{G}\mathbf{M} \tag{2}$$

One of the solutions to the above-mentioned overdetermined system of equations is the singular value decomposition. In order to apply the singular value decomposition, we decompose the matrix \mathbf{G} into $\mathbf{U}\mathbf{W}\mathbf{V}^T$, where \mathbf{W} is a diagonal matrix and \mathbf{U} and \mathbf{V} are orthogonal matrices. Finally, the solution vector (components of the moment tensor) is written as (Jost and Herrmann 1989):

$$\mathbf{M} = \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\mathbf{S} \tag{3}$$

Grid search for teleseismic moment tensors

The physical processes producing the non-DC components are not well understood. One of the key concepts in moment tensor computations is the seismogram inversion. In routine applications, we map the propagation and the source effects to generate the synthetic seismograms, in order to verify them with the corresponding recorded traces. Then, the best solution may be acquired by minimizing the difference between the synthetic and recorded data. In this section, we have estimated teleseismic moment tensors based on the procedure that has been developed by Zhu and Ben-Zion (2013) and implemented into the computer programs in seismology by Herrmann (2013). Zhu and Ben-Zion (2013) decomposed a general seismic potency tensor into isotropic, double-couple, and CLVD employing the eigenvalues and eigenvectors of the full tensor. Then, they have decomposed and parameterized the potency tensor to derive a general seismic moment tensor. They also have derived relations between different parameters of the potency and moment tensors in isotropic media. As the number of variables of this decomposition method is not high, we can utilize grid

search for implementing this method. The deviatoric grid search in this section is a particular instance when $ISO = 0$.

Source-type plot

The well-known beach balls display the fault plane orientations and the slip vector defining the shear dislocation across the faults. Another method for representing different kinds of seismic sources is the so-called source-type plot. This plot helps to visualize and to better understand the physical processes responsible for earthquakes. Several graphical methods based on different projections have been suggested for this kind of representation (Aso et al. 2016). We applied the diamond source-type plot that models the position of the source on a CLVD-ISO coordinate system (Vavrycuk 2015a, b). In this system, the DC components of the source are represented by color intensity. The sum of the absolute values of the CLVD and ISO is not larger than one, so the moment tensors lie inside the diamond. Shear sources are located around the origin and explosion/implosion sources at the top or bottom vertices of the diamond, and pure tensile or compressive cracks are plotted around the margins of the diamond (points in the first and third quadrants of the diamond correspond to tensile and shear sources, respectively). The points along the CLVD axis correspond to faulting on a non-planar fault. The elastic properties of the medium not only can control the angle of separation (Dufumier and Riverra 1997) but also have an effect on the ISO/CLVD ratio of the pure tensile and shear–tensile sources. In isotropic media, this ratio is:

$$C_{ISO}/C_{CLVD} = (4/3)(V_p/V_s)^2 - 1 \quad (4)$$

Scale factors C_{ISO} , C_{DC} , C_{CLVD} fulfill the following relation (Vavrycuk 2015a, b):

$$|C_{ISO}| + |C_{CLVD}| + DC = 1 \quad (5)$$

Data

We used well-recorded broadband GSN (Global Seismographic Network) and FDSN (Federation of Digital Seismograph Networks) waveforms of this earthquake, downloaded from IRIS DMC (Incorporated Research Institutions for Seismology: ds.iris.edu/wilber3, last accessed 1 May, 2017). We choose stations in the range of 30° – 90° from the epicenter to avoid waveform complexities caused by traveling in the upper mantle and core. The recorded traces were corrected for the instrument response, converted to ground velocity (m/s) between 0.004–5 Hz, low-pass filtered at 0.25 Hz, and interpolated to the sampling rate of 1 Hz. We selected the suitable traces after visual examination of the fitted seismograms to preserve the estimations against unstable solutions. Synthetic traces were calculated by employing

the AK135 Earth seismic velocity model (Montagner and Kennett 1996). We used a wide band of source depths, from 1 km to 20 km (in one-kilometer increments; for the first three methods) and from 1 to 10 km (in one-kilometer increments for the last two methods).

Results

We applied a grid search to find not only the best-fitting nodal planes of the 5 April 2017, Fariman, Iran, earthquake but also its teleseismic moment tensors (Fig. 1b). The details of the computations and results have been provided as a separate electronic supplement to this paper. By the grid search (Table 2), we systematically checked the entire range of source parameters (nodal planes and moment magnitude by depth) for their adaptability with the recordings (Fig. 2). We have checked the best nodal planes by the grid search over DC mechanisms with depth. For every grid point, we modeled the synthetic ground velocities and generated the L_2 norm of the misfit between the recorded and the estimated waveforms. The obtained minimum misfit provides the nodal planes of the best double couple with depth combination (Fig. 3a).

In order to estimate teleseismic moment tensors, we used different decomposition methods. The linear deviatoric moment tensor inversion found (Fig. 3b) the best-fitted deviatoric moment tensor (in terms of DC and CLVD percentages) and estimated 92.24% DC, 7.76% CLVD at a depth of 6 km. The grid search for the best double-couple moment tensor estimated 99.9% DC, 0.1% CLVD at a depth of 6 km (Fig. 3c). The grid search for the best deviatoric moment tensor computed 69.75% DC, 30.25% CLVD at a depth of 7 km (Fig. 3d). The grid search for the best full moment tensor calculated 68.07% DC, 29.37% CLVD and 2.56% ISO at a depth of 7 km (Fig. 3e). For direct inversion results, we made uncertainty analysis based on the standard error (Fig. 1b). Figure 5a–c shows the misfits and applied weights to each station component for direct inversion in this study. For grid search results, as the velocity model uncertainty biases the source parameter determination in an unknown way, we do not calculate uncertainty (Herrmann et al. 2011). Figure 4a–c illustrates the best fitted traces computed based on the grid search for the best double-couple nodal planes. Our estimated depths are near IRSC and less than GCMT and USGS W-phase depth estimations (Table 2). All the estimations predicted shallow depths for the main shock (below 10 km) and display the existence of regional compression.

The Kagan angle presents the 3D rotations by which one DC source can be rotated into another arbitrary DC source. Because of the symmetry of the DC sources, there will be four such rotations. For most practical purposes, if one of the rotations is small, we may pass over the other three

Table 2 Comparing teleseismic source parameters calculated from different methods employed in this study with other studies

Agency—Method	Strike (deg.)	Dip (deg.)	Rake (deg.)	Mag— M_0 (dyne \times cm)	%ISO—%DC—%CLVD	Depth (km)
IRSC	329	45	127	5.9 Mw	0	6
	101	56	58	9.81e+24	98.6 1.4	
USGS body wave	317	48	123	6 Mw	0	4
	93	51	95	1.22e+25	83.3 16.7	
USGS W-phase	316	20	120	6.1 Mww	0	11.5
	105	73	80	2.08e+25	87.93 12.07	
GCMT	312	53	117	6 Mw	0.04	12
	91	44	59	1.42e+25	91.44 8.52	
This study—grid-searched nodal planes	320	57	123	5.96 Mw	0	6
	90	45	50	1.01e+25	100 0	
This study—linear DEV-MT	314	55	117	5.96 Mw	~0.001	6
	93	43	57	1.09e+25	92.24 7.76	
This study—grid-searched DC-MT	320	55	120	5.94 Mw	0	6
	95	45	54	1.01e+25	99.97 0.03	
This study—grid-searched DEV-MT	311	46	117	5.931 Mw	~0.003	7
	95	50	65	1.06e+025	69.75 30.25	
This study—grid-searched FULL-MT	311	46	117	5.929 Mw	2.56	7
	95	50	65	1.05e+025	68.07 29.37	

DC double-couple moment tensor, *DEV* deviatoric moment tensor, *FULL* full moment tensor, *ISO* isotropic, *deg* degree, *mag* magnitude, M_0 scalar moment, *GCMT* global centroid moment tensor, *USGS* US geological survey, *IRSC* Iranian Seismological Center

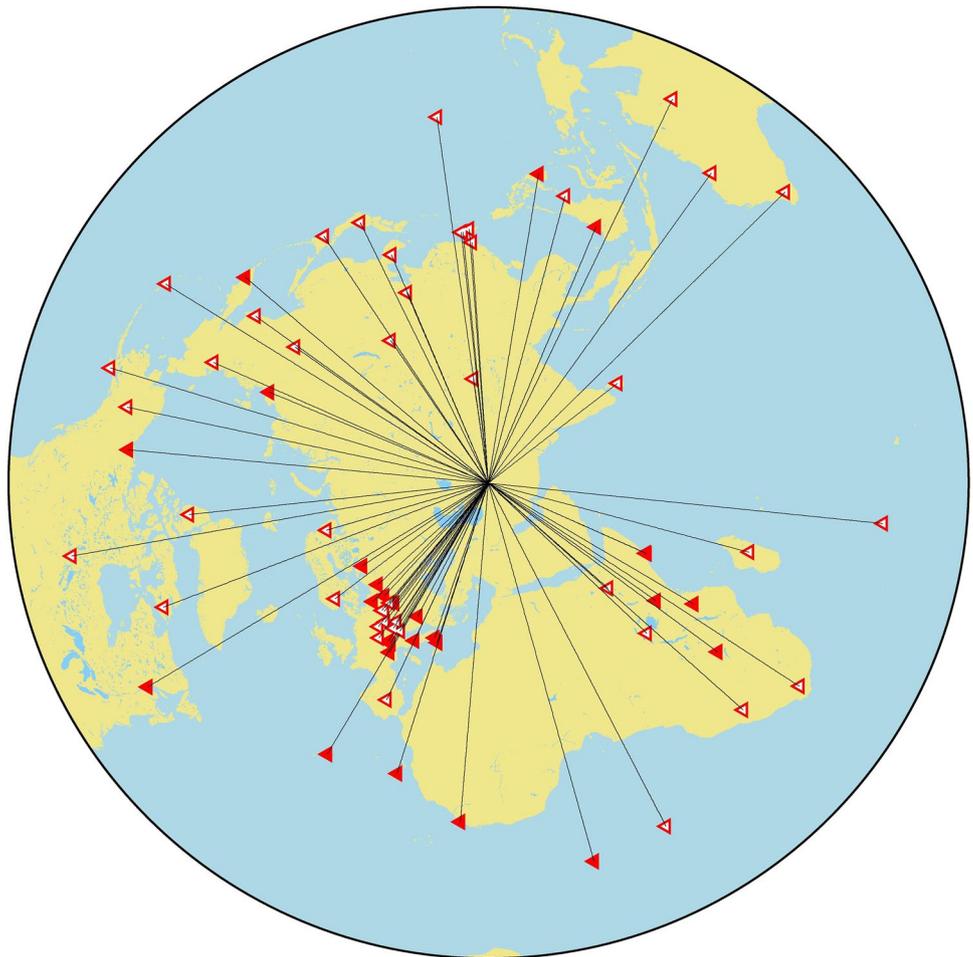
rotations (Kagan 1991). The complete agreement starts from 0 degrees to complete disagreement near 120 degrees. The values below 60 degrees may present acceptable agreement and above 60 degree a considerable mismatch (D'Amico et al. 2014). The minimum Kagan angle between the best-fitted nodal planes of the main shock derived from this study and the nodal planes of the GCMT is about 9.5 degrees. The minimum Kagan angle between this study's best-fitted nodal planes and the USGS W-phase is near 37 degrees.

Discussion

Shear faulting, represented by double-couple components, is usually considered as the most common mechanism for earthquakes. Physical processes responsible for the non-double-couple components are not well understood, yet

they have already been observed in landslides, volcanic and geothermal events, and mines (Miller et al. 1998). Non-DC earthquakes are valuable for enhancing our perception of how seismic events (natural or induced) work. Foulger and Julian (2015) gave a comprehensive review of the theory of non-DC events and several observed examples. The moment tensor inversions are routinely applied by the zero trace condition in order to stabilize the inversions and to justify shear earthquakes in isotropic media. Although, if the source is not shear or the focal area is anisotropic such inversions yield distorted results. The non-DC components can include both the ISO and CLVD components. The amount of the ISO and CLVD mechanisms depends on the strength and the symmetry of anisotropy and on the orientation of faulting (Vavrycuk 2005). Amount of CLVD estimated by linear DEV-MT in this study is comparable with GCMT result (Table 2).

Fig. 2 Global distribution of stations from 30 to 90 degrees around the epicenter: The solid red triangles show the P and a solid white triangle within them shows the SH and SV components used in this analysis



The teleseismic GCMT calculates the deviatoric moment tensor, the centroid location, and the time of an earthquake by minimizing the difference between the synthetic and observed waveforms in two steps. In the first step, the location is fixed to predetermined data (e.g., from PDE: Preliminary Determination of Epicenters) and then for the centroid location and time of the event. GCMT inverts three types of waves in different time windows and frequency bands: (1) body waveforms (40–150 s) which sample the focal sphere of the event in a relatively uniform manner, (2) large-amplitude medium-period surface waves (50–150 s) that can provide helpful information for relatively smaller earthquakes as they are less affected by noise, and (3) very long-period surface waves (125–350 s) known as mantle waves that are less sensitive to the spatial extent and temporal history of the source (Dziewonski et al. 1992; Ekström et al. 2012). Hjörleifsdóttir and Ekström (2010) showed that GCMT underestimated the seismic moment in areas of crustal thinning (10–20%), overestimated in areas of thick crust (15–20%) and overestimated in areas of low-angle thrust faulting associated with subduction zones (up to 54%). The results of

Konstantinou and Rontogianni (2011) agree in general with Hjörleifsdóttir and Ekström (2010) results.

The W-phase waves are very long-period interference of body waves that travel in the upper mantle and arrive after primary waves and before surface waves. The synthetic seismograms of the W-phase waves are estimated by the superposition of the normal modes. In practice, the time-domain deconvolution is used to derive W-phases from broadband seismograms. Its frequency band is around 100–1000 s. This phase is extracted from the vertical component instruments. The method applies linear inversion for a point source to calculate seismic moment and moment magnitude (Kanamori and Rivera 2008).

The epicentral area of this event is not well studied by seismicity and seismotectonic data, but based on the present information, the region is under the influence of the global compressional forces resulted from the convergence between the Africa and Eurasia. The area is locally predominated by the northeastward motion of the central Iran toward Eurasia. Regional tectonic and seismotectonic (Berberian et al. 2000) and instrumental source mechanisms confirm the existence of the mentioned compressional forces displayed

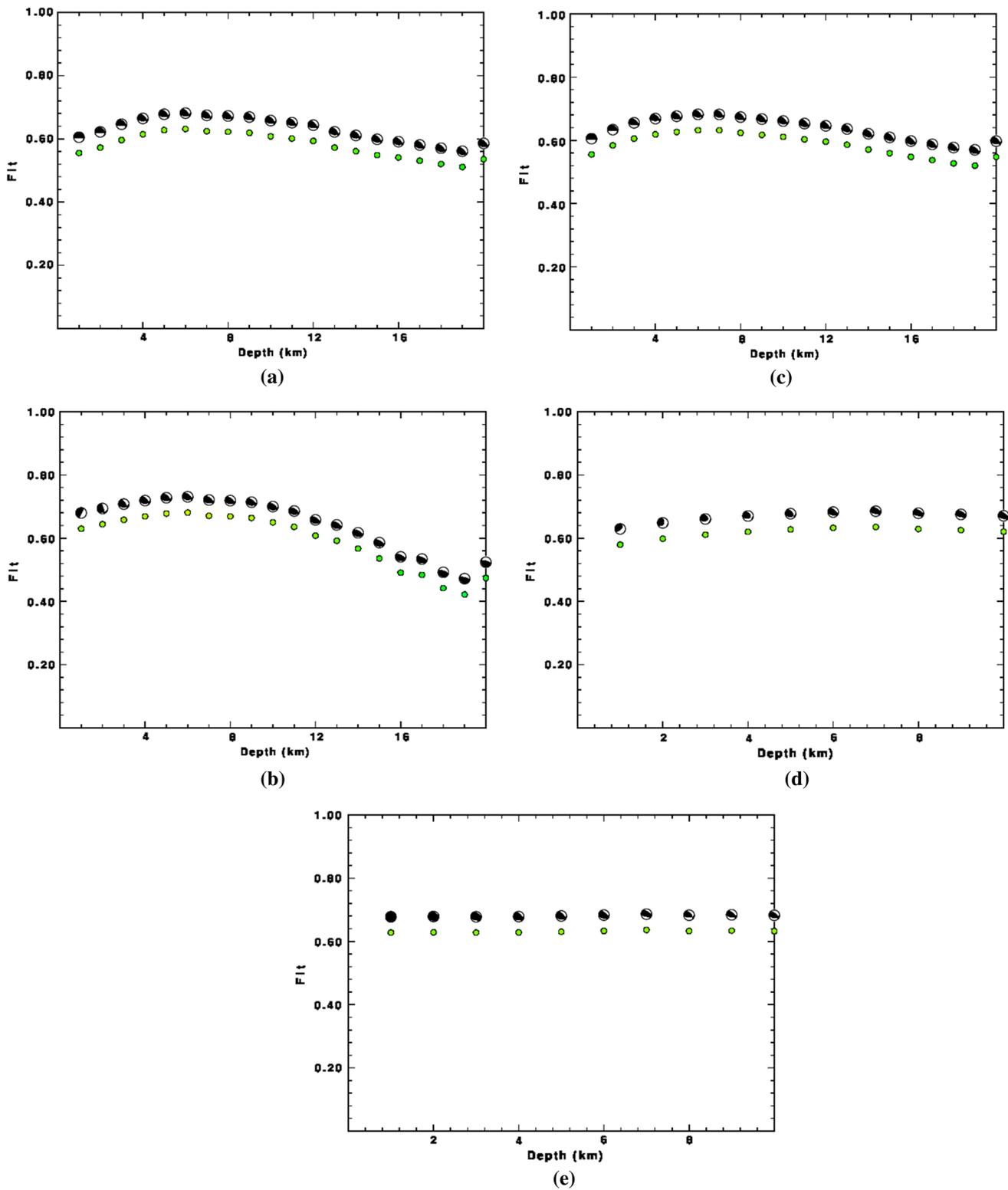
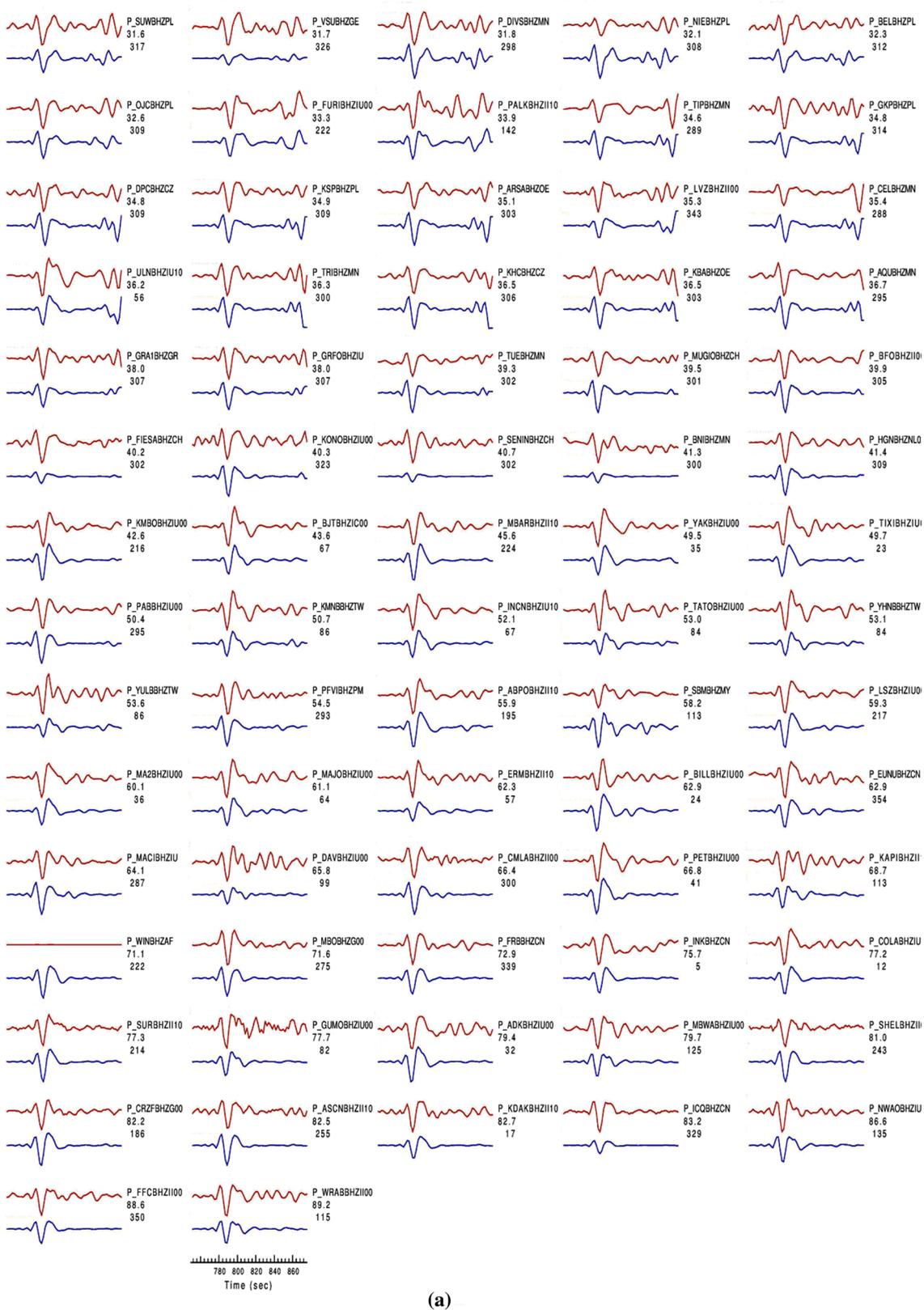


Fig. 3 Best-fitted teleseismic source mechanisms versus depth. **a** Grid-searched best-fitted nodal planes, **b** best linear deviatoric moment tensor, **c** grid-searched best-fitted DC-MT, **d** grid-searched

best-fitted DEV-MT, and **e** grid-searched best-fitted full-MT. Green circles show the numerical value of each result



(a)

Fig. 4 Best-fitted traces based on the grid search for: **a** P wave, **b** SH wave, and **c** SV wave components. The trace pairs are plotted by the station name, epicentral distance, and source to station azimuth in

degrees. The traces are drawn with the same scale, and name of the stations is written on the upper right side

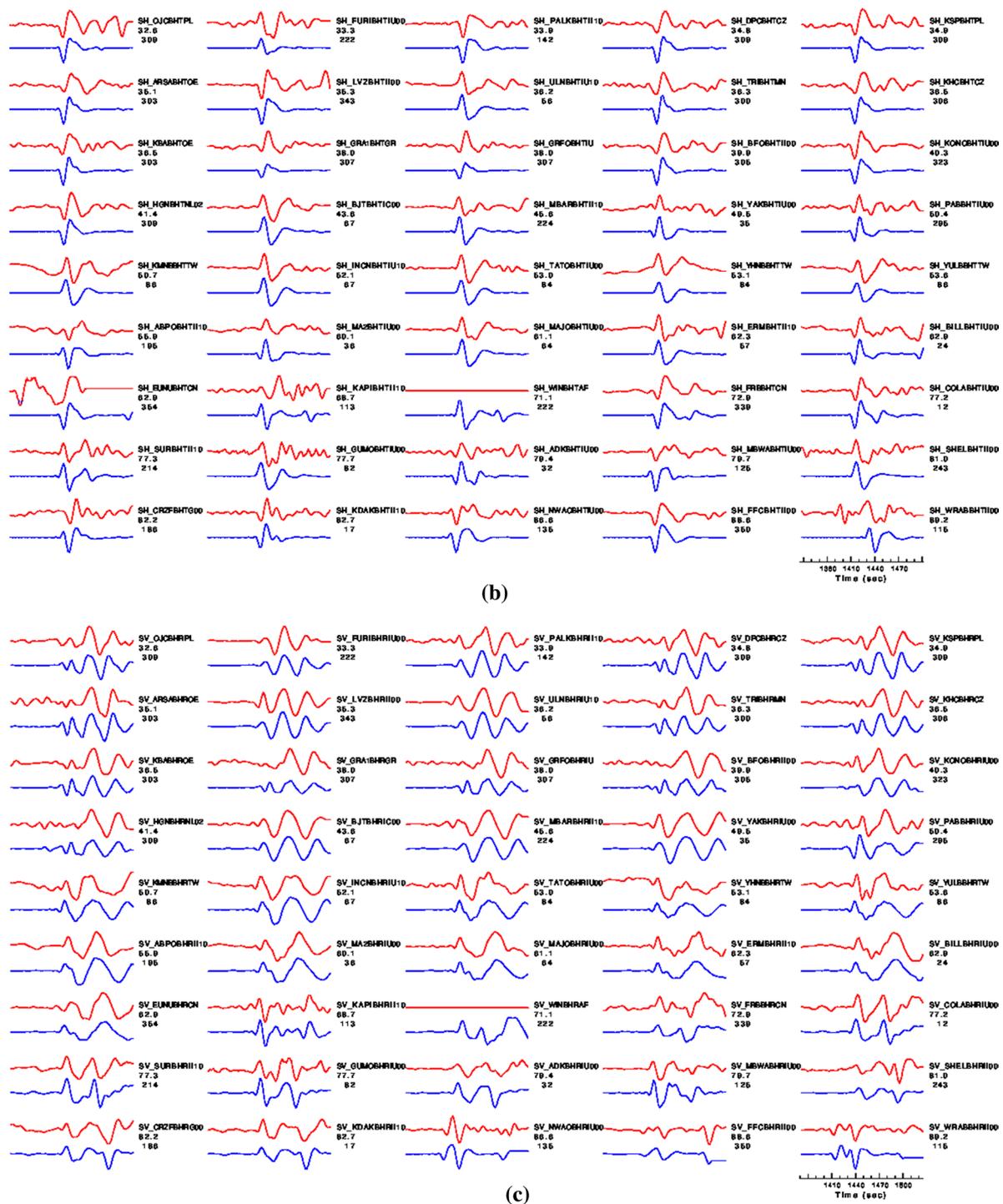


Fig. 4 (continued)

by reverse components (Fig. 1a). The focal mechanism for this earthquake is a combination of compression, with some right-lateral strike-slip motion that can be taken as an indication of a transpressional regime in this area. According to the teleseismic source parameters derived in this study (Fig. 1b, Table 2), the best-fitted nodal planes, deviatoric

moment tensors based on linear inversion, and the grid-searched moment tensors all show reverse components. Various factors including different selected stations so different azimuthal coverages, different frequency bands, the use of ground displacement versus ground velocity and different velocity models to compute Green's functions may

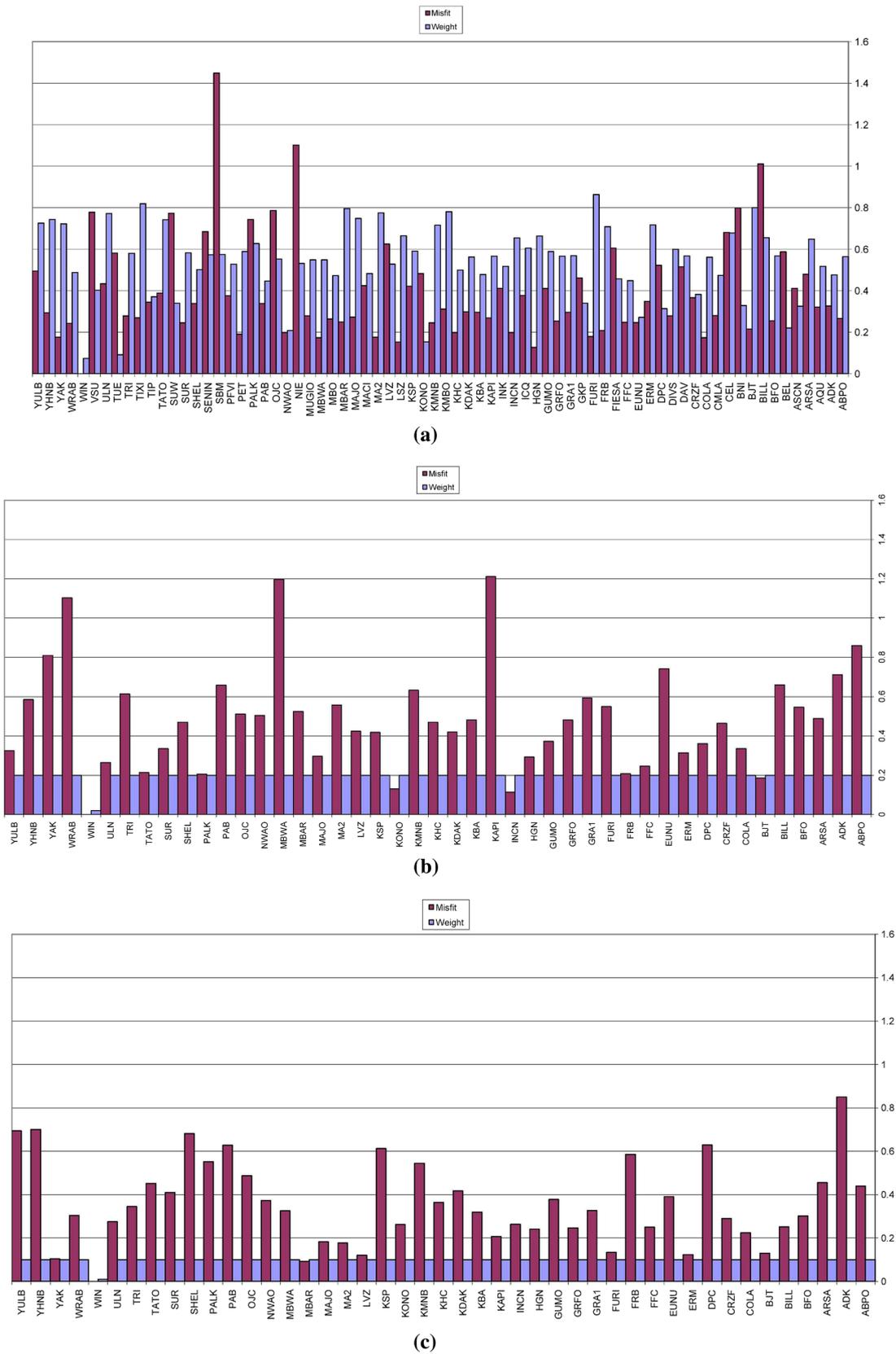


Fig. 5 Misfits and weights for each station components **a** P, **b** SV, and **c** SH as applied in this analysis

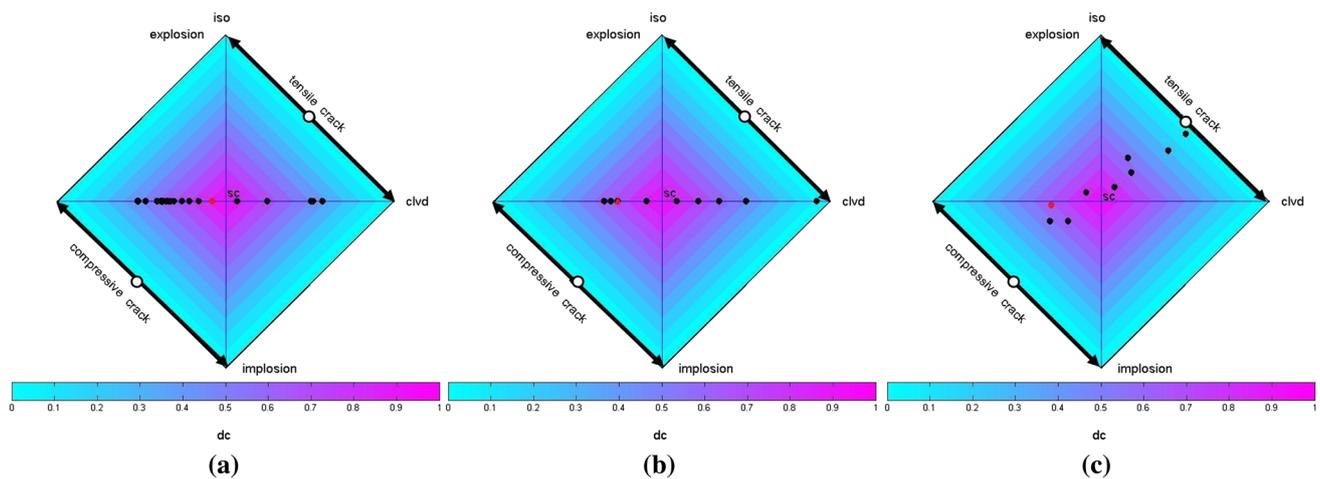


Fig. 6 Points plotted for different moment tensor results of this study based on diamond source-type plot. **a** Linear DEV-MT inversion, **b** DEV-MT, **c** full-MT. The moment tensor of each plotted point is

be considered to be responsible for the differences in the results.

As mentioned earlier on the diamond source-type plot (that is based on hexagonal bipyramid projection), elements of the mechanism (DC, \pm CLVD, and \pm ISO) are displayed by points. We draw our calculated moment tensors based on the linear inversion and grid search (DEV-MT and full-MT) on this CLVD-ISO source-type plot. The linear deviatoric moment tensor inversion (Fig. 6a), the best grid-searched deviatoric MT (Fig. 6b), and the best full grid-searched MT (Fig. 6c) all plotted the moment tensor near the CLVD axis in the third quadrant. The plotted points on Fig. 6a–c are based on the moment tensors computed by applying the linear deviatoric inversion, the grid-searched deviatoric MT, and the full grid-searched MT correspondingly (for more information, please look at the electronic supplement).

Non-DC components of the point source mechanisms are considered to be indications of the unmodeled complexity of the source. Several phenomena like shear faulting in a heterogeneous or anisotropic medium and multiple shear faulting or shear faulting on a non-planar fault are among the reasons for obtaining considerable CLVD component from moment tensor inversions (Vavrycuk 2002). Source-type plot analysis (Fig. 6a–c) shows that the plotted moment tensors are along the CLVD axis. For tectonic sources where slip occurs entirely at the fault surface, the components are reducible to five (Ortega et al. 2014). The double-couple assumption of a slip direction purely parallel to the fault plane may be approximative. The common tectonic models proposed to explain the earthquake moment tensors may imply rheological constraints that are not always realistic. Dufumier and Riverra (1997) suggested a physical model that includes non-tectonic and tectonic volumetric changes.

computed by linear deviatoric inversion, grid-searched deviatoric MT, and full grid-searched MT correspondingly. The details of calculations are presented on the electronic supplement

They explained that the trace of the observed moment tensor as the term of a unique volumetric change at the source is misleading so that the decomposition in non-tectonic and tectonic isotropic components is more reliable and informative.

Acknowledgments I would like to thank Sean Ford and anonymous reviewers for their constructive comments and valuable suggestions. I acknowledge Ramón Zúñiga (co-editor-in-chief), Zbigniew Wiśniewski (managing editor), and Eleftheria Papadimitriou (chief editor) for their helps. I acknowledge IRSC, IRIS, ISC, GCMT, and USGS for providing their data for analysis. Computer programs in seismology (Herrmann 2013), moment decomposition code (Vavrycuk 2015a), and GMT (Wessel and Smith 1995) were used for computations and drawing some figures.

References

- Ashtari Jafari M (2008) Lessons learned from the Bam urban earthquake. In: 14th world conference on earthquake engineering. Beijing, China, <https://doi.org/10.13140/rg.2.1.1892.9125>
- Ashtari Jafari M (2013) Spatial distribution of seismicity parameters in the Persian Plateau. *Earth Planets Space* 65:863–869. <https://doi.org/10.5047/eps.2013.02.006>
- Ashtari Jafari M (2016) Lessons learned from the recent earthquakes in Iran. In: D’Amico S (ed) *Earthquakes and their impact on society*. Springer International Publishing. https://doi.org/10.1007/978-3-319-21753-6_18
- Aso N, Ohta K, Ide S (2016) Mathematical review on source-type diagrams. *Earth Planets Space* 68:52. <https://doi.org/10.1186/s40623-016-0421-5>
- Berberian M, Ghorashi M, Shoja-Taheri J, Talebian M (2000) Seismotectonic and earthquake-fault hazard investigations in the Mashhad-Neyshabour region. Report No. 72, Geological Survey of Iran
- D’Amico S, Cammarata L, Cangemi M, Cavallaro D, Di Martino R, Firetto Carlino M (2014) Seismic moment tensors and regional

- stress in the area of the December 2013–January 2014, Matese earthquake sequence (Central Italy). *J Geodyn* 82:118–124
- Dufumier H, Riverra L (1997) On the resolution of the isotropic component in moment tensor inversion. *Geophys J Int* 131:595–606. <https://doi.org/10.1111/j.1365-246X.1997.tb06601.x>
- Dziewonski AM, Ekström G, Salganik MP (1992) Centroid-moment tensor solutions for July–September 1991. *Phys Earth Planet Inter* 72:1–11
- Ekström G, Nettles M, Dziewonski AM (2012) The global CMT project 2004–2010: centroid-moment tensors for 13017 earthquakes. *Phys Earth Planet Inter* 200:1–9
- Foulger GR, Julian BR (2015) Non-double-couple earthquakes. *Encycl Earthq Eng*. https://doi.org/10.1007/978-3-642-36197-5_290-1
- Herrmann RB (2013) Computer programs in seismology: an evolving tool for instruction and research. *Seismol Res Lett* 84:1081–1088
- Herrmann RB, Benz H, Ammon CJ (2011) Monitoring the earthquake source process in north America. *Bull Seismol Soc Am* 101:2609–2625
- Hjörleifsdóttir V, Ekström G (2010) Effects of three-dimensional Earth structure on CMT earthquake parameters. *Phys Earth Planet Inter* 179:178–190
- Jost ML, Herrmann RB (1989) A student's guide to and review of moment tensors. *Seismol Res Lett* 60:37–57
- Julian BR, Miller AD, Foulger GR (1998) Non-double-couple earthquakes 1. Theory *Rev Geophys* 36:525–549
- Kagan YY (1991) 3-D rotation of double-couple earthquake sources. *Geophys J Int* 106:709–716
- Kanamori H, Rivera L (2008) Source inversion of W phase: speeding up seismic tsunami warning. *Geophys J Int* 175:222–238
- Konstantinou KI, Rontogianni S (2011) A comparison of teleseismic and regional seismic moment estimates in the European-mediterranean region. *Seismol Res Lett* 82:188–200
- Miller AD, Foulger GR, Julian BR (1998) Non-double-couple earthquakes 2. Observations. *Rev Geophys* 36:551–568
- Montagner JP, Kennett BLN (1996) How to reconcile body-wave and normal-mode reference earth models. *Geophys J Int* 125:229–248. <https://doi.org/10.1111/j.1365-246X.1996.tb06548.x>
- Ortega R, Quintanar L, Rivera L (2014) Full moment tensor variations and isotropic characteristics of earthquakes in the Gulf of California transform fault system. *Pure Appl Geophys* 171:2805–2817. <https://doi.org/10.1007/s00024-013-0758-7>
- Shabanian E, Bellier O, Siame L, Arnaud N, Abbassi MR, Cocheme JJ (2009) New tectonic configuration in NE Iran: active strike-slip faulting between the Kopeh Dagh and Binalud mountains. *Tectonics*, 28, TC5002, <https://doi.org/10.1029/2008tc002444>
- Vavrycuk V (2002) Non-double-couple earthquakes of 1997 January in West Bohemia, Czech Republic: evidence of tensile faulting. *Geophys J Int* 149:364–373
- Vavrycuk V (2005) Focal mechanisms in anisotropic media. *Geophys J Int* 161:334–346. <https://doi.org/10.1111/j.1365-246X.2005.02585.x>
- Vavrycuk V (2015a) Moment tensor decompositions revisited. *J Seismol* 19:231–252
- Vavrycuk V (2015b) Moment tensors: decomposition and visualization. In: Beer M, Kougioumtzoglou I, Patelli E, Au SK (eds) *Encyclopedia of earthquake engineering*. Springer, Berlin. https://doi.org/10.1007/978-3-642-36197-5_288-1
- Wessel P, Smith WHF (1995) New version of the generic mapping tools released. *EOS* 76:329
- Zhu L, Ben-Zion Y (2013) Parameterization of general seismic potency and moment tensors for source inversion of seismic waveform data. *Geophys J Int* 194:839–843



Cyclic loading–unloading creep behavior of composite layered specimens

Bo Hu¹ · Sheng-Qi Yang¹ · Peng Xu² · Jian-Long Cheng³

Received: 10 June 2018 / Accepted: 2 February 2019 / Published online: 9 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

Soft and hard interbedded rocks show obvious time-dependent deformation after deep tunnel excavations, and it is therefore necessary to research the mechanical behavior of the layered rock. However, it is hard to obtain ideal transversely isotropic rocks in fields, so rock-like specimens were poured by using artificial materials. Cyclic loading–unloading creep experiments were performed on the artificial layered cemented specimens with various layer angles (0°, 30°, 60° and 90°) at a 20 MPa confinement. Time-independent deformations and time-dependent deformations of the rock-like specimens were distinguished to investigate the visco-elasto-plastic deformation characteristics. Instantaneous elastic strain and instantaneous plastic strain had linear correlations with stress ratio, whereas creep strain, including visco-elastic strain and visco-plastic strain, increased nonlinearly with an increasing stress ratio. The specimens with a small layer angle had more noticeable time-independent and time-dependent deformations and larger steady-state creep rates than those of the specimens with a large layer angle. Attenuation creep and secondary creep could be observed at relative low stress levels, whereas accelerating creep until failure occurred at the creep failure stress level. The time for creep failure can be predicated according to the axial steady-state creep rate or volumetric creep curve. Damage in the rock-like specimens showed linear correlation with the stress ratio. Dip angle has a significant effect on the creep failure mode under cyclic loading–unloading conditions.

Keywords Rock-like material · Anisotropy · Creep tests · Visco-elasto-plastic deformation · Creep failure

Introduction

In recent years, to meet the requirements of economic development, an increasing number of deep engineering projects have been planned and constructed. At the same time, the strength and deformational behaviors of surrounding rocks become more complicated (Ramamurthy 1993; Malan and Spottiswoode 1997; Vogel and Andrast 2000; He et al. 2005;

Cho et al. 2012; Liu et al. 2013; Xie et al. 2015; Yang et al. 2017, 2018). Anisotropy has been a long-standing issue in rock engineering (Amadei et al. 1983; Amadei 1996; Barla 1974), particularly when actual excavation engineering frequently encounters weak hard uneven strata, showing distinct anisotropic characteristics and causing great engineering difficulty (e.g., tunnel boring machine (TBM) entrapment) (Sanio 1985; Babendererde et al. 2004; Gong et al. 2006; Shang et al. 2007; Stille and Palmström 2008; Barla 2016). The anisotropic characteristics of most sedimentary rock and metamorphic rock (e.g., mudstone, sandstone, shale, slate, phyllite, schist and gneiss) are distinct, and some of them are generally considered transversely isotropic characteristics (Colak and Unlu 2004). Moreover, Tien et al. (2006) reported the idea of tunnel excavation in a foliated rock mass, which can be approximately regarded as transversely anisotropic rock. For a better understanding of the influence of the anisotropy of a layered rock mass on the deformation and failure behaviors of composite rocks, numerous experimental tests have been conducted. Donath (1964) did a groundbreaking investigation on the strength

✉ Sheng-Qi Yang
yangsqi@hotmail.com

¹ State Key Laboratory for Geomechanics and Deep Underground Engineering, School of Mechanics and Civil Engineering, China University of Mining and Technology, Xuzhou 221116, People's Republic of China
² China Construction Third Engineering Design Bureau Co., Ltd, Wuhan 430064, Hubei, People's Republic of China
³ State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, Sichuan, People's Republic of China

and deformational and failure behaviors of stratified rocks (e.g., the Martinsburg slate) under complex stress conditions. The test results showed that the specimens with dip angles of 0° and 60° (direction perpendicular to the discontinuities) have the highest and the lowest strengths, respectively. Nasser et al. (1997) investigated the strength and failure mechanism of four intact schistose rocks and analyzed the anisotropic behaviors and failure patterns of the specimens in macro- and microscales at varying orientations. Niandou et al. (1997) investigated the mechanical behaviors of Tournemire shale under hydrostatic and triaxial compressive conditions. The results showed that confining pressure and loading orientation greatly affected the failure behavior of the specimen. To calculate the elastic parameters for transverse isotropic rocks, Nunes (2002) developed a new method to determine the orientation and deformation parameters of transversely isotropic cylinders of rock under isotropic biaxial loading. Gonzaga et al. (2008) presented a new methodology to determine the elastic parameters for transverse isotropic rocks from a single standard cylindrical specimen. Cho et al. (2012) investigated the deformation and strength properties of anisotropy rocks and presented the method to determine the independent elastic constants. Gong et al. (2015) developed a new algorithm to improve understanding about nonlinear behavior of a simulated roadway tunnel in horizontally stratified rock masses.

However, ideal composite rock specimens (with uniform thickness and various specified dip angles) are extremely difficult to obtain from the field. Artificial materials are widely used to establish rock-like specimens with a wide range of mechanical properties and different specified dip angles in laboratory testing. Tien and Tsao (2000) used dried powder materials (e.g., cement, sand, microsilica and kaolinite) and water in various ratios to prepare stratified rock and drilled composite rock specimens with various dip angles. Then, Tien and Kuo (2001) presented a new failure criterion for transversely isotropic rocks based on sliding mode and non-sliding mode. Tien et al. (2006) reported the failure mechanism of artificial transversely isotropic rock specimens with different dip angles during uniaxial compressive experiments. Moreover, the failure modes of the composite specimens under triaxial compressive conditions were also classified systematically. For large-scale model experiments, He et al. (2009) used gypsum and water in various ratios as similar materials to investigate the mechanism of roadway instability in deep mines. He et al. (2010) simulated roadway excavation in geologically horizontal strata at great depth using a physical modeling experiment. He (2011) also used a physical modeling approach and infrared thermography to study roadway excavation in geologically 45° inclined rock strata. Numerical test results showed that the weak plane (transverse isotropic plane) significantly affects the deformation, strength and permeability of anisotropy rocks

(McLamore and Gray 1967; Barla 1974; Nova 1980; Colak and Unlu 2004; Naumann et al. 2007; Gonzaga et al. 2008; Khanlari et al. 2015). Moreover, the creep deformation of composite rock strata cannot be ignored for long-term stability in tunnel engineering. Han et al. (2007) studied the effect of stress and loading path on the rheological behaviors of interbedded strata of soft and hard rock under uniaxial and triaxial tests and simulated experimental results using FLAC^{3D} (ITASCA Consulting China Ltd., Wuhan, China). Wang et al. (2010) studied the mechanics interaction and creep deformation behavior of bedded salt rock under uniaxial and low confinement conditions.

However, in the past, fewer researches are rarely related to the time-dependent deformation of composite rocks, which is very important for predicting the long-term stability and safety of rock engineering (Yang and Hu 2018). Moreover, the visco-elasto-plastic deformation behaviors considering the timescale of the composite rocks with different dip angles have rarely been reported. Therefore, to better understand the effects of dip angle on the time-independent and time-dependent deformation and failure behaviors of the composite strata rock, triaxial cyclic loading–unloading creep experiments were conducted on artificial composite rock specimens.

Experiments

Preparation of rock-like specimens

The rock-like specimens (also be termed as artificial transversely isotropic rocks or composite rocks) are made up of ‘hard’ rock and ‘weak’ rock with two different strengths and stiffnesses. Following the research by Cheng et al. (2017), the composite rock specimens were composed of ordinary Portland cement, gypsum, kaolinite powder and fine quartz sand, proportions for which are listed in Table 1.

As Cheng et al. (2017) reported, the rock-like specimen blocks were formed in a rectangular stiff plastic mold with 150 mm width, 150 mm height and 550 mm length. The thickness of the two kinds of rock was designed to be 10 mm. Furthermore, 80 g black ink was put into the ‘weak’ rock to distinguish the two rocks. The mixture slurry of ‘hard’ rock with designed quality was poured into the mold and vibrated 2 min before being allowed to set 10 min to achieve its initial strength. Subsequently, the mixture slurry of ‘weak’ rock with designed quality was poured on the initially hardened ‘hard’ rock and vibrated for same amount of time; then, the specimens were allowed to set for 50 min to attain the initial specified strength. The next layers of mixture slurry were laid repeating the above procedures. The cast-in-place composite rock block was hardened for 4 h and then placed in water with the mold at room temperature

Table 1 Two material ratios for the preparation of the rock-like specimen (Cheng et al. 2017)

Material	Portland cement	Water	Gypsum	Kaolinite	Quartz sand
<i>'Weak' rock</i>					
Ratio	1	0.75	0.15	0.3	1.2
Material	Sulfate aluminum cement	Water	Gypsum	Quartz sand	Iron powder
<i>'Hard' rock</i>					
Ratio	1	0.355	0.9	0.9	0.1

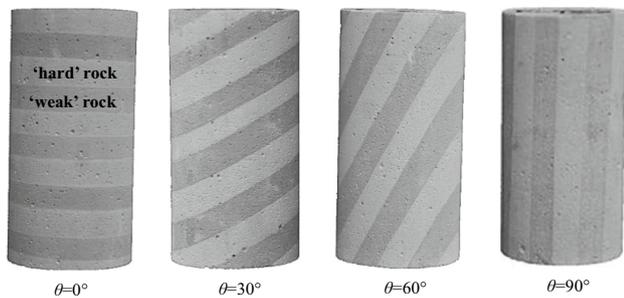


Fig. 1 Formed specimens of artificial transversely isotropic rocks

for 7 days. It is worth noting that the pouring time for the materials in the preparation stage was strictly controlled to decrease the discrete error.

Basic mechanics tests were performed after 28 days of curing at room temperature. The molded test piece was drilled with a 50-mm-diameter drill perpendicular to the horizontal or vertical bedding plane to obtain specimens with an inclination angle of $\theta = 0, 30, 60$ and 90° . Then, the specimens were cut and polished to form cylinders with diameters of 50 mm and heights of 100 mm. The tested specimens are shown in Fig. 1. The relevant physical and mechanical parameters of the two materials are given in Table 2.

Testing procedure

Cyclic loading and unloading creep experiments were conducted on servo-controlled rock triaxial testing equipment. To obtain the elastic strain, viscosity strain and plastic strain of the composite rock, multi-step loading and unloading cycles were used. First, the rock-like specimens

sealed within a nitrile rubber sleeve were subjected to a hydrostatic pressure of 20 MPa at rate of 5 MPa/min. Then, deviatoric stresses were applied to the configured levels by stress-controlled mode at a rate of 7.37 MPa/min for 2 days. Subsequently, deviatoric stresses were removed. After 1 day, the next deviatoric stress was applied until failure of the rock-like specimens. The loading and unloading procedure is shown in Fig. 2. Tables 3, 4 list the basic experimental information of the two types of rock material.

During creep experiments, the recoverable strains and irrecoverable strains can be separated by cyclic loading and unloading methods. Recoverable strain is composed of instantaneous elastic strain (ϵ_{me}) and visco-elastic strain (ϵ_{ce}), whereas irrecoverable strain is composed of instantaneous plastic strain (ϵ_{mp}) and visco-plastic strain (ϵ_{cp}). After one loading–unloading cycle, the stress–strain curve is as shown in Fig. 3a (segment OABDE), and the strain–time curve is

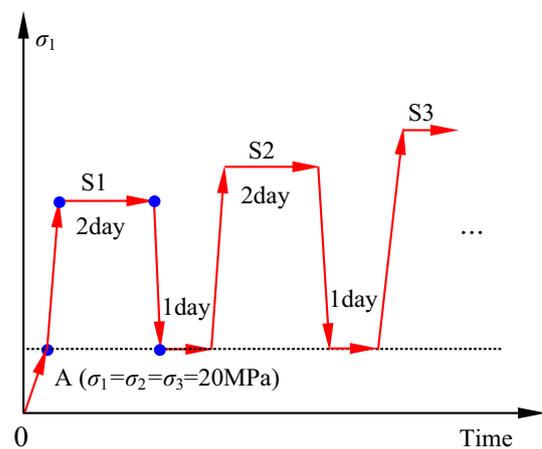


Fig. 2 Loading and unloading procedure for tests

Table 2 Physical and mechanical parameters of the two materials (Cheng et al. 2017)

	UCS (MPa)	E (GPa)	ν	C (MPa)	φ ($^\circ$)	ρ (kg m^{-3})
'Weak' rock	16.44	3.60	0.23	5.21	24	1627
'Hard' rock	68.43	23.69	0.34	18.17	40	1912

UCS uniaxial compressive strength, E Young's modulus, ν Poisson's ratio, C cohesion, φ , friction angle, ρ density

Table 3 Basic physical parameters of the tested composite rock specimens in this research

No.	Dip (°)	<i>L</i> (mm)	<i>D</i> (mm)	<i>M</i> (g)	ρ (kg m ⁻³)	v_p (m s ⁻¹)	v_s (m s ⁻¹)
F-0-2	0	100.85	50.03	391.97	1.98	2857	1923
F-30-2	30	100.85	50.02	382.11	1.93	2837	1815
F-60-2	60	99.41	49.89	386.75	1.99	2950	1851
F-90-2	90	99.09	50.01	388.61	2.00	3030	2222

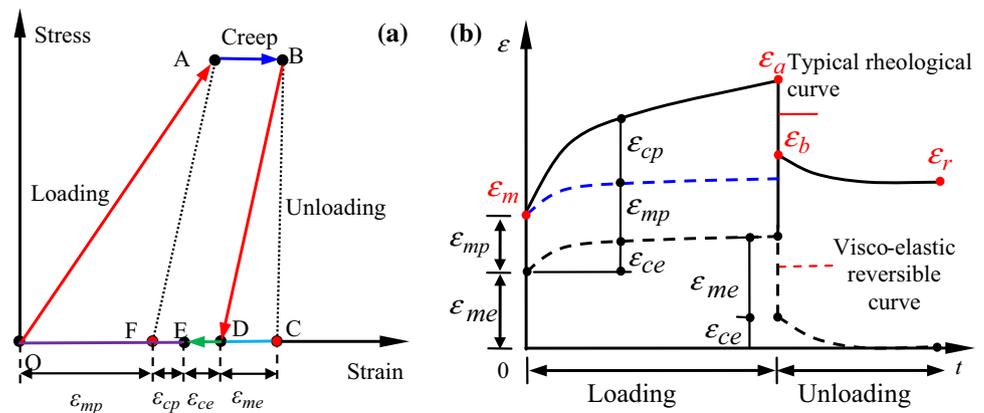
L length, *D* diameter, *M* mass, ρ density, v_p *P*-wave velocity, v_s *S*-wave velocity

Table 4 Cyclic loading–unloading creep experiment conditions

No.	Dip (°)	σ_3 (MPa)	σ_p (MPa)	SD	Creep stress levels (MPa)					
					S1	S2	S3	S4	S5	S6
F-0-2	0	20	115.74	6.95	73.46	85.74	97.91	110.05	122.16	134.43
F-30-2	30	20	101.84	3.65	63.48	74.12	84.62	95.26	–	–
F-60-2	60	20	70.50	9.08	40.85	47.63	54.34	61.14	68.03	73.24
F-90-2	90	20	125.35	2.36	73.86	86.05	98.38	108.81	122.93	135.32

σ_p peak strength (peak deviatoric stress), *SD* standard deviation

Fig. 3 A typical deformation curve of rock under 1 cycle loading and unloading. **a** Stress–strain curve; **b** strain–time curve (Xu and Yang 2015)



plotted in Fig. 3b. In Fig. 3a, segments OA and BD represent loading and unloading curves, respectively, and segment AB represents time-dependent deformation under the creep stress condition. When creep stress is removed (0 MPa), the axial strain curve first decreases from *B* to *D* and then recovers to *E* with time. Therefore, instantaneous elastic strain ϵ_{me} (segment CD), visco-elastic strain ϵ_{ce} (segment DE), visco-plastic strain ϵ_{cp} (segment EF) and instantaneous plastic strain ϵ_{mp} (segment OF) can be determined. Specifically, total strain (ϵ) can be divided into time-independent strain (ϵ_m) and time-dependent strain (ϵ_c) (Xia and Zhong 1989; Zhao et al. 2017), which can be expressed as Eqs. (1) and (2)

$$\epsilon_m = \epsilon_{me} + \epsilon_{mp} \tag{1}$$

$$\epsilon_c = \epsilon_{ce} + \epsilon_{cp} \tag{2}$$

Results and discussion

Rheological deformation behavior

Multi-step cyclic loading and unloading tests were performed on the composite rocks with four dip angles (0°, 30°, 60° and 90°). The axial stress–strain curves and strain–time curves of the rock-like specimens are plotted in Figs. 4 and 5, respectively. It is clear that the axial strains of the 0° and 30° dip angle composite rock specimens were larger than those of the 60° and 90° specimens. When the first stress level was applied, the specimens with 0° and 30° dip angle showed longer yield phase, whereas the specimens with 60°–90° dip angle showed less distinct yield phase. The irreversible deformations also showed similar variations, which were increasing with increasing loading–unloading cycles. Axial deformation increased with loading stress linearly besides in the first loading stage. The rock-like specimens deformed gradually over

Fig. 4 Axial stress–strain curves of rock-like specimens under cyclic loading–unloading path

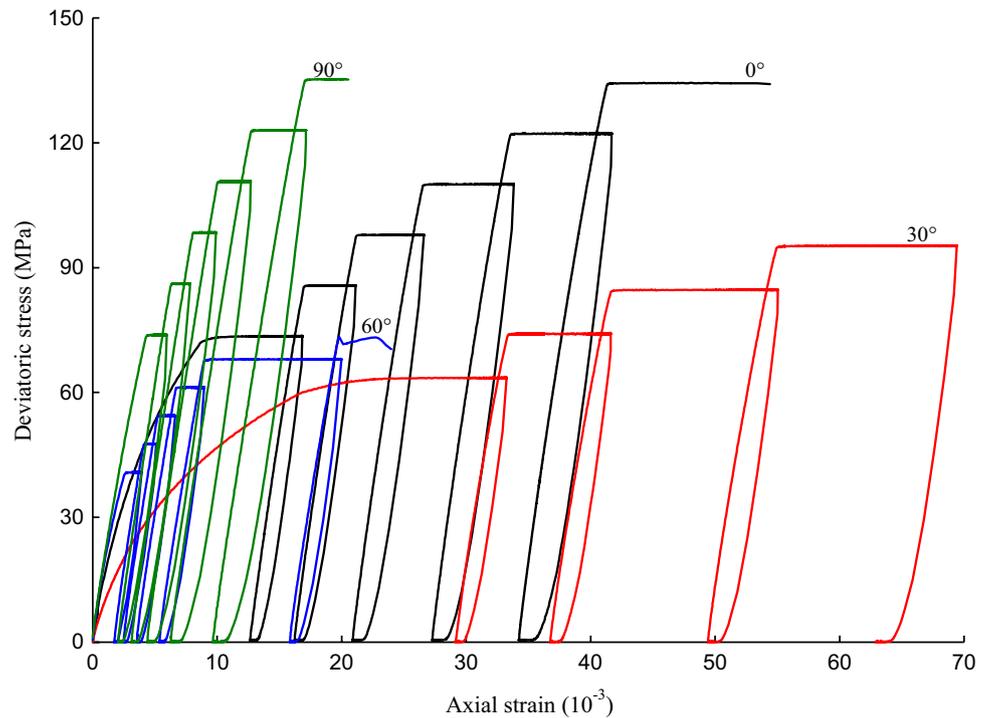
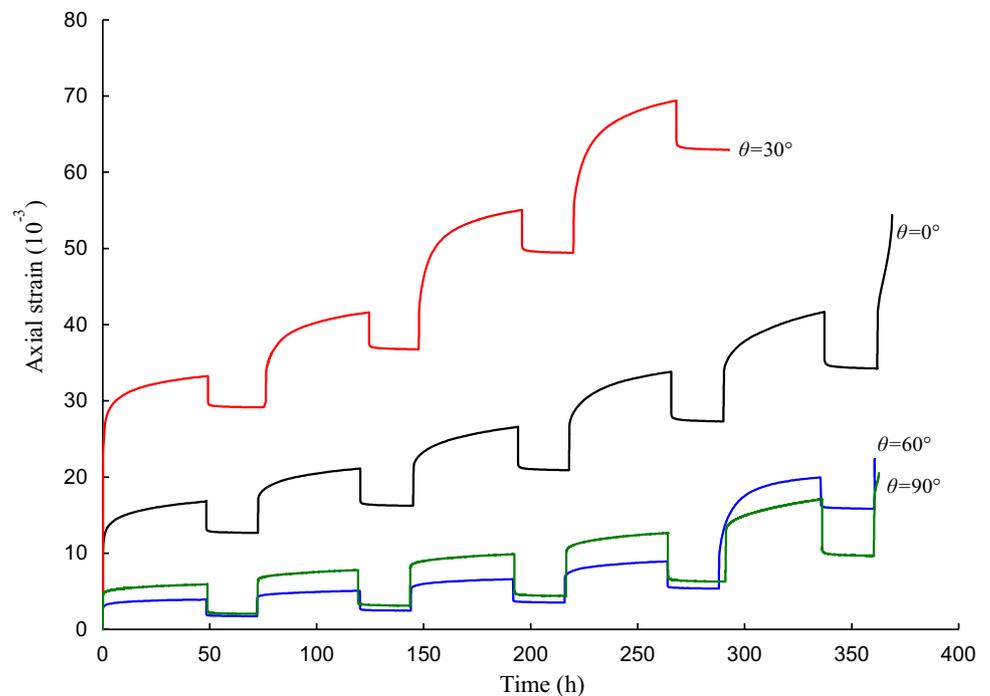


Fig. 5 Axial strain–time curves of rock-like specimens under cyclic loading–unloading path



time for approximately 2 days after the deviatoric stress was applied to the configure levels. When the deviatoric stresses were removed, axial deformations gradually decreased to some values (residual strain) during the unloading stage. It is worth noting that the specimen with a 30° dip angle did not incur creep failure because the

axial deformation was large and might be beyond the sensor range; therefore, the final loading–unloading cycle was unable to be performed.

The axial strain–time curves of the composite rock specimens at a confining pressure of 20 MPa under cyclic loading–unloading condition in relation to dip angle are

plotted in Fig. 5. It can be seen that the axial strain of the 30° dip angle specimen was nearly two times that of the 0° dip angle specimen and approximately 4–6 times that of the 60° and 90° dip angle composite rock specimens. Under loading conditions, axial strain curves showed not only noticeable primary but also a steady-state creep stage, as well as tertiary creep at the final stress levels. Accelerating creep failure over short times at the onset of accelerated creep stage occurred rapidly. When the constant stresses were removed, the deformations of the rock-like specimens recovered over time, and finally, residual strains were unable to be reversed. Moreover, the 0° and 30° dip angle specimens yielded a longer primary creep stage than those of the two high-dip-angle specimens. Therefore, the cyclic

loading and unloading creep tests illustrated the influence of loading history on deformations. Commonly, the axial strain under loading conditions increased with increasing deviatoric stress, and the residual strain also increased with increased loading history.

The curves of the axial, lateral and volumetric strains of the transversely isotropic synthetic rocks in relation to deviatoric stress and time are shown in Fig. 6. In Fig. 6, the axial strains increased with increasing deviatoric stress and time at the loading stage. The irrecoverable strains increased with the number of loading–unloading cycles. The variations in axial creep strain and volumetric creep strain were pronounced, whereas the lateral creep strains varied indistinctively. Specifically, for the 0° dip

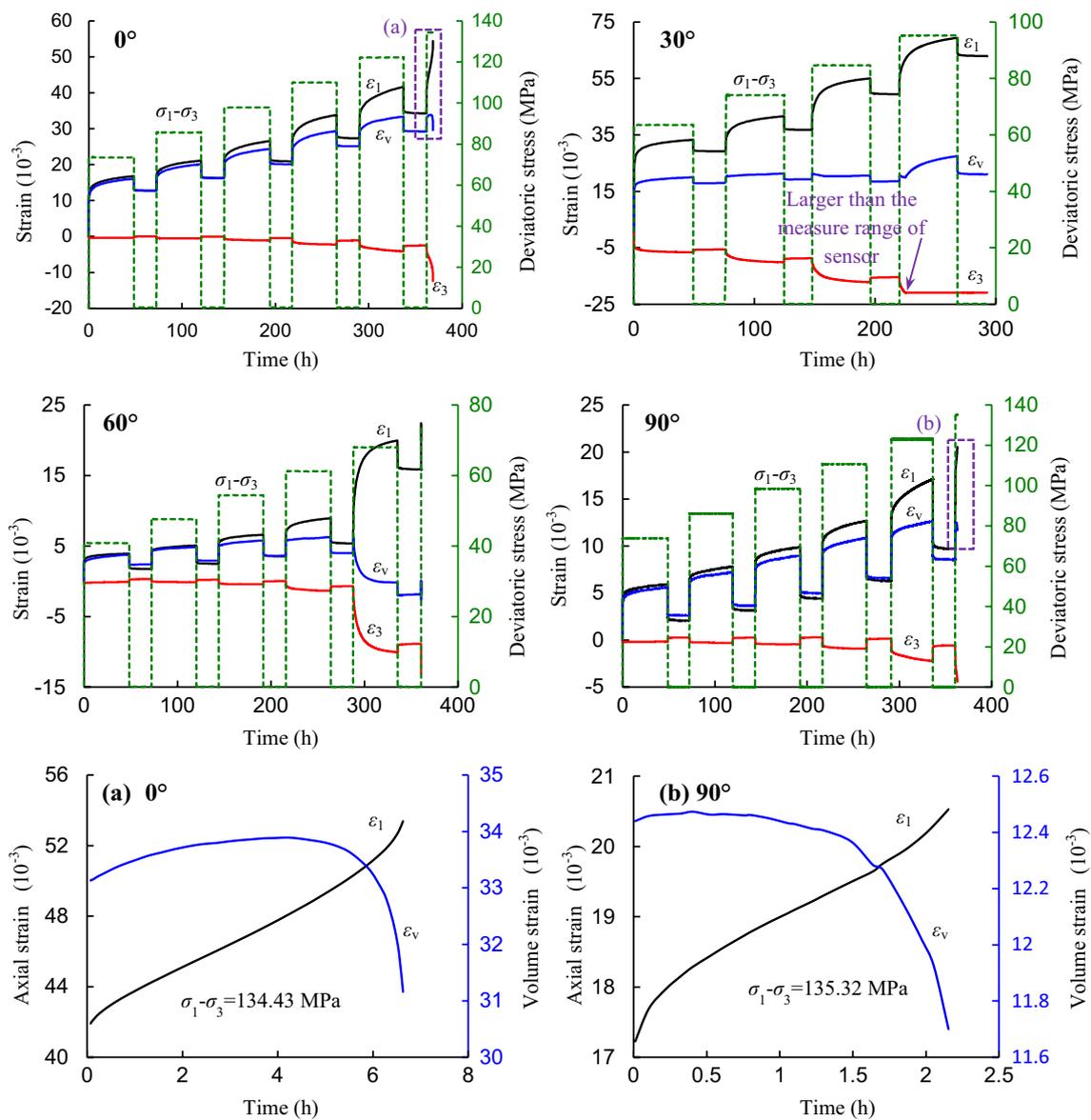


Fig. 6 Strain-time curves of rock-like specimens under cyclic loading–unloading

angle specimen, the axial and volumetric strains increased with increasing deviatoric stress and time under stress levels below creep failure stress, whereas the volumetric strain transformed from increasing to decreasing when the stress level reached 134.43 MPa. This means that the specimen transitioned from having a compaction-dominated deformation to a dilatancy-dominated deformation. For the 30° dip angle specimen, the lateral strain increased with increasing deviatoric stress and time, which was larger than the effective range of the sensor under a stress level of 95.26 MPa. Hence, the lateral strain and volumetric strain values were invalid. Moreover, the axial deformation was so large under this stress level causing it to possibly exceed the range of the sensor, so the last loading–unloading cycle was not completed. For the 60° dip angle specimen, the axial and lateral strains had a sudden rise under the failure stress level. The volumetric strain began to decrease. After unloading for 1 day and then continued application of a stress level of 73.24 MPa, tertiary creep occurred in a short time. For the 90° dip angle specimen, variations in the axial strain, lateral strain and volumetric strain were similar to the results of the 0° dip angle rock-like specimen. Three-stage creep occurred in a short time, and volumetric strain decreased gradually over time at 135.32 MPa, as shown in Fig. 6a, b. The reason the creep failure stresses were larger than the peak stresses will be discussed in Discussion section of this paper.

Instantaneous elasto-plastic deformations

Under compressive creep loading, rock specimens yield reversible deformation and irreversible deformation. When the loading is removed, a partial of deformation will recover with time, but another part of deformation is unrecoverable. By using a loading–unloading method, the elastic deformation, viscosity deformation and plastic deformation of the rock specimen can be divided. This has important practical value for reasonably predicting the rheological deformation and rupture stress of transversely isotropic rocks. The strain separation results are listed in Table 5.

Figure 7 illustrates the relationships between the axial instantaneous strain, instantaneous elastic strain and instantaneous plastic and stress ratio of the specimens with dip angle ranging from 0° to 90°. It is clear that axial deformation had a linear correlation with the stress ratio. In Fig. 7, the dotted straight lines represent the fitting results of the experimental data. The fitting equation can be expressed as Eq. 3, and the fitting parameters are listed in Table 6. The slopes of the fitting lines in relation to dip angle are plotted in Fig. 7d. The slope of the fitting line of instantaneous elastic strain first increased as dip angle increased from 0° to 60° and then decreased as the dip angle increased to 90°. However, the slopes of the fitting lines of instantaneous strain and instantaneous plastic strain first decreased as the dip angle increased from 0° to 30° and then increased with increasing dip angle

$$(\sigma_1 - \sigma_3)/\sigma_p = A \cdot \varepsilon + B. \quad (3)$$

Table 5 Axial visco-elasto-plastic strain separation results for the rock-like specimens

No.	$\sigma_1 - \sigma_3$ (MPa)	$\varepsilon_m (10^{-3})$	$\varepsilon_c (10^{-3})$	$\varepsilon_e (10^{-3})$	$\varepsilon_p (10^{-3})$	$\varepsilon_{me} (10^{-3})$	$\varepsilon_{mp} (10^{-3})$	$\varepsilon_{ce} (10^{-3})$	$\varepsilon_{cp} (10^{-3})$
F-0-2	73.46	11.44	5.37	4.18	12.63	3.65	7.79	0.53	4.84
	85.74	17.27	3.78	4.88	16.17	4.21	13.06	0.68	3.10
	97.91	21.40	5.25	5.70	20.96	4.88	16.52	0.82	4.43
	110.05	26.63	7.28	6.49	27.43	5.58	21.06	0.91	6.37
	122.16	33.42	8.19	7.36	34.25	6.45	26.97	0.91	7.28
F-30-2	63.48	24.24	8.94	4.08	29.10	3.39	20.85	0.69	8.25
	74.12	33.30	8.36	4.95	36.71	4.06	29.24	0.90	7.46
	84.62	42.29	12.62	5.31	49.59	4.54	37.75	0.77	11.85
	95.26	55.10	14.44	6.48	63.06	5.16	49.94	1.33	13.11
F-60-2	40.85	2.98	0.95	2.18	1.75	1.98	1.00	0.21	0.75
	47.63	4.21	0.90	2.62	2.50	2.40	1.82	0.22	0.68
	54.34	5.33	1.28	3.07	3.54	2.75	2.58	0.32	0.96
	61.14	6.87	2.09	3.57	5.39	3.20	3.66	0.36	1.73
	68.03	9.53	10.33	4.07	15.80	3.50	6.03	0.57	9.76
F-90-2	73.86	4.58	1.36	3.55	1.03	0.34	1.02	3.89	2.05
	86.05	6.41	1.39	4.20	2.21	0.47	0.92	4.67	3.12
	98.38	8.12	1.77	4.90	3.22	0.55	1.22	5.46	4.44
	108.81	10.08	2.58	5.67	4.41	0.70	1.88	6.37	6.29
	122.93	12.54	4.55	6.55	5.99	0.85	3.70	7.40	9.69

Fig. 7 Relationships between stress ratio and time-independent deformations of rock-like specimens

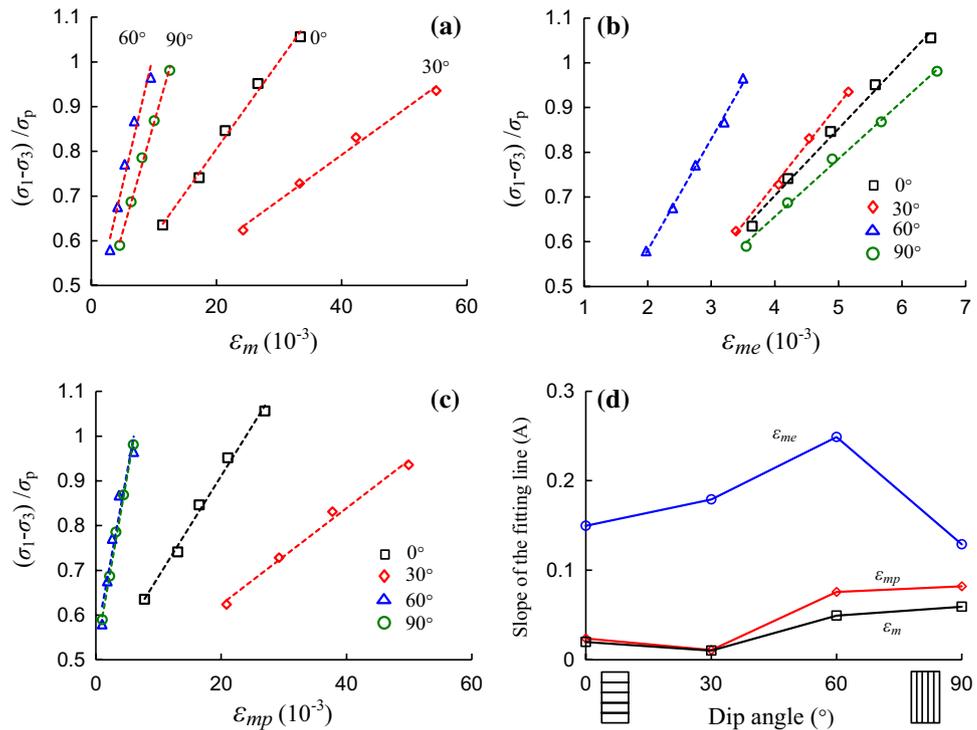


Table 6 Fitting results of Eq. 3

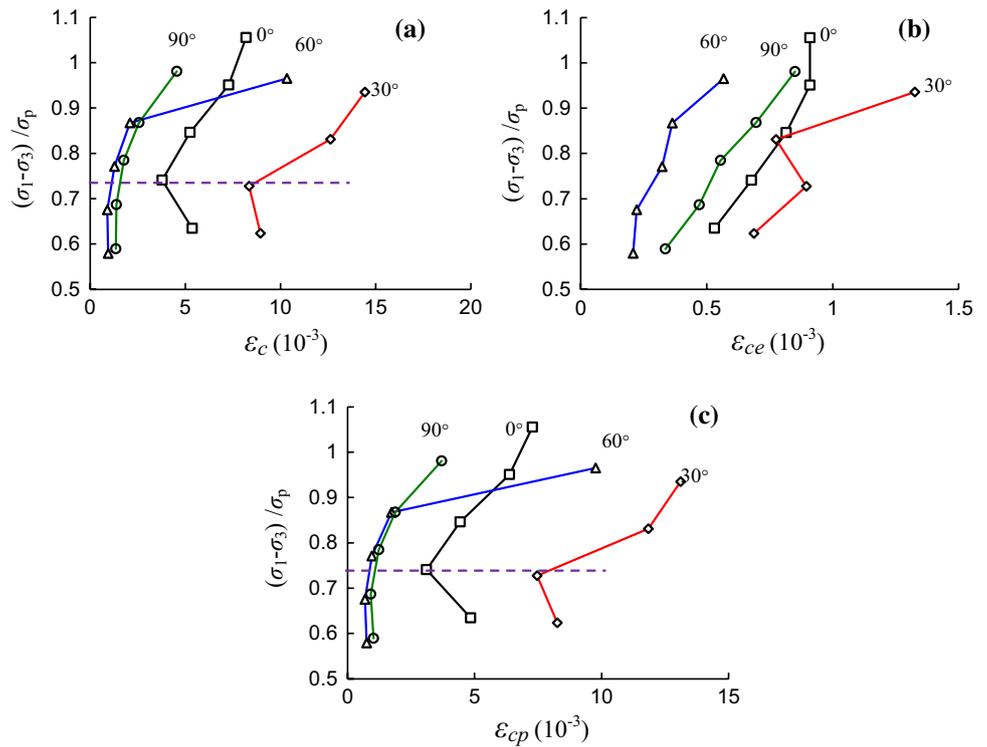
Dip angle (°)	ϵ_m			ϵ_{me}			ϵ_{mp}		
	A	B	R^2	A	B	R^2	A	B	R^2
0	0.019	0.414	0.99	0.149	0.105	0.99	0.023	0.461	0.99
30	0.010	0.386	0.99	0.179	0.121	0.99	0.011	0.409	0.99
60	0.051	0.372	0.99	0.249	0.085	0.99	0.076	0.543	0.93
90	0.061	0.430	0.96	0.129	0.141	0.99	0.082	0.524	0.99

Visco-elastic and visco-plastic deformations

Figure 8 illustrates the relationships between the time-dependent strains and the stress ratio of the rock-like specimens with four different dip angles. Figure 8a shows that axial creep strain first increased as the dip angle increased from 0° to 30° and then decreased when the dip angle increased to 60°. When the dip angle augmented to 90°, it had a slight increase in creep strain. The visco-plastic strain showed similar variation to the creep strain, as shown in Fig. 8b. The visco-elastic strain first increased as the dip angle increased to 30°, then decreased as the angle increased to 60° and finally increased as the dip angle increased to 90°. Moreover, the time-dependent strains of the 0° and 30° dip angle specimens were larger than those of the 60° and 90° dip angle specimens. There were no noticeable variations in the creep strain and visco-plastic

strain for the 60° and 90° dip angle composite rock specimens when the stress ratio increased from near 0.6 to 0.8. However, for the 0° and 30° dip angle specimens, the creep strains and visco-plastic strains reduced apparently. Subsequently, they increased with increasing stress ratio, as shown in Fig. 8c. The reason for these results may be that axial compressible deformations for low-dip-angle specimens were larger than those of the high-dip-angle specimens. At the first stress levels, rock-like specimens were commonly compacted at first. However, the maximum compressible deformation of the low dip angles was still not achieved. The large deformation of the ‘weak’ rock was dominant. In contrast, when the second stress levels were applied, the compressible deformations of the low-dip-angle specimens were relatively smaller than those at the former stress levels. Therefore, the time-dependent deformations of the low-dip-angle specimens were much smaller. After that, specimens began lateral expansion as the stress level increased.

Fig. 8 Relationships between the time-dependent strains and stress ratio of rock-like specimens



Creep rate

Creep rate reflects the relationship between rock deformation and time. When the applied stress level is low, rocks only yield primary and steady-state creep deformation, whereas it occurs tertiary creep as stress level reaches creep failure stress. Hence, it is significant to study the evolution in steady-state creep rate to predict rock creep deformation under compression conditions and its long-term strength.

The relationships between deviatoric stress and steady-state creep rate of the composite rock specimens are shown in Fig. 9. Figure 9a clearly shows that steady-state creep rate increased slightly with increasing deviatoric stress when the stress level was less than some value, whereas it augmented obviously as deviatoric stress was larger than that value, which can be termed as long-term strength. The steady-state

creep rate had an approximate exponential function relationship with deviatoric stress. Figure 9b plots the relationships between the stress ratio and the steady-state creep rate for the tested specimens. It can be seen that the steady-state creep rate had an approximate linear correlation with the stress ratio as the ratio was less than approximately 0.9. In contrast, the axial steady-state creep rate of the specimen with dip angle of 30° was the largest.

Figure 10 illustrates the variations in axial and volumetric strains and their evolution rates of the 0° dip angle rock-like specimen. It is clear that the axial strain rate and volumetric strain rate decreased with increasing creep time, which means that volumetric compression continued with creep time, while the volumetric strain rate decreased over time. When the stress levels increased to creep failure stresses, the axial and volumetric strain curves and corresponding strain

Fig. 9 Relationships between steady-state creep rate and stress level for rock-like specimens

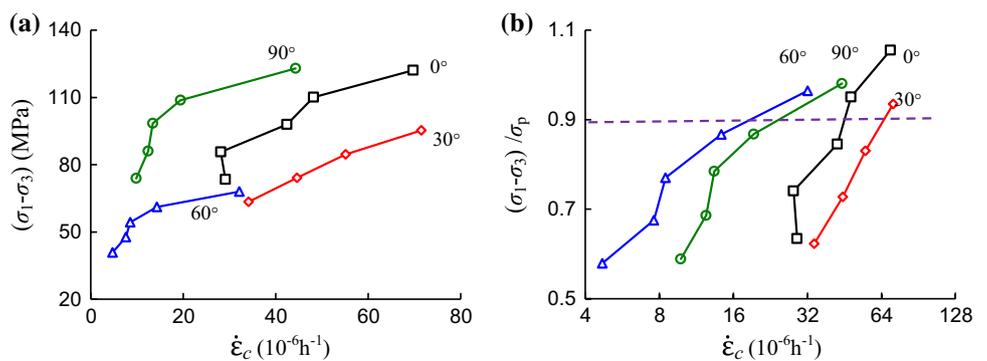


Fig. 10 Variations in axial and volumetric creep rates over time of 0° dip angle rock-like specimen

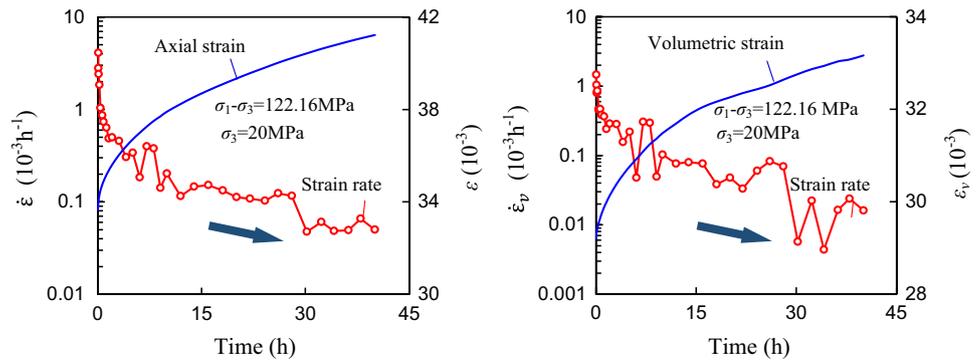
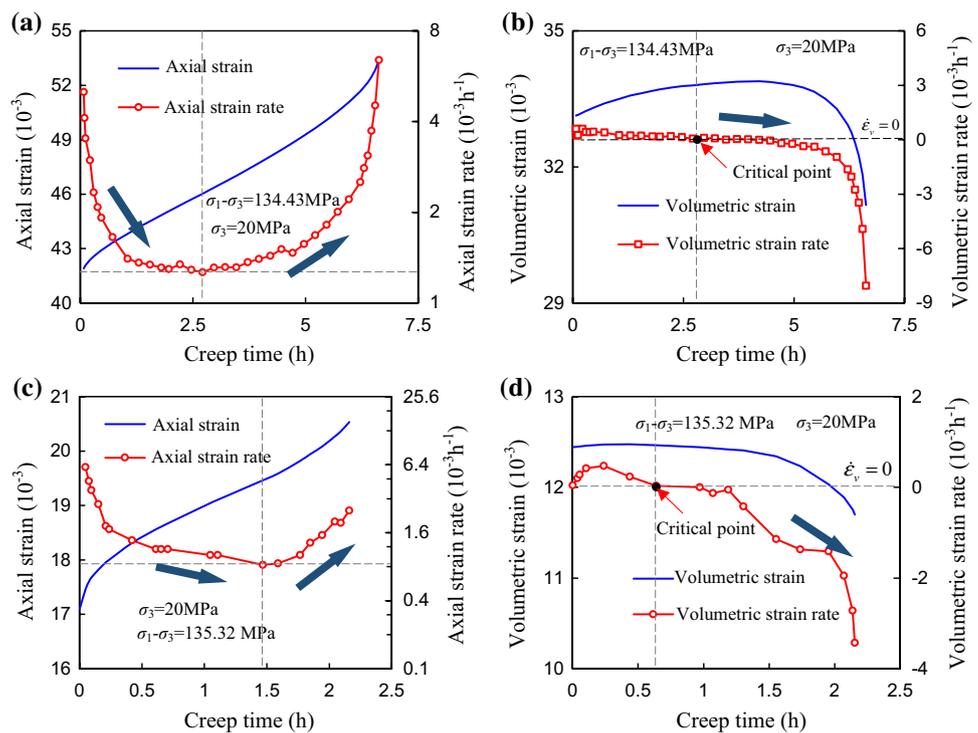


Fig. 11 Variations in axial and volumetric creep rates over time of 0° and 90° dip angle rock-like specimens

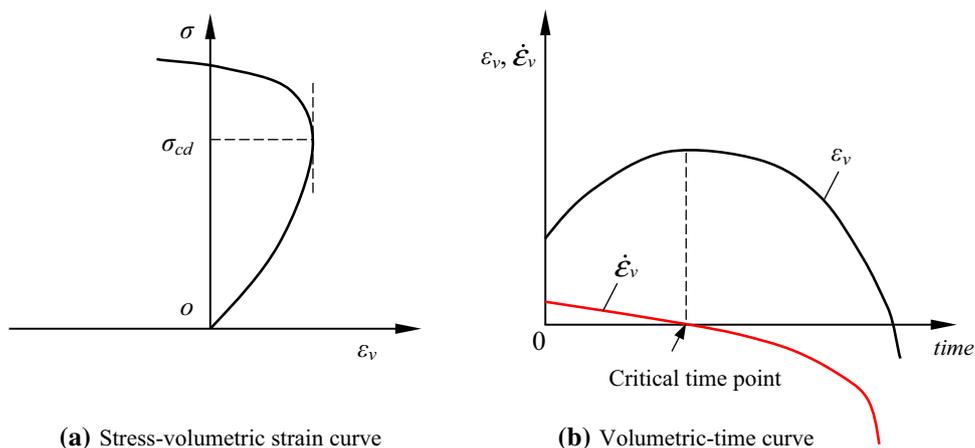


rate curves of the rock-like specimens with dip angles of 0° and 90° are as shown in Fig. 11. From Fig. 11a, b, the axial strain showed primary creep, steady-state creep and tertiary creep stages, whereas the volumetric strain first increased and then decreased at creep failure stress. During the three-stage creep, the axial strain rate first decreased to the minimum value at $t = 2.71$ h and then increased nonlinearly until accelerating creep occurred. Correspondingly, as creep time increased to 2.71 h, the volumetric strain rate decreased to $0 \times 10^{-3} \text{ h}^{-1}$; after that, it became a negative value. These results indicate that the specimen's deformation was mainly dominated by volumetric compression before $t = 2.71$ h and then was mainly dominated by shear dilatancy after $t = 2.71$ until creep failure. Therefore, the turning point of the axial strain rate can be used as a critical point for the unstable deformation of rock-like specimens. Figures 11c, d illustrate

the axial and volumetric strain curves and corresponding strain rate curves of the specimen with a dip angle of 90° at creep failure stress. They were similar to the results of the 0° specimen, whereas the critical time point of the axial strain rate curve was different than that of the volumetric strain rate curve. It is sometimes difficult to predicate the turning point of the axial strain curve; hence, the critical point of the volumetric strain curve can also be used to predicate creep failure.

When the volumetric strain rate equals 0, compressive volumetric deformation reaches its maximum value, which is termed as damage strength (σ_{cd}) in the stress–strain curve (Martin and Chandler 1994), as shown in Fig. 12a. The rock specimen deforms stably when the deviatoric stress is less than σ_{cd} , whereas cracks occur due to unstable expansion as the deviatoric stress exceeds σ_{cd} . Therefore, σ_{cd} can be

Fig. 12 Volumetric strain curve and critical creep failure point



regarded as a strength parameter for the instability failure of the specimen. However, creep deformation of the composite rock specimen is dependent on time when the constant stress is applied to σ_{cd} . It is necessary to determine the critical point of creep failure according to the volumetric strain rate curve. In a creep test, the corresponding time point at which the volumetric strain rate equals 0 can be regarded as the creep instability point, as shown in Fig. 12b.

Damage evolution and creep failure behavior

Damage evolution

To describe the deterioration of material, Kachanov (1958) initially introduced the damage concept, and the damage model of Kachanov was extended to anisotropic damage by Murakami (1983). This concept was applied to rock mass by Kyoya (1985) and was called damage tensor. Damage variable (D) can quantitatively characterize the deterioration degree of rock during the deformation process, which is difficult to measure directly. Therefore, D is calculated using other physical or mechanical parameters of materials, such as elastic modulus, ultrasonic wave velocity, density and severity, energy, strain and acoustic emission cumulants (Jin et al. 2011). However, one vital principle to definite the damage variable is that it should be easy measured to build a relationship with the macroscopic mechanics. Hence, the damage variable of the composite rock under the cyclic loading–unloading condition can be calculated using the elastic strain and the plastic strain, as shown in Eq. 4

$$D = 1 - \frac{\epsilon^e}{\epsilon}, \tag{4}$$

where ϵ^e and ϵ represent the elastic strain and total strain of the rock-like specimen, respectively. Due to the uneven lateral deformations of the ‘hard’ rock and ‘weak’ rock, D is calculated using the axial strain of the specimen.

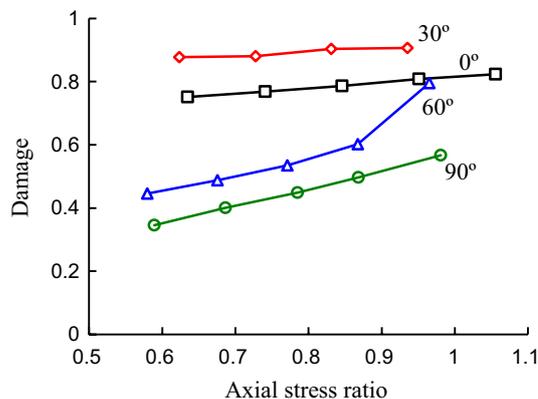


Fig. 13 Relationship between damage and stress ratio

The results of D under different deviatoric stresses are plotted in Fig. 13. From Fig. 13, D increased with a linearly increasing axial stress ratio. At an identical stress ratio, the damage degree of the rock-like specimen first increased as the dip angle increased from 0° to 30°, and then, the damage degree decreased with an increasing dip angle. As a result, the damage of the composite rock specimen with a dip angle of 30° was the largest.

Creep failure behavior

As shown in Fig. 14, under cyclic loading–unloading at a confining pressure of 20 MPa, a macroscopic shear crack crossed the bedding planes for the 0° dip angle specimen. Expansion failure took place due to ‘weak’ rocks that were squeezing under a lateral constraint. Hence, shear–expansion deformation in the ‘weak’ rock layer near loading was relatively obvious compared to the other layers. In Fig. 14a, a straight shear crack crossed the second ‘hard’ rock layer and the second ‘weak’ rock layer, whereas this crack did not propagate along the original straight line but stopped in shear–expansion failure, as shown in

Fig. 14 Failure mode after cyclic creep loading and unloading

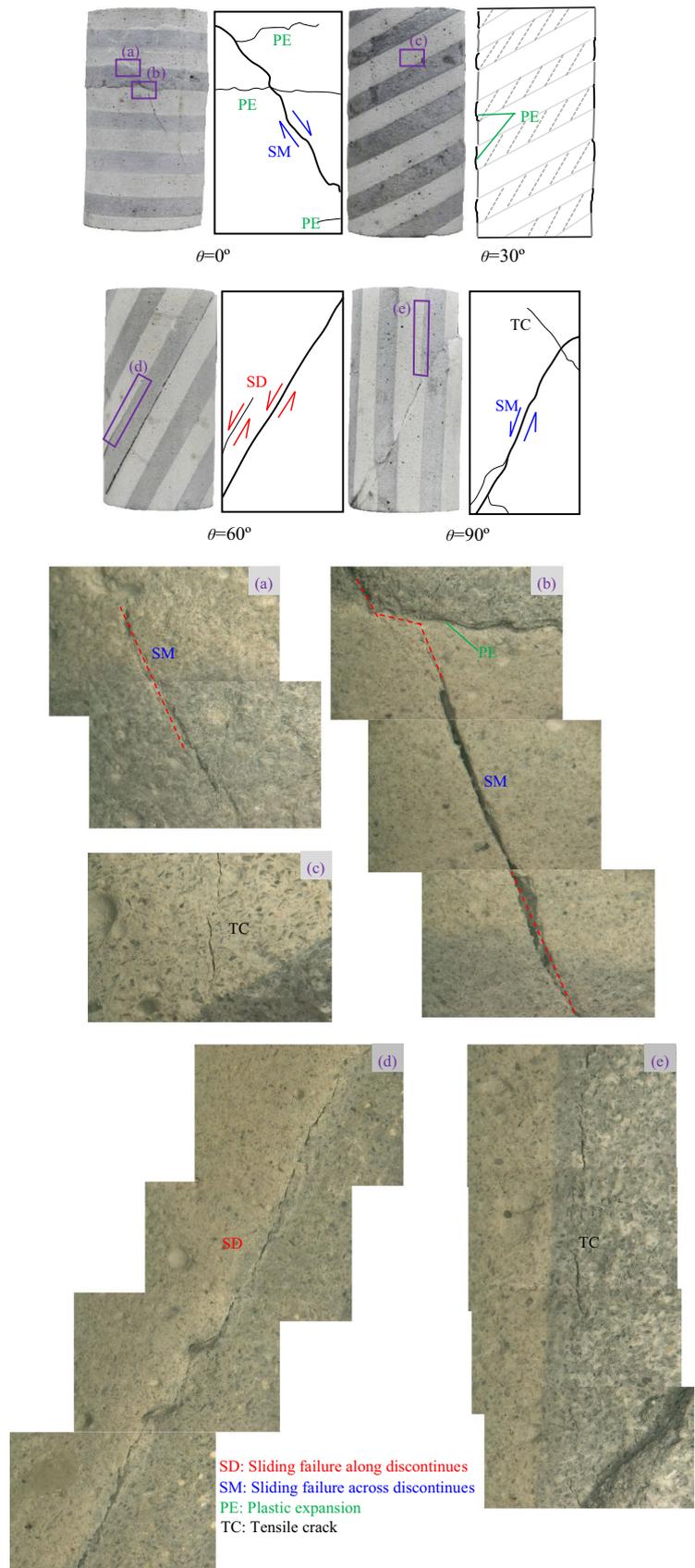


Fig. 14b. This may be due to the strain localization in the ‘weak’ rock material and failure in the material leading to stress transfer. New crack initiation and propagation were induced by the former stress transfer. The ‘hard’ rock had higher strength than the ‘weak’ rock; hence, the crack was much easy to propagate in the ‘weak’ rock material. Finally, the crack crossed the ‘hard’ rock and ‘weak’ rock as a straight line. In other words, this failure is not influenced by the rock discontinuities.

For the 30° dip angle rock-like specimen, there were no noticeable macro-cracks that appeared on the specimen under cyclic loading–unloading. However, the axial compaction and lateral expansion deformations within the ‘weak’ rocks were obvious as shown in Fig. 14. Microtensile cracks parallel to the direction of deviatoric stress within the ‘hard’ rock layers, which were adjacent to the ‘weak’ rock layers, appear under lateral 20 MPa confinement, as shown in Fig. 14c. Due to the axial cyclic compaction induced by cyclic loading and unloading, the deformation of the rock-like specimen was mainly generated in ‘weak’ rock material, and the cumulative shear displacement around discontinuity was smaller than its limit displacement. Therefore, there was no sliding crack along the discontinuities, whereas the lateral dilation of ‘weak’ rock was more obvious than the ‘hard’ rock. This lateral expansion (shear displacement) in ‘weak’ rock layer exerted tensile stress on the ‘hard’ rock, which induces tensile cracks parallel to the direction of deviatoric stress in the ‘hard’ rock around the interface of ‘hard’ and ‘weak’ rock layers.

For the composite rock specimen with a dip angle of 60°, it failed along the discontinuity in the middle part of the specimen under cyclic loading and unloading. The mechanism of this result is that the strength of the rock-like specimen was controlled by the strength of the interface, and the shear stress acting on the discontinuity was larger than the normal stress acting on the discontinuity. Therefore, the fracture orientation is same as that of the bedding plane (Tien et al. 2006). Moreover, due to the uncoordinated deformation between the ‘weak’ rock and ‘hard’ rock around the sliding failure surface, a short sliding crack appears along the interface of the ‘hard’ rock and ‘weak’ rock above the main sliding failure surface, as shown in Fig. 14d.

In the 90° dip angle composite specimen, shear failure occurred under identical loading conditions. A macro-shear sliding crack across the discontinuity appeared in the middle location of the rock-like specimen. Confining pressure provided a lateral constraint preventing the development of tensile cracks parallel to the direction of major stress. However, an unnoticeable tensile crack appear along the interface of the ‘hard’ rock and the ‘weak’ rock, which is induced by tensile stress generated by the shear sliding failure, as shown in Fig. 14e. An unobvious tensile fracture across discontinuities can be seen on the upper end of the main shear sliding

crack. This may be induced by a torque generated by the shear failure.

Therefore, dip angle has significant effect on the creep failure modes under cyclic loading and unloading conditions. When the dip angles were 0° and 90°, shear sliding failure occurred in rock-like specimens across the discontinuities, whereas plastic flow mainly occurred in the ‘weak’ rock material as the dip angle was of 30°. However, it mainly ruptured along the discontinuity when the dip angle equaled 60°.

Discussion

According to the previous results, the creep failure stresses were higher than their peak stresses, except in the specimen with a 30° layer. Considering that the axial deformation may reach a maximum value of the displacement sensor, we stop applying cyclically loadings. Therefore, the final stress level applied on the 30° specimen was lower than its peak strength. Why were the creep failure stresses higher than σ_p ? One reason may be related to specimen differences, and another may be the specificity of the composite rock stratum specimen. However, according to the results of the repetitious triaxial compression experiment as shown in Fig. 15, the peak deviatoric stresses are still larger than their creep failure stresses. This finding indicates that specimen differences are not the reason for this phenomenon. From another point of view, the loading–unloading history has great influence on the bearing capacity of rock-like specimens. The cyclic loading and unloading path may enhance the failure strength of the specimen. The specimen was compacted and gradually became much denser at 20 MPa confinement under cyclical loading conditions. Frictional work was dominant between grains. Strain accumulation occurred mainly in the soft layers. Therefore, deviatoric stress can load to higher values. Under cyclically loading, failure stress was close to

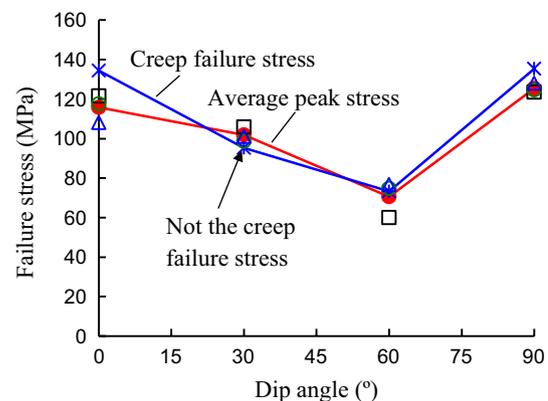
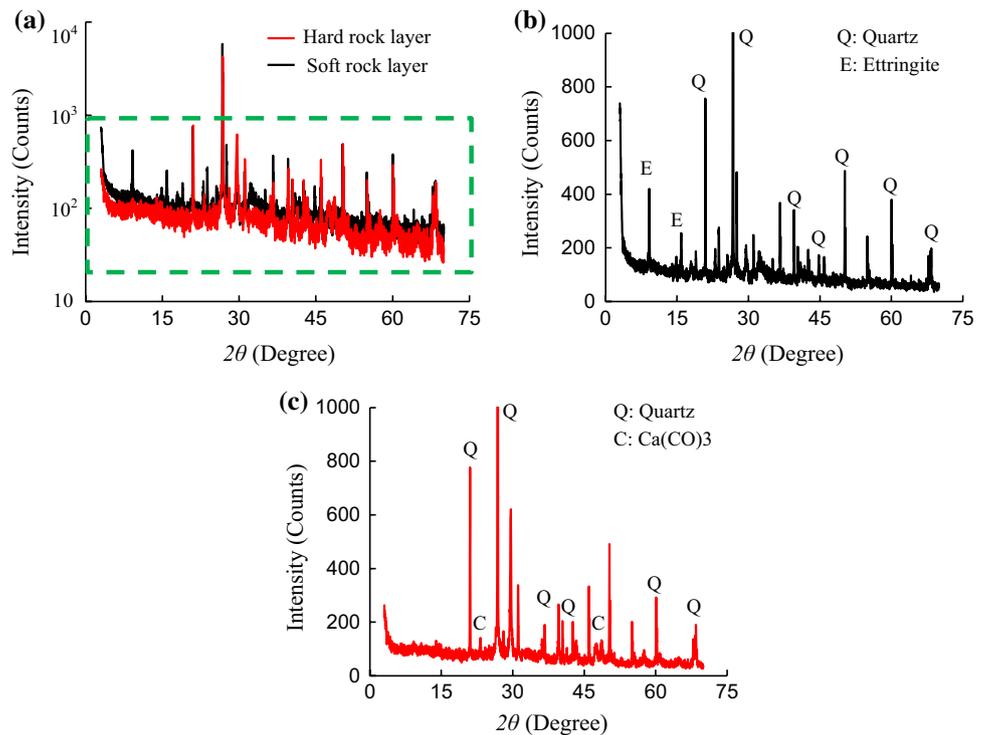


Fig. 15 Failure stresses of the rock-like specimens

Fig. 16 XRD results of the rock-like specimens



the monotonic strength and could lower the strength (Eberhardt et al. 1999; Meng et al. 2016; Yang et al. 2017). However, Yang found that cycles of loading–unloading increased the rock strength (Yang et al. 2015). Therefore, the final material failure strength under cyclical loading depends on the type of rock (Cerfontaine and Collin 2018).

Under a 20 MPa confinement, the layered rock-like specimens were gradually compacted through cyclical loading. The large deformations of the soft layers were dominant. For the rock-like specimen with 0° dip angle, the applied axial stress was acting on the layers vertically. The layers were compacted axially, and lateral expansion and squeezing existed in the soft layers. As dip angle increased to 30°, the axial stress can be divided into normal stress acting vertically on the layers and shear stress acting on the layers parallel to the layer direction. This shear stress cannot induce sliding between hard layers and soft layers. However, the soft layers obviously underwent shear sliding deformation. Therefore, the axial deformation of the specimen with 30° dip angle was the largest. For the 60° specimen, the divided shear stresses acting on layers were larger than that of the 30° specimen. When the specimen was compacted under cyclical loading, the accumulation of the deformation may reach the deformation limits between hard and soft rock layers. When the shear stress reached the friction strength between the two different layers, shear sliding occurred between the soft and hard rock layers. As a result, the deformation and strength of the 60° specimen were lower than those of the other

specimens. The applied axial stress was parallel to the layers of the rock-like specimen with 90° dip angle. The confining pressure constrains the development of tensile cracks parallel to the direction of major stress. The soft rock and hard rock yield the same axial deformation under various creep loadings; hence, the hard rock would carry more load than the soft rock and further be subjected to moment. The deformation of hard layers was dominant, so the axial deformation of the 90° specimen was the smallest. Gatelier et al. (2002) reported that compaction and microcracking are the two competing mechanisms of the layered rocks. So, natural initial anisotropy should be taken into account (Royer-Carfagni and Salvatore 2000; Gatelier et al. 2002).

Because the diffraction intensity of the quartz sand is much higher than that of other main components, the vertical axis is used by logarithmic coordinates. Figure 16b, c plots the main components of the soft and hard layers of the rock-like specimen, respectively. Quartz and ettringite are the main components of the soft layers, whereas quartz and CaCO₃ are the main components of the hard layers. The Kaolinite in soft layers produced ettringite, which can improve the ultimate strain of the specimen (Klimesch and Ray 1998; Qian and Li 2001). The hard layers had CaCO₃, which can enhance the stiffness of the specimens. As a result, the soft rock layers had larger deformation than that of the hard rock layers. Therefore, the deformation of the soft rock layers takes up a large proportion in the creep deformation of the rock-like samples.

Conclusions

Rocks anisotropy always bring difficulties to the excavation engineering (e.g., TBM entrapment). It is very important to study its mechanical behaviors especially the rheological mechanical properties. It is also very difficult practically to obtain ideal anisotropy rocks with uniform thickness and various specified dip angles. However, artificial materials can be used to make anisotropy rock specimens with a wide range of mechanical properties and specified different dip angles conveniently in the laboratory, which are also reproducible. This method has been proved to be feasible.

The time-independent (instantaneous elastic and instantaneous plastic) and time-dependent (visco-elastic and viscoplastic) deformations of the rock-like specimens with different dip angles can be divided by cyclic loading–unloading creep experiments. The results show that time-independent strains have linear correlations with the applied stress ratio, whereas time-dependent strains have nonlinear correlations with the stress ratio. The creep strain (rate) depends on the applied stress and dip angle. At relative low stress ratio, rock-like specimens show primary and secondary creep, whereas tertiary creep can be observed after secondary creep at a higher stress ratio. The deformation of the 30° dip angle specimen is the largest, whereas that of the 90° specimen is most unobvious at identical stress ratios.

The damage variable values at various stress ratios are calculated using axial strain, linearly increasing with increasing axial stress ratio. The variable values of the 30° and 90° dip angle composite rock specimens are the maximum and the minimum, respectively. Moreover, dip angle has a significant effect on the failure modes of rock-like specimens.

Acknowledgements This research was supported by the Fundamental Research Funds for the Central Universities (2015XKZD05). The authors would like to express their sincere gratitude to the editor and two anonymous reviewers for their valuable comments, which have greatly improved this paper.

References

- Amadei B (1996) Importance of anisotropy when estimating and measuring in situ stresses in rock. *Int J Rock Mech Min Sci Geomech Abstr* 33(3):293–325
- Amadei B, Rogers JD, Goodman RE (1983) Elastic constants and tensile strength of anisotropic rocks. In: *Proceedings of the 5th international congress of rock mechanics*, pp A189–A196
- Babendererde S, Hoek E, Marinos P, Cardoso AS (2004) Characterization of granite and the underground construction in metro do Porto, Portugal. In: *Proceeding of international conference on site characterization*, Porto, pp 19–22
- Barla G (1974) Rock anisotropy: theory and laboratory testing. *Rock Mech* 114:131–169
- Barla G (2016) Full-face excavation of large tunnels in difficult conditions. *J Rock Mech Geotech Eng* 8(3):294–303
- Cerfontaine B, Collin F (2018) Cyclic and fatigue behavior of rock materials: review, interpretation and research perspectives. *Rock Mech Rock Eng* 51(2):391–414
- Cheng JL, Yang SQ, Chen K, Ma D, Li FY, Wang LM (2017) Uniaxial experimental study of the acoustic emission and deformation behavior of composite rock based on 3D digital image correlation (DIC). *Acta Mech Sin* 33(06):999–1021
- Cho JW, Kim H, Jeon S, Min KB (2012) Deformation and strength anisotropy of Asan gneiss, Boryeong shale, and Yeoncheon schist. *Int J Rock Mech Min Sci* 50:158–169
- Colak K, Unlu T (2004) Effect of transverse anisotropy on the Hoek–Brown strength parameter ‘mi’ for intact rocks. *Int J Rock Mech Min Sci* 41(6):1045–1052
- Donath FA (1964) A strength variation and deformational behavior of anisotropic rocks. *State of stress in the Earth’s crust*. Elsevier, New York, pp 281–298
- Eberhardt E, Stead D, Stimpson B (1999) Quantifying progressive pre-peak brittle fracture damage in rock during uniaxial compression. *Int J Rock Mech Min Sci Geomech Abstr* 36(3):361–380
- Gatelier N, Pellet F, Loret B (2002) Mechanical damage of an anisotropic porous rock in cyclic triaxial tests. *Int J Rock Mech Min Sci Geomech Abstr* 39(3):335–354
- Gong QM, Jiao YY, Zhao J (2006) Numerical modelling of the effects of joint spacing on rock fragmentation by TBM cutters. *Tunn Undergr Space Technol* 21(1):46–55
- Gong WL, Peng YY, Sun XM (2015) Enhancement of low-contrast thermograms for detecting the stressed tunnel in horizontally stratified rocks. *Int J Rock Mech Min Sci* 74:69–80
- Gonzaga GG, Leite MH, Corthésy R (2008) Determination of anisotropic deformability parameters from a single standard rock specimen. *Int J Rock Mech Min Sci* 45(8):1420–1438
- Han B, Wang ZY, Ding XL, Xu P (2007) Numerical simulation for rheological characteristics of interbedded strata of soft and hard rock. *Chin J Yangtze River Sci Res Inst* 24(2):25–29 (in Chinese)
- He MC (2011) Physical modeling of an underground roadway excavation in geologically 45° inclined rock using infrared thermography. *Eng Geol* 121(3–4):165–176
- He MC, Xie HP, Peng SP, Jiang YD (2005) Study on rock mechanics in deep mining engineering. *Chin J Rock Mech Eng* 24(16):2803–2813 (in Chinese)
- He MC, Gong WL, Li DJ, Zhai HM (2009) Physical modeling of failure process of the excavation in horizontal strata based on IR thermography. *Int J Min Sci Technol* 19(6):689–698
- He MC, Gong WL, Zhai HM, Zhang HP (2010) Physical modeling of deep ground excavation in geologically horizontal strata based on infrared thermography. *Tunn Undergr Space Technol Inc Trenchless Technol Res* 25(4):366–376
- Jin JF, Li XB, Yin ZQ, Zou Y (2011) A method for defining rock damage variable by wave impedance under cyclic impact loadings. *Chin J Rock Soil Mech* 32(5):1385–1393+1410
- Kachanov LM (1958) Time of the rupture process under creep conditions. *Izv Akad Nauk SSSR Tech Nauk* 8:26–31
- Khanlari G, Rafiei B, Abdilor Y (2015) Evaluation of strength anisotropy and failure modes of laminated sandstones. *Arab J Geosci* 8:3089–3102
- Klimesch DS, Ray AS (1998) Autoclaved cement-quartz pastes with metakaolin additions. *Adv Cem Res* 7:109–118
- Kyoya T (1985) A damage mechanics theory for discontinuous rock mass. In: *Proceedings of the 5th international conference on numerical methods in geomechanics*, Nagoya, 1–5 April, pp 469–480
- Liu QS, Shi K, Huang X (2013) Feasibility of application of TBM in construction of deep coal mine and its key scientific problems. *Chin J Min Saf Eng* 30(5):633–641 (in Chinese)

- Malan DF, Spottiswoode SM (1997) Time-dependent fracture zone behavior and seismicity surrounding deep level stopping operations. *Rockburst and seismicity in mines*. A. A. Balkema, Rotterdam, pp 173–177
- Martin CD, Chandler NA (1994) The progressive fracture of Lac du Bonnet granite. *Int J Rock Mech Min Sci* 31(6):643–659
- McLamore R, Gray KE (1967) The mechanical behavior of anisotropic sedimentary rocks. *J Eng Ind* 89:62–73
- Meng Q, Zhang M, Han L, Pu H, Nie T (2016) Effects of acoustic emission and energy evolution of rock specimens under the uniaxial cyclic loading and unloading compression. *Rock Mech Rock Eng* 49(10):3873–3886
- Murakami S (1983) Notion of continuum damage mechanics and its application to anisotropic creep damage theory. *J Eng Mater Technol* 105(2):99–105
- Nasseri MH, Rao KS, Ramamurthy T (1997) Failure mechanism in schistose rocks. *Int J Rock Mech Min Sci* 34(3–4):219
- Naumann M, Hunsche U, Schulze O (2007) Experimental investigations on anisotropy in dilatancy, failure and creep of Opalinus Clay. *Phys Chem Earth* 32:889–895
- Niandou H, Shao JF, Henry JP, Fourmaintraux D (1997) Laboratory investigation of the mechanical behavior of Tournemire shale. *Int J Rock Mech Min Sci* 34:3–16
- Nova R (1980) The failure of transversely isotropic rocks in triaxial compression. *Int J Rock Mech Min Sci Geomech Abstr* 17:325–332
- Nunes AL (2002) A new method for the determination of transverse isotropic orientation and associated elastic parameters for intact rock. *Int J Rock Mech Min Sci* 9:257–273
- Qian XQ, Li ZJ (2001) The relationships between stress and strain for high-performance concrete with metakaolin. *Cem Concr Res* 31:1607–1611
- Ramamurthy T (1993) Strength, modulus responses of anisotropic rocks. In: Hudson JA (ed) *Compressive rock engineering*, vol 1. Pergamon, Oxford, pp 313–329
- Royer-Carfagni G, Salvatore W (2000) The characterization of marble by cyclic compression loading: experimental results. *Mech Cohes Frict Mater* 5(7):535–563
- Sanio HP (1985) Prediction of the performance of disc cutters in anisotropic rock. *Int J Rock Mech Min Sci Geomech Abstr Geomech Abstr* 22(3):153–161
- Shang YJ, Yang ZF, Zeng QL, Sun YC, Shi YY, Yuan GX (2007) Retrospective analysis of TBM accidents from its poor flexibility to complicated geological conditions. *Chin J Rock Mech Eng* 26(12):2404–2411 (in Chinese)
- Stille H, Palmström A (2008) Ground behaviour and rock mass composition in underground excavations. *Tunn Undergr Space Technol* 23(1):46–64
- Tien YM, Kuo MC (2001) A failure criterion for transversely isotropic rocks. *Int J Rock Mech Min Sci* 38:399–412
- Tien YM, Tsao PF (2000) Preparation and mechanical properties of artificial transversely isotropic rock. *Int J Rock Mech Min Sci* 37:1001–1012
- Tien YM, Kuo MC, Juang CH (2006) An experimental investigation of the failure mechanism of simulated transversely isotropic rocks. *Int J Rock Mech Min Sci* 43:1163–1181
- Vogel M, Andrast HP (2000) Alp transit-safety in construction as a challenge: health and safety aspects in very deep tunnel construction. *Tunn Undergr Space Technol* 15(4):481–484
- Wang AM, Li XG, Yang CH, Huang ZQ (2010) Study of interaction between creep deformation of bedded salt rock. *Chin J Rock Soil Mech* 31(12):3965–3970 (in Chinese)
- Xia CC, Zhong SY (1989) Experimental data processing method in consideration of influence of loading history on rock specimen deformation. *J Cent South Inst Min Metall* 20(1):18–24 (in Chinese)
- Xie HP, Gao F, Ju Y (2015) Research and development of rock mechanics in deep ground engineering. *Chin J Rock Mech Eng* 34(11):2161–2178 (in Chinese)
- Xu P, Yang SQ (2015) Study of visco-elasto-plastic constitutive model of coal under cyclic loading. *Chin J Rock Mech Eng* 34(3):537–545 (in Chinese)
- Yang SQ, Hu B (2018) Creep and long-term permeability of a red sandstone subjected to cyclic loading after thermal treatments. *Rock Mech Rock Eng* 51:2981–3004
- Yang SQ, Ranjith PG, Huang YH, Yin PF, Jing HW, Gui YL, Yu QL (2015) Experimental investigation on mechanical damage characteristics of sandstone under triaxial cyclic loading. *Geophys J Int* 201(2):662–682
- Yang SQ, Tian WL, Ranjith PG (2017) Experimental investigation on deformation failure characteristics of crystalline marble under triaxial cyclic loading. *Rock Mech Rock Eng* 50(11):2871–2889
- Yang SQ, Tian WL, Huang YH (2018) Failure mechanical behavior of pre-holed granite specimens after elevated temperature by particle flow code. *Geothermics* 72:124–137
- Zhao YL, Zhang LY, Wang WJ, Wan W, Li SQ, Ma WH, Wang YX (2017) Creep behavior of intact and cracked limestone under multi-level loading and unloading cycles. *Rock Mech Rock Eng* 50(6):1409–1424



Forecasting seismic activity rates in northwest Himalaya through multivariate autoregressive forecast of seismicity algorithm

Prasanta Chingtham¹ · Anurag Tiwari¹ · Arabind Kumar Yadav²

Received: 14 September 2017 / Accepted: 19 February 2019 / Published online: 27 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

In this study, a model based on multivariate autoregressive forecast of seismicity (MARFS) algorithm is adopted to forecast seismic activity rates in northwest Himalaya, using the compiled homogenized moment magnitude (M_W) based catalogue. For this purpose, each source zone delineated by Yadav et al. (Pure Appl Geophys 170:283–295, 2012) is divided into a spatial grid interval of $0.5^\circ \times 0.5^\circ$ while the entire catalogue span (1975–2010) is segregated into six time periods/grids to estimate seismic activity rates spatially and temporally. These seismic activity rates which are estimated from spatial density map of hypocenters exhibit high values in Chaman Fault (Zone 1), Hindukush-Pamir region (Zone 3) and the mega thrust systems, i.e., Main Central Thrust, Main Boundary Thrust and Himalayan Frontal Thrust (Zone 4). Then, the seismic activity rates during 2011–2016 could be forecasted by extrapolating (through auto-regression procedure) those observed for previous time periods. The forecast seismic activity rates are estimated within the values of 0 and 7.57 with high values primarily observed in Hindukush-Pamir region of Zone 3 and the gently north-dipping thrust fault systems (Main Central Thrust, Main Boundary Thrust, Himalayan Frontal Thrust) of Zone 4. Finally, the associated area under the curve of receiver operating characteristics graph suggests the superiority of forecasting model with respect to random prediction, whereas results of the data-consistency test, i.e., N test of our model, exhibit consistency in between the observed and simulated likelihoods. Moreover, the hypothetical t test performed in between the spatial grids of forecast seismic activity rates and observed seismic activity rates confirms that the former is consistent with the latter.

Keywords Seismicity · Autoregression · Forecast model · Northwest Himalaya

Introduction

Northwest Himalaya is situated in the seismically active young Himalayan mountain belt formed during the continuous convergence of about 45 mm/year between the Indian–Eurasian continental plates (DeMets et al. 1994; Bilham et al. 1997). The region has experienced major destructive earthquakes like 1998 M_W 6.6 Afghanistan–Tajikistan earthquake, 2002 M_W 6.3 Hindukush earthquake and 2005 M_W 7.6 Kashmir earthquake in the recent past. The triggering mechanism of these earthquakes in this region can be primarily attributed to crustal shortening caused by

subducting Indian plate under the sedimentary wedges of Himalaya (Bilham and Gaur 2000) and remnant slab of oceanic lithosphere (Chandra 1978). Historical documents of the destructive earthquakes in this area can also be traced back to eighth century (Ambraseys and Bilham 2003). The study region is comprised of zones V, IV and III, respectively, in seismic hazard zonation map of India (BIS 2002) that corresponds to peak ground acceleration value greater than or equal to 0.4 g, 0.25 g and 0.2 g respectively. Moreover, other seismic hazard studies in this region have estimated higher value of seismic hazard level (Bhatia et al. 1999; Rout et al. 2015). The region has susceptibility to large earthquakes which poses a serious threat to millions of people living in and around this gigantic mountain belt. Therefore, successful forecasting of seismic activity rates in this particular region can provide an important input for mitigating the disastrous effect of earthquakes.

Recently, research works on earthquake forecast models have tremendously increased in order to prepare mitigating

✉ Prasanta Chingtham
prasantachingtham@gmail.com

¹ National Center for Seismology, Ministry of Earth Sciences, New Delhi 110003, India

² Geological Survey of India, Kolkata 700016, India

plans for considerably reducing the casualties incurred by earthquakes (Marzocchi and Lombardi 2009). As such, regional earthquake likelihood models (RELM) and the collaborative study of earthquake predictability (CSEP) emerged as two foremost groups that aimed at providing the best earthquake forecast model through testing the reliability and skill of different proposed models in different seismic regimes of the world (Field 2007; Gerstenberger and Rhoades 2010; Zechar et al. 2010). The proposed models include epidemic type aftershock (ETAS) model (Ogata 1998), every earthquake a precursor according to scale (EEPAS) model (Rhoades and Evison 2004, 2005, 2006; Console et al. 2006; Rhoades 2007) and proximity to past earthquakes (PPE) model (Jackson and Kagan 1999; Rhoades and Evison 2004). These seismicity-based models differ from each other on the assumption of its different generating processes underlying the earthquake occurrences. Besides these, several other models have been developed using geophysical parameters such as local kinematic of surface velocities, Coulomb stress changes, geodetically observed strain rate, fault geometry and slip rate (Kagan et al. 2003; Schorlemmer et al. 2005; Bird and Zhen 2007; Helmstetter et al. 2007; Shen et al. 2007; Marzocchi and Zechar 2011; Toda and Enescu 2011; Marzocchi et al. 2012) for forecasting large damaging future earthquakes. However, the reliability and skill of these forecasting models should be precisely examined to select the best one for successful forecasting of earthquakes (International Commission on Earthquake Forecasting for Civil Protection 2009).

Smyth and Mori (2011) developed a statistical model based on multivariate autoregressive forecast of seismicity (MARFS) algorithm to forecast the number of earthquakes and b value of Gutenberg–Richter (GR) distribution in Japan. The multivariate mixture model is considered to identify high-density areas from earthquake clustering for calculating high seismic activity rates. The autoregressive procedure is also used to detect temporal variations of earthquake occurrences and their magnitude distributions. Then, rate adjustment for higher magnitudes is finally incorporated in this model for considering the increased probability of large earthquake occurrences after large elapsed time since previous earthquake. Therefore, adoption of spatial and temporal variations of earthquake occurrences and spatial density estimation technique makes this model different from other statistical earthquake clustering models (Ogata 1998; Console and Murru 2001; Smyth and Mori 2011).

In this present study, forecasting of seismic activity rates, expressed in terms of number of events per year, for earthquakes of magnitude $M_w \geq 4$ has been accomplished by using a modified MARFS algorithm in northwest Himalaya during the period 2011–2016. As such, a homogeneous earthquake catalogue in moment magnitude M_w , covering the period 1975–2016, compiled by Chingtham et al. (2017),

has been used. Since the study region is comprised of an area having a different level of seismicity and seismic hazard (BIS 2002), the earthquake source associated seismicity also varies (Chandra 1978). In order to have an area having uniform source mechanism and varying seismicity, the source zones given by Yadav et al. (2012) have been considered as five zones for application of MARFS algorithms.

Seismotectonic background

The study region is situated at western extremity of the Himalayan belt where it exhibits a semicircular arc (Chandra 1978). The compressive lithospheric block of this region is primarily surrounded by the Indian plate, the Afghan block, the Turan block and the Tarim block from south, southwest, west and northeast directions respectively (Koulakov and Sobolev 2006; Yadav et al. 2011). This lithospheric block is considered to be formed during continuous continent–continent collision between the Indian and Eurasian plates. With reference to Eurasian plate, the Indian plate is moving at a speed of 45 mm/year in this region (DeMets et al. 1994). The continuous collision has also resulted in crustal shortening, thereby forming the successive zones of tectonic deformation manifested along the major boundary faults/features (Gansser 1964; Lyon-Caen and Molnar 1983; Seeber and Armbruster 1984; Valdiya 2003). The major fault systems, from north to south, include the Main Central Thrust (MCT), Main Boundary Thrust (MBT) and Himalayan Frontal Thrust (HFT). Toward north, these gently north-dipping faults, i.e., MCT, MBT, HFT merge into a detachment plane known as Main Himalayan thrust (MHT) (Arora et al. 2012; Gupta and Gahalaut 2014). Besides these, various features/folds such as ridges, fracture zones and transverse faults also exist and control the locations and magnitudes of the earthquakes (Scholz and Small 1997; Robinson et al. 2006). In this region, MCT and MBT manifested the crustal shortening along the Himalayan belt while HFT represents the most active thrust during the quaternary period. During the Cenozoic and Mesozoic era, evidences of intensive folding and thrusting that represent the present-day structures and lithology of Himalaya are observed in the study region (Gansser 1964; Philip et al. 2014). All these existing folds and faults are along the NW to SE trend, and their structural formation or reactivation controls the earthquake activities in the study region. Figure 1 depicts the seismotectonic map of the study region drawn with the compiled earthquake catalogue having minimum magnitude, $M_w \geq 4$, during the period 1975–2010.

On the basis of current earthquake occurrences, tectonic features, orientation of focal planes and paleoseismological data, the study region is broadly divided into five potential source zones by Yadav et al. (2012). These zones include Sulaiman–Kirthar ranges (Zone 1), Northern Pakistan and

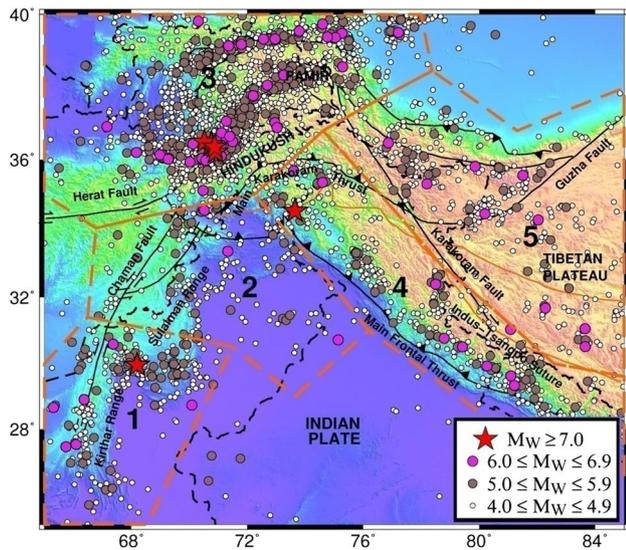


Fig. 1 A seismotectonic map of the study region prepared with the epicentral distribution of independent earthquakes (mainshocks) of $M_W \geq 4$ occurred during the period 1975–2010. Major fault systems of the northwest Himalaya are shown by black lines. Dashed lines indicate the political boundary. Red dashed lines indicate five broad source zones delineated on the basis of seismicity, tectonics and focal mechanisms of earthquakes as suggested by Yadav et al. (2012). All zones are indicated by number, i.e., 1, 2, 3, 4 and 5 (color figure online)

Hazara Syntaxis (Zone 2), Hindukush-Pamir (Zone 3), Himalayan Frontal Thrust (HFT) (Zone 4) and Tibetan Plateau region (Zone 5). Considering the similar focal depth distribution, Zone 3 is further subdivided into shallow and intermediate-depth zones by Yadav et al. (2012). However, these delineated two zones are considered as a single zone for forecasting purpose in the present study. Zone 1 comprises the southwestern Balochistan province of Pakistan where the devastating earthquake of magnitude M_W 7.7 occurred near Quetta, Pakistan, in 1935. The ongoing India–Eurasia plate interaction caused a transcurrent movement along the strike-slip Chaman Fault (Fig. 1). Zone 2 exhibits diffused seismicity along the Salt-Ranges and Hazara Syntaxis. The subduction of Indian plate beneath the Himalayan rock inflicted internal deformation within this zone (Armbruster et al. 1978; Bernard et al. 2000). Zone 3 consists of most seismically active regions of Hindukush-Pamir thrust. This active source zone is mainly associated with northward-subducting Indian plate below Hindukush and southward-subducting Eurasian plate beneath Pamir (Chatelain et al. 1980; Burtman and Molnar 1993; Fan et al. 1994). Here, shallow- to intermediate-depth earthquakes occur with predominantly reverse and strike-slip motion. Zone 4 is characterized by highly elevated Himalayan mountain belt formed by low dipping Indian Plate beneath the Eurasian landmass. North-dipping mega thrusts such as MCT, MBT,

HFT are the primary contributors of seismicity to this zone. 1905 M_W 7.8 Kangra, 2005 M_W 7.6 Kashmir, 1991 M_W 6.8 Uttarkashi and 1999 M_W 6.5 Chamoli are the major devastating earthquakes that occurred in this zone. Zone 5 lies on the Tibetan Plateau region where seismicity is related to the existing strike-slip Karakoram Fault, the Kunlun Fault and the Altyn–Tagh Fault (Yadav et al. 2012). Both normal and right-lateral strike-slip faulting earthquakes are quite prevalent in this zone.

Earthquake database

For this study, data are downloaded from International Seismological Center (ISC) and Global Centroid Moment Tensor (GCMT) databases covering the time period 1975–2016. However, the data comprises of various magnitude scales such as body-wave magnitude (m_b), surface-wave magnitude (M_S), local magnitude (M_L) and seismic moment magnitude (M_W). All these scales, i.e., m_b , M_S and M_L , exhibit saturation at different higher magnitudes, and these subsequently require conversion of these scales into single-magnitude scale (M_W) for avoiding erroneous results in the analysis (Bormann et al. 2007; Thingbaijam et al. 2008, 2009; Loannis et al. 2011; Yadav et al. 2011; Chingtham et al. 2014, 2015, 2016, 2017). This conversion can be accomplished by correlating different magnitude scales documented for the same earthquakes, thereby establishing the region-specific regression relationships among the different magnitude scales. Here, the regression techniques and equations developed by Chingtham et al. (2017) for northwest Himalaya are employed for compiling the homogeneous earthquake catalogue in M_W . A declustering method comprising of spatial and temporal window technique established by Knopoff (2000) is used to remove the foreshocks and aftershocks from mainshocks.

Methodology

For incorporating the temporal variability of seismic activity rates into forecast seismic activity rates, time span of earthquake catalogue (i.e., 1975–2010) is segregated into six time periods or grids, i.e., 1975–1980, 1981–1986, 1987–1992, 1993–1998, 1999–2004 and 2005–2010. Here, selection of particular time period is a trade-off between seismicity and the number of time windows in the entire catalogue. This implies that the number of time windows that depends on both the length of time period and the number of earthquakes should not be small enough to miss the local variations in seismicity arising from insufficient number of earthquakes. However, large number of earthquakes will produce fewer number of time windows and eventually this will give

unauthentic results in the forecast seismic activity rates produced by autoregression model. Also, the study region is also divided into spatial grids of $0.5^\circ \times 0.5^\circ$ for considering the spatial variations of seismic activity rates into forecast seismic activity rates. During the particular time grid 1975–1980, the spatial density map of each source zone is calculated to examine the earthquake clustering. In each source zone, the spatial density map can be accomplished through a well-defined clustering technique known as mixture model (Everitt 1993). This novel technique considers that hypocenters of the earthquakes can be considerably described by probability density functions given by a mixture of M component density functions in unknown proportions. This model best explains the earthquakes by finding the M mean vectors, and consequently, the entire space of the hypocenters can be represented by its density through estimating the M covariance matrices as given by following equation:

$$f(x, \theta) = \sum_{m=1}^M \delta_m \phi\left(x, \mu_m, \Sigma_m\right), \quad 0 \leq \delta_m \leq 1, \quad \sum_{m=1}^M \delta_m = 1 \quad (1)$$

where δ_m gives the mixing proportions and θ represents the vector of unknown parameters $\delta_m, \mu_m, \Sigma_m, m = 1, \dots, M$. The component density functions being a multivariate normal can be written as follows:

$$\begin{aligned} & \phi\left(x, \mu_m, \Sigma_m\right) \\ &= \frac{1}{2\pi^{d/2}} \left| \Sigma_m \right|^{-1/2} \exp\left(-\frac{1}{2(x - \mu_m)^T \cdot \Sigma_m^{-1} (x - \mu_m)}\right) \end{aligned} \quad (2)$$

where d denotes the number of independent variables and both μ_m and Σ_m are the m th component mean and covariance matrix. The associated parameters can be calculated via expectation–maximization (EM) algorithm (Dempster et al. 1977; McLachlan and Ng 2009).

By doing so, all the spatial grids in each source zone are assigned density so that density at all the spatial grids in the space is nonzero as also guaranteed by mixture model. Then, the spatial density map in each source zone is normalized to obtain the sum of density at all the spatial grids to be one. Finally, the seismic activity rate at each spatial grid is calculated by multiplying the density of each grid by seismic activity of the zone during the time grid 1975–1980. This computation process is repeated for the remaining seismic sources, and thus, seismic activity rates during the period 1975–1980 are obtained as shown in Fig. 2a. Similar procedure has been adopted to compute seismic activity rates in the region for the remaining time grid, i.e., 1981–1986, 1987–1992, 1993–1998, 1999–2004 and 2005–2010 as depicted in Fig. 2b–f. In other words, we can say that each spatial grid in the study region has calculated six seismic

activity rates corresponding to six different time grids during the period 1975–2010. Each time grid is equal to one time grid having time span of 6 years. These seismic activity rates in each spatial grid can be extrapolated during the forecast period 2011–2016 by employing the autoregression model. Extrapolation of seismicity, i.e., correlation of future and past seismicity, can be done because several researchers believed that the time of occurrences of future earthquakes largely depends on the size and the time of occurrences of the last earthquakes. In other words, we can say that earthquakes in fault zones are caused by the constant buildup and release of strain in the earth's crust (Gardner and Knopoff 1974; Nishenko and Singh 1987). Hence, the autoregression model predicts the future variable from the discrete time grid, comprising the random variables at equally spaced points. In this process, the variable on the time grid is regressed on its own previous values of the same time grid, thereby considering that the value of the variable at any time can be estimated from the weighted sum of the previous values (Chatfield 2004) as given by equation written below:

$$\chi^T = \sum_{i=1}^q \beta_i \chi^{T-i} + \varepsilon^T \quad (3)$$

where β and χ imply the vector of predictor coefficients and the calculated seismic activity rates at each spatial grid of the particular time grid respectively, and χ^T represents the χ value of the time grid prior to the interested time grid of the forecasting period, i.e., the most recently observed value. The order (q) of the autoregressive model can be reliably estimated from the Akaike Information Criterion formulated by Akaike (1974). A Yule–Walker equation is hereby adopted to estimate the vector β as given by the following equation:

$$\begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ \vdots \\ \vdots \\ r_q \end{pmatrix} = \begin{pmatrix} 1 & r_1 & r_2 & \cdots & r_{q-1} \\ r_1 & 1 & r_1 & \cdots & r_{q-2} \\ r_2 & r_1 & 1 & \cdots & r_{q-3} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ r_{q-1} & r_{q-2} & \cdot & \cdots & 1 \end{pmatrix} * \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \cdot \\ \cdot \\ \cdot \\ \beta_q \end{pmatrix} \quad (4)$$

where r_i is the observed correlation between the variables in the time grid that are separated by i steps (Chatfield 2004). After precisely determining the values of β , we can calculate the value at each grid during the forecasting period by simply replacing T by $T+1$ in Eq. 3. In this present work, the optimal order of autoregression model, i.e., q value, is estimated from the partial autocorrelation function given by Yule–Walker equation (Eq. 4).

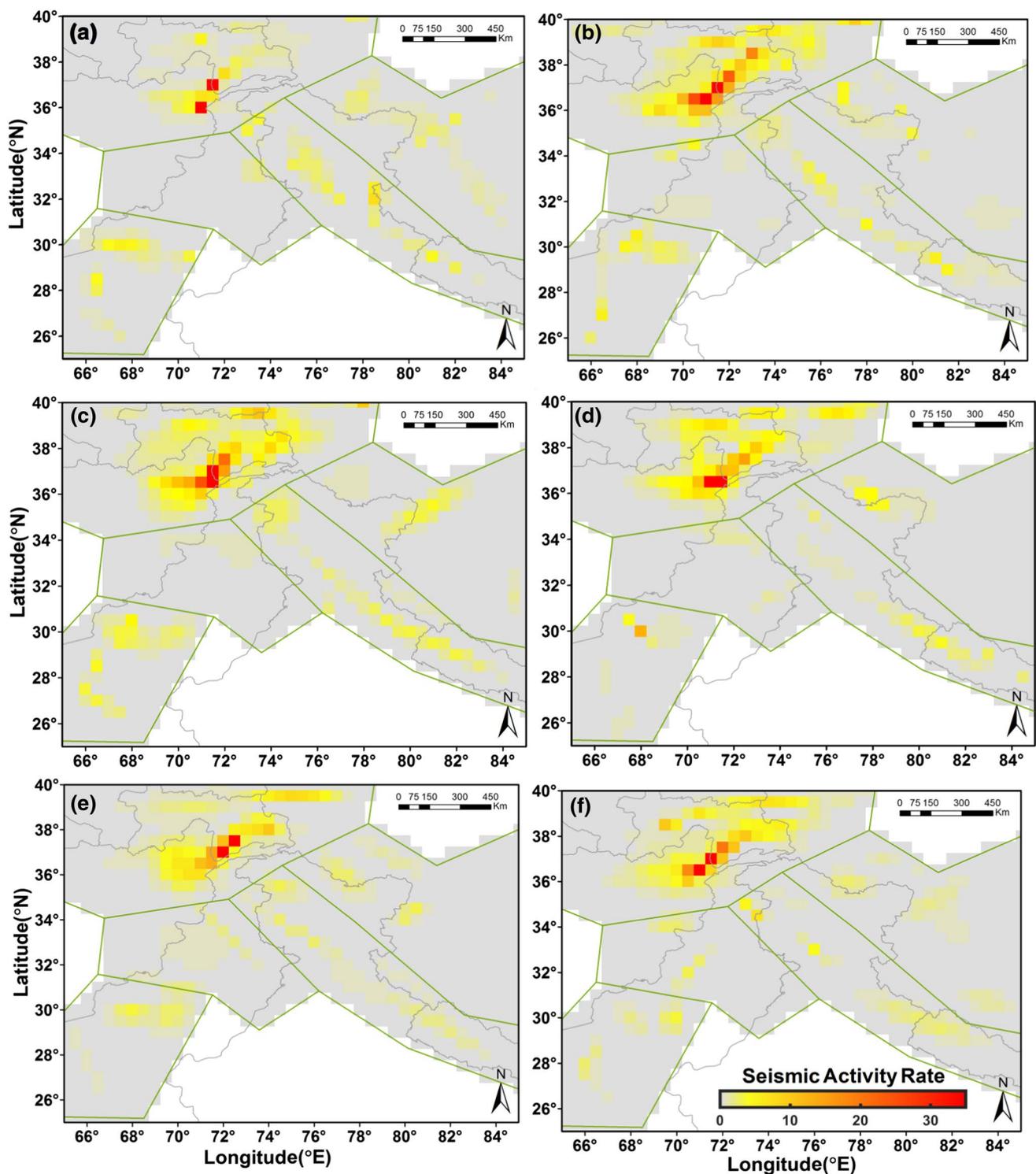


Fig. 2 Seismic activity rate map of northwest Himalaya for $M_w \geq 4.0$ during the time grids **a** 1975–1980 **b** 1981–1986 **c** 1987–1992 **d** 1993–1998 **e** 1999–2004 and **f** 2005–2010 through the normalized multivariate Gaussian mixture model

As depicted in Fig. 3, the coefficient value is found to be decreasing with increasing lags and become insignificant after lag three. Therefore, we have selected q value of 3 for

estimating the seismic activity rates at every spatial grids in the study region (Fig. 4a). This implies that increased seismic activity rates tend to decrease the forecast seismic

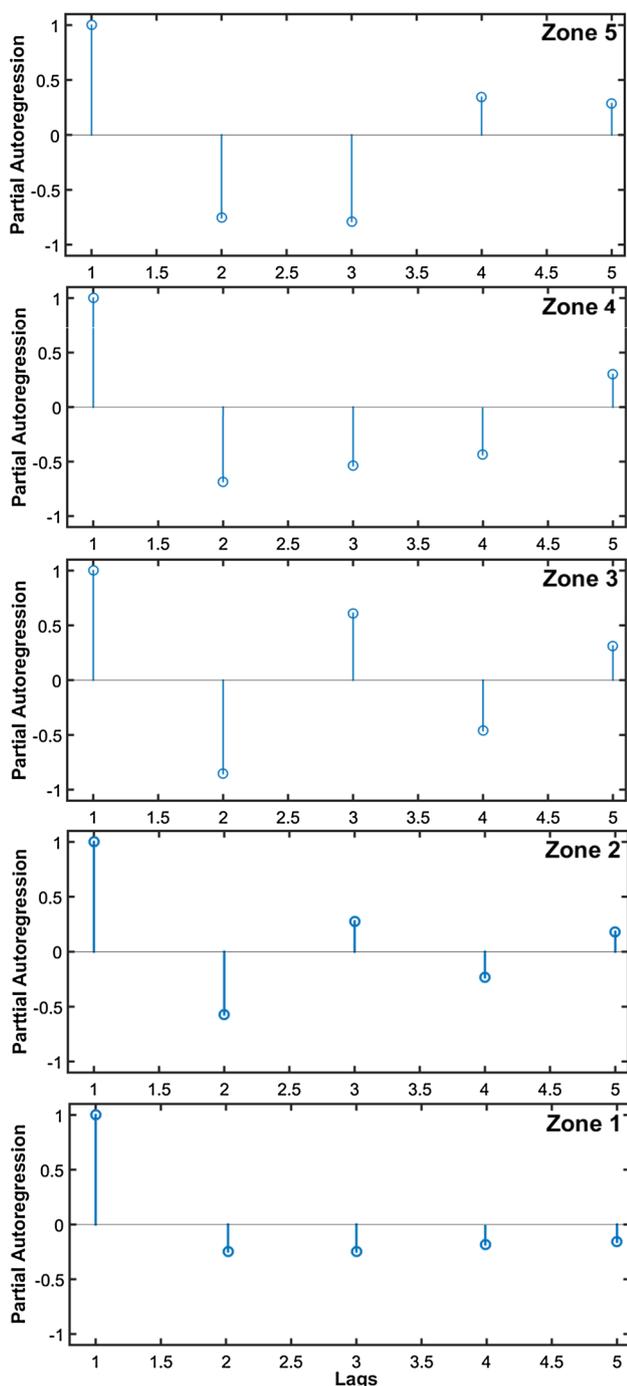


Fig. 3 Variations in partial autocorrelation coefficients for autoregressive model of various lags at particular five grid points, each grid point representing each source zone in the study region

activity rates, even though the decrease rate in each source zone is found to be unequal. A flowchart describing the steps of our methodology is shown in Fig. 5.

For comparing the performance of our model, an inhomogeneous Poisson model is constructed from the seismic activity rates estimated from the spatial density map of our

statistical model based on MARFS algorithm. In this model, forecast seismic activity rates can be calculated by simply taking the average of the previous value of seismic activity rates at the same spatial grid and hence forecast seismic activity rates can be estimated from the following relation:

$$X^T = \frac{1}{N} \sum_{i=1}^{T-N} B_i^{T-N} \tag{5}$$

where X gives the average of the previous values at the same spatial grid; B_i^{T-N} denotes the values of seismic activity rates at each spatial grid, whereas N represents the number of time grids, respectively. The forecast seismic activity rate map reproduced by inhomogeneous Poisson model is drawn with the observed seismic activity rates for magnitude $M_W \geq 4.0$ during the time period 2011–2016 (Fig. 4b and c).

Validation of our earthquake forecasting model in terms of reliability and skill is an important and essential procedure before applying this model to different tectonic regimes of the world.

Here, we will also demonstrate the ability of forecast model during the period 2011–2016 by utilizing receiver operating characteristic (ROC) curve (Fawcett 2006; Murru et al. 2009). The ROC curve is defined by the plot between true positive rate and false positive rate. The true positive rate can be estimated from $\frac{\gamma}{\gamma+\vartheta}$, whereas the false positive rate is calculated from $\frac{\xi}{\xi+\omega}$. The value ϑ is the sum of all spatial grids with forecast seismic activity rates less than a very small assumed alarm rate. Whereas, all the spatial grids with forecast seismic activity rates greater than the particular alarm rate are summed up to give the value γ . These alarm rates are produced by varying the threshold values, and earthquakes are assumed to be expected or not expected in any spatial bins with seismic activity rates greater or lower than the threshold respectively. Similar fashion can be used to compute ξ and ω . This process is repeated by increasing the alarm rate where each alarm rate corresponds to a point on the ROC curve. The idea behind the use of both measures, i.e., true positive and false positive prediction rates, lies on the fact that when using these measures in isolation, one can simply optimize them to obtain zero false positive rates for specifying a prediction that covers the entire experiment space-time. Finally, area under the curve (AUC) is computed and its value is compared with the AUC obtained from random prediction (i.e., $AUC=0.5$) to validate our forecasting of seismic activity rates. Figure 6a depicts the ROC curves comparing the predicted seismic activity rates by random model and our forecast model during the four different time grids, i.e., 1993–1998, 1999–2004, 2005–2010 and 2011–2016. The stability of this forecasting model can be primarily investigated from these ROC curves, thereby examining the considerably fluctuated seismic activity rates

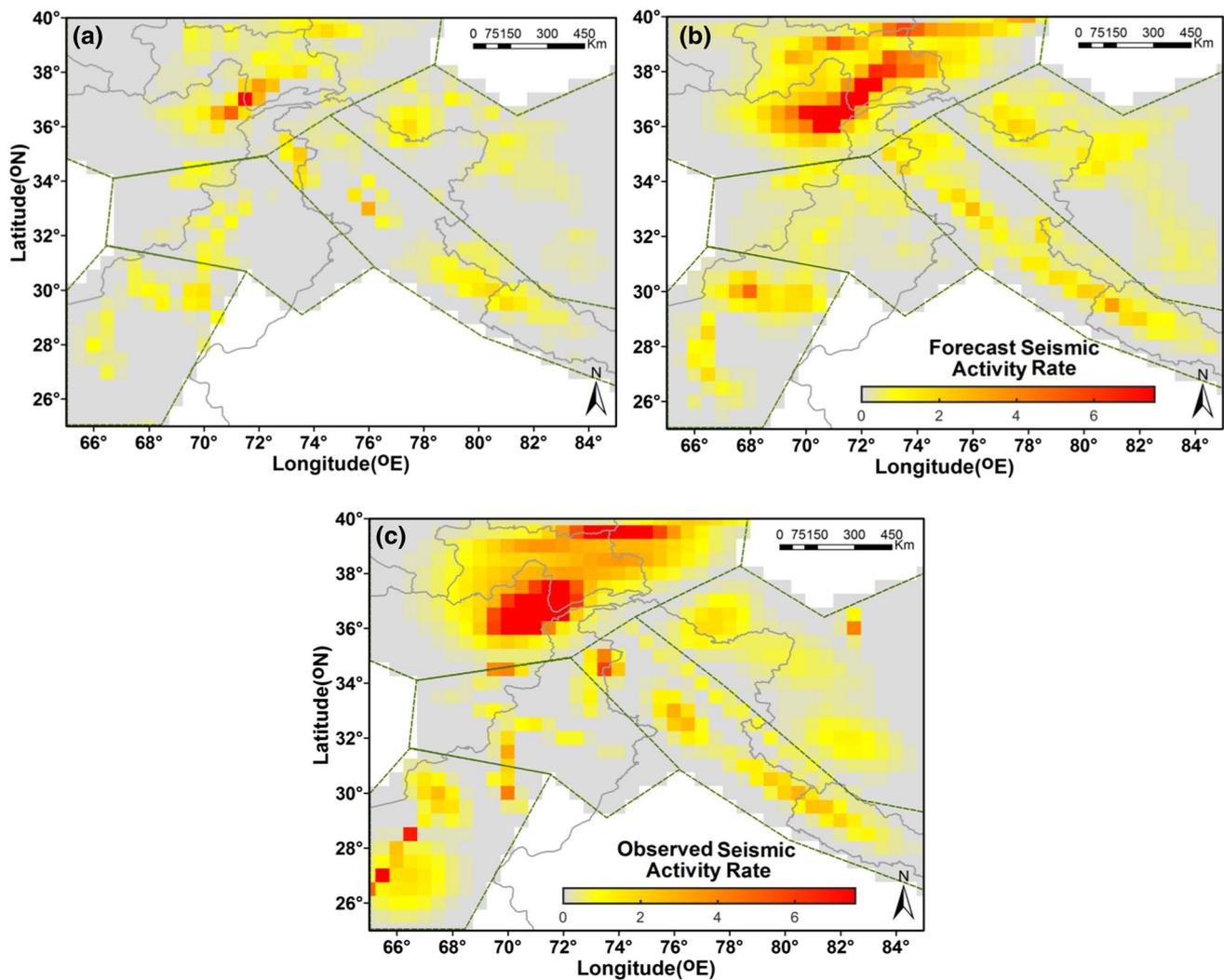


Fig. 4 Forecast seismic activity rate maps reproduced by **a** our model and **b** inhomogeneous Poisson model for magnitude $M_w \geq 4.0$ during the time period 2011–2016 and **c** estimated seismic activity rates from observed earthquakes during the period 2011–2016

over the short time periods. Also, AUC for the predicted seismic activity rates by our forecast model and inhomogeneous Poisson model is presented to compare the performance of these two models during the forecast time period 2011–2016 (Fig. 6b).

Furthermore, a likelihood test, i.e., N test, that is implemented in the CSEP Testing Center software is also performed to investigate the consistency of forecast seismic activity rates with observed seismic activity rates during the forecast period 2011–2016. N test comprises of comparing the number of events between observed and simulated catalogues (Schorlemmer and Gerstenberger 2007; Schorlemmer et al. 2010; Zechar et al. 2010; Rhoades et al. 2011). Figure 7 depicts the data-consistency test performed between observation and simulations at a significant level of 0.1 during the time period 2011–2016. This data-consistency test can be accomplished by investigating the quantile scores (δ

for N test) computed from each perturbed catalogue of the simulated ones (Schorlemmer and Gerstenberger (2007) and Zechar et al. (2010)).

Besides, two-sample t test is also incorporated to check the statistical significance between spatial coordinates of N peak points taken from the forecast and observed seismic activity rate maps during the period 2011–2016. In this hypothetical t test, two random samples of independent observations are investigated for their independency based on the calculated mean, each having an underlying normal distribution (Snedecor and William 1989). On the basis of computed t value, hypothesis is rejected if the calculated value is greater than the tabulated value for equal degree of freedom and level of significance ($\alpha = 0.05$), otherwise is failed to reject. Table 1 depicts the values of t test applied to various peak points of the above-mentioned two maps.

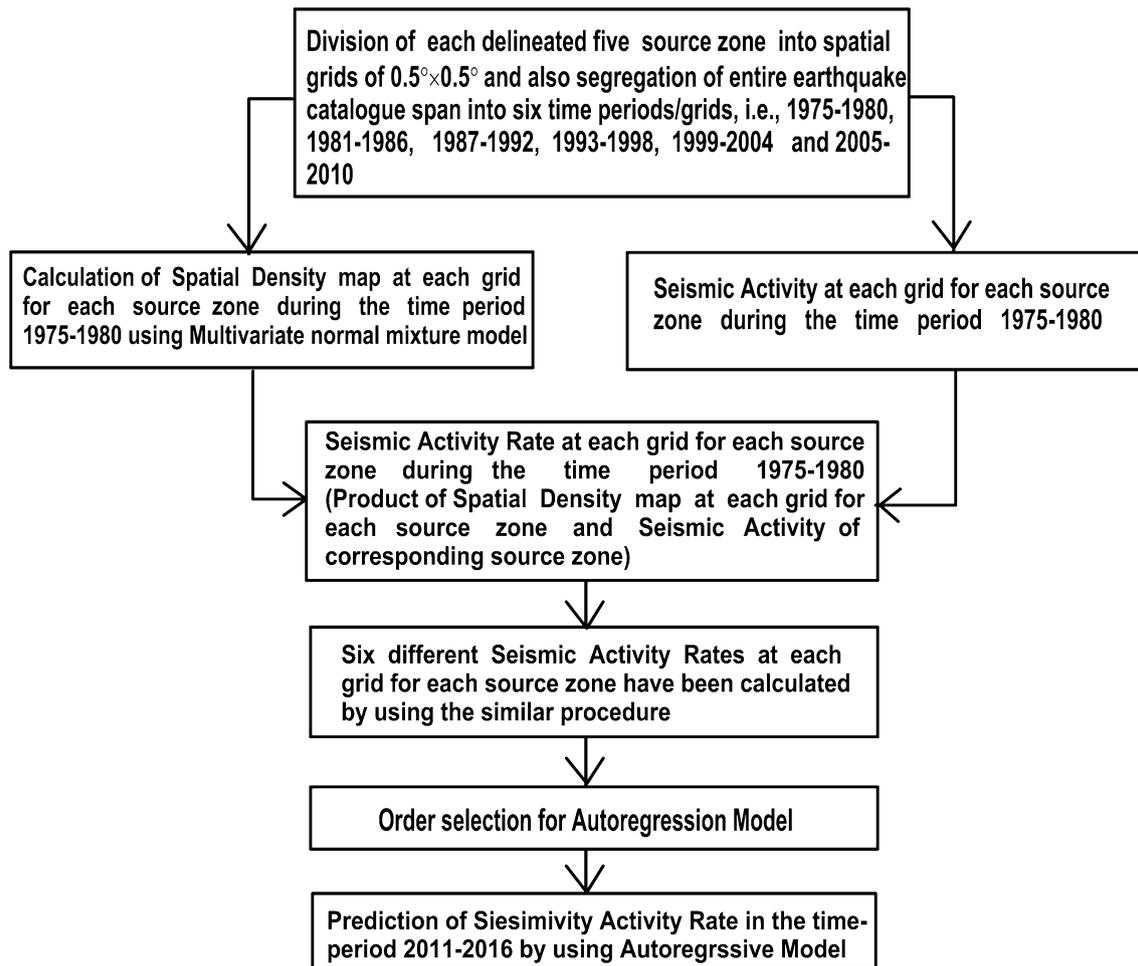


Fig. 5 Flowchart for estimating the forecast seismic activity rate map in our study region

Results and discussion

In this present study, a seismicity-based earthquake forecasting model is devised to forecast seismic activity rates via an autoregressive process and a multivariate normal mixture model in northwest Himalaya during the period 2011–2016. Figure 2a–f depicts the estimated seismic activity rates in each source zone of the study region during the different six time grids. In the seismic activity map shown in Fig. 2a, the values of seismic activity rates are scattered in between 0 and 34.86 throughout the study region. Low values of estimated seismic activity rates during the period 1975–1980 can be easily co-related to high magnitude of completeness, m_t , values. Pockets of high values of seismic activity rates are mostly observed in Hindukush-Pamir Himalaya of Zone 3, even though Zones 1 and 4 also exhibit the same high seismic rate. The high activity rate in Hindukush-Pamir Himalaya may be attributed to increased stress accumulation caused by subducting Indian plate beneath the overlying Himalayan

wedges (Yadav et al. 2012; Chingtham et al. 2014, 2015). Along the major strike-slip Chaman Fault, the transcurrent movement due to plate interaction between the Indian and Eurasian plates caused high activity rate in Zone 1, whereas MCT, MBT and HFT are primarily responsible for high values in Zone 4. Zone 5 is predominantly dominated by moderate values of seismic activity rates. The moderate values of seismic activity rates in this zone can be co-related to the normal and right-lateral earthquakes emanating from Karakoram Fault in the Tibetan Plateau region. However, Zone 2 that represents thin-skinned internal deformation within the Indo-Pakistan Plate exhibits low seismic activity rates (Yadav et al. 2011).

Similarly, the seismic activity rate maps and its corresponding values during the time grids 1981–1986, 1987–1992, 1993–1998, 1999–2004 and 2005–2010 are also depicted in Fig. 2b–d. During each time grid, Zones 3 and 4 exhibit almost similar pattern of seismic activity rates while the remaining source zones show dissimilar pattern of seismic activity rates due to dispersed seismicity (Fig. 2a–f).

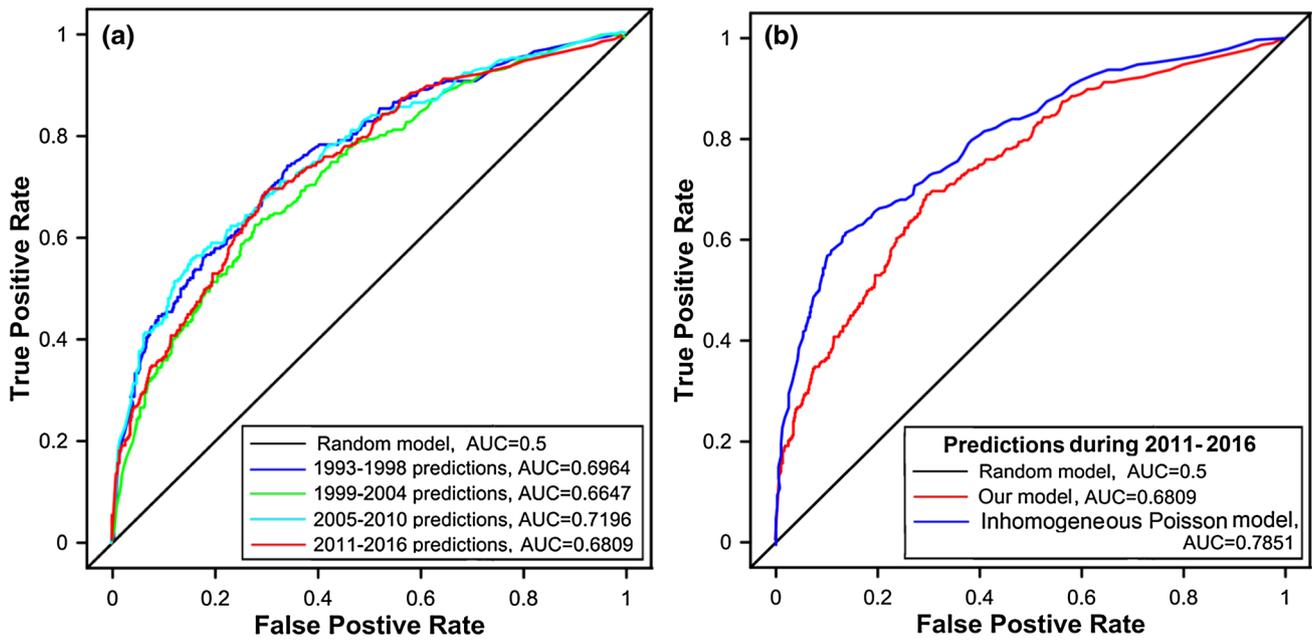


Fig. 6 Receiver operating characteristics (ROC) curves exhibiting the performance of the models. **a** ROC curves for the forecast seismic activity rates produced by our model for various time periods. Blue, green, cyan and red lines give the ROC curves for the time periods,

1993–1998, 1999–2004, 2005–2010 and 2011–2016 respectively, and **b** comparison of ROC curves between our model (red line) and inhomogeneous Poisson model (blue line) during the forecast time period 2011–2016 (color figure online)

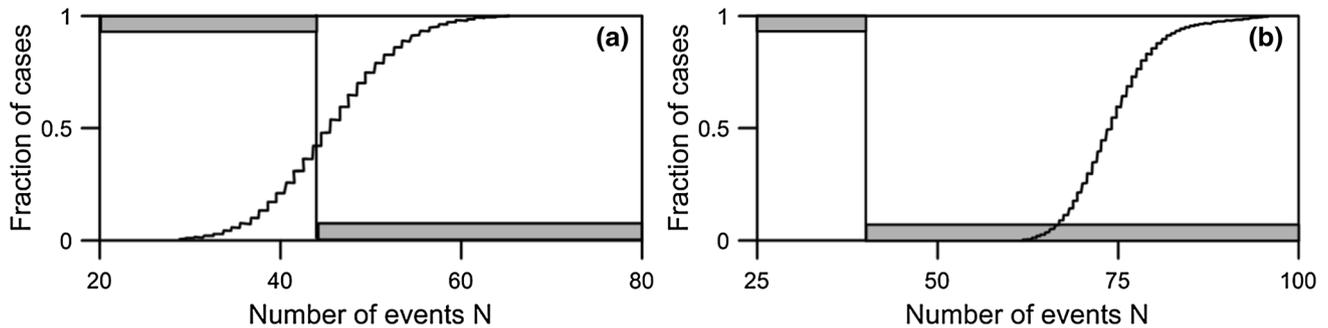


Fig. 7 Results of the data-consistency tests, i.e., N test for **a** our model and **b** inhomogeneous Poisson model during the forecast period 2011–2016. The black curve indicates the cumulative distribu-

tion of numbers computed from simulated events. Rejection bars are marked by the gray patches (color figure online)

Table 1 Results of the t test computed from the coordinates of N grid points having high seismic activity rates in the forecast and observed seismic activity rate maps during the time period 2011–2016

Sl. no.	Number of spatial grid points (N)	Seismic activity estimated from the forecasting model	Seismic activity calculated from the observed events	t value ($\alpha=0.05$)	Decision (0 & 1 to reject and fail to reject the hypothesis)
1	10	Latitude	Latitude	0.275	0
		Longitude	Longitude	0.083	0
2	20	Latitude	Latitude	0.053	0
		Longitude	Longitude	0.111	0
3	30	Latitude	Latitude	0.400	0
		Longitude	Longitude	0.068	0
4	40	Latitude	Latitude	0.129	0
		Longitude	Longitude	0.348	0

Hence, seismic activity rates are calculated for all the spatial grids in each source zone during the time grids 1975–1980, 1981–1986, 1987–1992, 1993–1998, 1999–2004 and 2005–2010.

Figure 4a represents the forecast seismic activity rates drawn along with observed seismic activity rates in the study region during the period 2011–2016. The study region exhibits low to high seismic activity rates within the range 0–7.57. However, our forecast model predicts high values of seismic activity rates in Zones 3 and 4. Low to moderate values of seismic activity rates are mostly predicted by our adopted model in the remaining zones. It is also noteworthy to mention that the observed high seismic activity rates mostly occurred on the spatial grids having moderate to high seismic activity rate predicted by our model. Discrepancy between observed high seismic activity rates and forecast seismic activity rates is also observed in Fig. 4a and c.

Figure 4b represents the estimated forecast seismic activity rates produced by inhomogeneous Poisson model in the study region. The forecast seismic activity rates calculated by inhomogeneous Poisson model are found to be relatively high in comparison with our forecast model (Fig. 4b). These high seismic activity rates are primarily due to averaging the previous values of the same spatial grid in the six time grids. This is so because earthquakes form a stochastically independent sequence of events in time and space, thus exhibiting a memoryless process. On the contrary, our model utilizes the linear dependency of values at the same spatial grid in the six time grids to produce forecast seismic activity rates in the testing period 2011–2016. Furthermore, forecast seismic activity rates not only depend on their most recent value but also on the previous optimal order ($q = 3$) lags. However, performance between our model and inhomogeneous Poisson model can be better examined through ROC curves and likelihood-based N test.

Figure 6a also exhibits ROC curves estimated from Eqs. 4 and 5 for assessing the reliability and skill of forecast model relative to random prediction in the study region. The associated AUC values are found to be 0.6964, 0.6647, 0.7196 and 0.6809 during the time grids 1993–1998, 1999–2004, 2005–2010 and 2011–2016, respectively. This graph also illustrated that the considered model has a better forecasting power than random prediction ($AUC = 0.5$), particularly with larger database. This also implies that the highly fluctuated seismic activity rates during the assumed short time period do not have any adverse effect on the stability of forecasting model. Through visual inspection, we can deduce that the high seismic activity rates are compatible with the observed high seismic activity rates during the period 2011–2016 (Fig. 4a). Also, from the ROC curves depicted in Fig. 6b, AUC for the seismic activity rates produced by

inhomogeneous Poisson model is found to be relatively larger than AUC for the seismic activity rates produced by our model during the forecasting period 2011–2016. This implies that inhomogeneous Poisson model forecasts better locations of observed earthquakes in comparison with our model. However, before stating that the inhomogeneous Poisson model is superior to our model, likelihood-based N test is conducted to this model for comparing the number of simulated and observed events.

In N test, the curve depicting the cumulative distribution of number of simulated events intersects the number of observed events at $\delta = 0.47$ (Fig. 7a). In this test, we also observed that the observed events vary in between 27–63 where the number of events participating in this test is found to be 43. Result of this likelihood-based test concluded that our forecast model is found consistent with observations. On the other hand, the inhomogeneous Poisson model over-predicts the number of events during the forecast period 2011–2016, as observed from the result of N test and hence this inhomogeneous Poisson model can be rejected at the given significance level (Fig. 7b).

Moreover, results of the t test obtained from the spatial coordinates of various N peaks of forecast seismic activity rates and observed earthquakes are depicted in Table 1. In the first test, ten spatial coordinates for both the forecast and observed seismic activity rates have been selected and t test was performed in between the latitudes and longitudes of the forecast and observed seismic activity rates. The results give t value more than the significance level ($\alpha = 0.05$), and thus the null hypothesis will be rejected for these ten spatial grid points. Similarly, t values for these 20, 30 and 40 spatial grid points have been conducted and found to be greater than significance level ($\alpha = 0.05$), thereby rejecting the null hypothesis for these 20, 30 and 40 spatial grid points. Magnitude distributions of the forecast are not formulated in this model, and therefore, we have omitted the discussion of magnitude distributions in the revised manuscript. Though this lack of magnitude component provides a less detailed (and effectively weaker) evaluation of the forecasts, our model predicts better number or occurrence rate and spatial coordinates of observed events, as confirmed from N and t tests. Therefore, it can be summarized that the results obtained from our forecast model are promising in a certain well-defined sense. Hence, this statistical model describing earthquake clustering can be incorporated for real-time probabilistic forecasting of earthquake in timely and effective manner by updating the catalogue on routine basis. Moreover, this modeling approach will substantially provide a useful contribution in decision making and public information for saving millions of lives living in and around the study region.

Conclusions

The major advantage of this modified MARFS model lies on the underlying assumptions of non-stationarity and its density estimation technique. The forecast model, therefore, unambiguously captures both changes in seismic activity rates at $0.5^\circ \times 0.5^\circ$ bins by exploiting spatial and temporal distributions of previous earthquakes in northwest Himalaya through multivariate mixture model and autoregression technique. Though this model successfully predicts better number or occurrence rate of observed events, inhomogeneous Poisson model outperforms this model in forecasting better locations of observed earthquakes in northwest Himalaya. Moreover, time dependency adjustment factor for large earthquakes can also be incorporated in this model by assuming the hypothesis that the probabilities of occurrence of large earthquake increase as time elapses since the previous earthquake. Taking into account of its performance and reliability, this kind of forecast modeling approach is expected to provide useful information based on inferred seismic activity rates for real-time earthquake hazard assessment in northwest Himalaya.

Acknowledgements The authors are thankful to Dr. V. K. Gahalaut, director, National Center for Seismology (NCS), for constant support and encouragement. We are grateful to anonymous reviewers for their detailed comments that tremendously helped to improve the manuscript. Some of the figures were plotted with the Generic Mapping Tools (Wessel and Smith 1998).

References

- Akaike H (1974) New look at statistical-model identification. *IEEE Trans Autom Control* 19:716–723
- Ambraseys NN, Bilham R (2003) Earthquakes and crustal deformation in northern Baluchistan 1892–2001. *Bull Seismol Soc Am* 93(4):1573–1600
- Armbruster J, Seeber L, Jacob KH (1978) The northern termination of the Himalayan mountain front: active tectonics from microearthquakes. *J Geophys Res* 83(B1):269–282
- Arora BR, Gahalaut VK, Kumar N (2012) Structural control on along-strike variation in the seismicity of the Northwest Himalaya. *J Asian Earth Sci* 57:15–24
- Bernard M, Shen-Tu B, Holt WE, Davis DM (2000) Kinematics of active deformation in the Sulaiman Lobe and Range Pakistan. *J Geophys Res* 105:13253
- Bhatia SC, Kumar MR, Gupta HK (1999) A probabilistic hazard map of India and adjoining regions. *Ann Geofis* 42:1153–1164
- Bilham R, Gaur VK (2000) Geodetic contributions to the study of seismotectonics in India. *Curr Sci* 79:1259–1269
- Bilham R, Larson K, Freymueller J, Idylhim P (1997) GPS measurements of present-day convergence across the Nepal Himalaya. *Nature* 386:61–64
- Bird P, Zhen L (2007) Seismic hazard inferred from tectonics: California. *Seismol Res Lett* 78(1):37–48
- BIS (2002) Is 1893 (part 1)—2002: Indian standard criteria for earthquake resistant design of structures, part 1—general provisions and buildings. Bur Indian Stand, NewDelhi
- Bormann P, Liu R, Ren X, Gutdeutsch R, Kaiser D, Castellaro S (2007) Chinese national network magnitudes, their relation to NEIC magnitudes, and recommendations for new IASPEI magnitude standards. *Bull Seismol Soc Am* 97:114–127
- Burtman V, Molnar P (1993) Geological and geophysical evidence for deep subduction of continental crust beneath the Pamir. *Geol Soc Am Spec Paper* 281. <https://doi.org/10.1130/SPE281-p1>
- Chandra U (1978) Seismicity, earthquake mechanisms and tectonics along the Himalayan mountain range and vicinity. *Phys Earth Planet Inter* 16:109–131
- Chatelain JL, Roeker SW, Hatzfeld D, Molnar P (1980) Micro-earthquake seismicity and fault plane solutions in the Hindu-kush region and their tectonic implications. *J Geophys Res* 85:1365–1387
- Chatfield C (2004) The analysis of time series. Chapman and Hall, Boca Raton, Florida
- Chingtham P, Chopra S, Baskoutas I, Bansal BK (2014) An assessment of seismicity parameters in northwest Himalaya and adjoining regions. *Nat Hazards* 71:1599–1616
- Chingtham P, Yadav RBS, Chopra S, Yadav AK, Gupta AK, Roy PNS (2015) Time-dependent seismicity analysis in the northwest Himalaya and its adjoining regions. *Nat Hazards* 80(30):1783–1800. <https://doi.org/10.1007/s11069-015-2031-0>
- Chingtham P, Sharma B, Chopra S, Roy PNS (2016) Statistical analysis of aftershock sequences related with two major Nepal earthquakes: April 25, 2015, M_w 7.8, and May 12, 2015, M_w 7.2. *Ann Geophys* 59(5):S0540. <https://doi.org/10.4401/ag-7025>
- Chingtham P, Prajapati SK, Gahalaut VK, Chopra S, Roy PNS (2017) Forecasting seismicity rate in the north-west Himalaya using rate and state dependent friction law. *Geomat Nat Haz Risk* 8(2):1643–1661. <https://doi.org/10.1080/19475705.2017.1369168>
- Console R, Murru M (2001) A simple and testable model for earthquake clustering. *J Geophys Res* 106:8699–8711
- Console R, Murru M, Catalli F (2006) Physical and stochastic models of earthquake clustering. *Tectonophysics* 417(1–2):141–153
- DeMets C, Gordon RG, Argus DF, Stein S (1994) Effect of recent revisions to the geomagnetic reversal timescale on estimates of current plate motions. *Geophys Res Lett* 21:2191–2194
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via algorithm. *J Roy Stat Soc Ser B (Stat Methodol)* 1:1–38
- Everitt BS (1993) Cluster analysis. Edward Arnold, London
- Fan G, Ni JF, Wallace TC (1994) Active tectonics of the Pamir and the Karakoram. *J Geophys Res* 99:7131–7160
- Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognit Lett* 27:861–874
- Field EH (2007) Overview of the working group for the development of regional earthquake likelihood models (RELM). *Seismol Res Lett* 78(1):7–16
- Gansser A (1964) Geology of the Himalayas. Inter-Science, London
- Gardner JK, Knopoff L (1974) Is the sequence of earthquakes in southern California with aftershocks removed, Poissonian? *Bull Seismol Soc Am* 64:1363–1367
- Gerstenberger MC, Rhoades DA (2010) New Zealand earthquake forecast testing centre. *Pure appl Geophys* 167:877–892. <https://doi.org/10.1007/s00024-010-0082-4>
- Gupta HK, Gahalaut VK (2014) Seismotectonics and large earthquake generation in the Himalayan region. *Gondwana Res* 25(1):204–213
- Helmstetter A, Kagan YY, Jackson DD (2007) High-resolution time-independent grid-based forecast for $M > 5$ earthquakes in California. *Seismol Res Lett* 78:78–86

- International Commission on Earthquake Forecasting for Civil Protection (2009) Operational earthquake forecasting: state of knowledge and guidelines for utilization released by the Dipartimento Della Protezione Civile, Rome, Italy, 2 Oct 2009
- Jackson DD, Kagan YY (1999) Testable earthquake forecasts for 1999. *Seismol Res Lett* 70:393–403
- Kagan YY, Rong YF, Jackson DD (2003) Probabilistic forecasting of seismicity. In: Mulargia F, Geller RJ (eds) *Earthquake science and seismic risk reduction*. Kluwer Academic Publishing, Dordrecht, pp 185–200
- Knopoff L (2000) The magnitude distribution of declustered earthquakes in southern California. *Proc Natl Acad Sci USA* 97(22):11880–11884. <https://doi.org/10.1073/pnas.190241297>
- Koulakov I, Sobolev S (2006) A tomographic image of Indian lithosphere break-off beneath the Pamir–Hindukush region. *Geophys J Int* 164:425–440
- Loannis B, George P, Chingtham P, Bansal BK (2011) Temporal variation of seismic parameters in the western part of the India-Eurasia plate collision zone. *Res Geophys* 1(e3):8–12
- Lyon-Caen H, Molnar P (1983) Constraints on the structure of the Himalaya from an analysis of gravity anomalies and a flexural model of the lithosphere. *J Geophys Res* 88:8171–8192
- Marzocchi W, Lombardi AM (2009) Real-time forecasting following a damaging earthquake. *Geophys Res Lett* 36:L21302. <https://doi.org/10.1029/2009GL040233>
- Marzocchi W, Zechar JD (2011) Earthquake forecasting and earthquake prediction: different approaches for obtaining the best model. *Seismol Res Lett* 82:442–448
- Marzocchi W, Amato A, Akinci A, Chiarabba C, Lombardi AM, Pantosti D, Boschi E (2012) A ten-year earthquake occurrence model for Italy. *Bull Seismol Soc Am* 102:1195–1213
- McLachlan G, Ng S (2009) The EM algorithm. In: Wu X, Kumar V (eds) *The top-ten algorithms in data mining*. Chapman and Hall/CRC, Boca Raton, pp 93–115
- Murru M, Console R, Falcone G (2009) Real time earthquake forecasting in Italy. *Tectonophysics* 470:214–223
- Nishenko SP, Singh SK (1987) Conditional probabilities for the recurrence of large and great interplate earthquakes along the Mexican subduction zone *Bull Seismol Soc Am* 77:2094–2114
- Ogata Y (1998) Space-time point-process models for earthquake occurrences. *Ann Inst Stat Math* 50(2):379–402
- Philip G, Suresh N, Bhakuni SS (2014) Active tectonics in the northwestern outer Himalaya: evidence of large-magnitude palaeoearthquakes in Pinjaur Dun and the frontal Himalaya. *Curr Sci* 106:211–222
- Rhoades DA (2007) Application of the EEPAS model to forecasting earthquakes of moderate magnitude in southern California. *Seism Res Lett* 78(1):110–115
- Rhoades DA, Evison FF (2004) Long-range earthquake forecasting with every earthquake a precursor according to scale. *Pure appl Geophys* 161:47–71
- Rhoades DA, Evison FF (2005) Test of the EEPAS forecasting model on the Japan earthquake catalogue. *Pure appl Geophys* 162:1271–1290
- Rhoades DA, Evison FF (2006) The EEPAS forecasting model and the probability of moderate-to-large earthquakes in central Japan. *Tectonophysics* 417:119–130
- Rhoades DA, Schorlemmer D, Gerstenberger MC, Christophersen A, Zechar JD, Imoto M (2011) Efficient testing of earthquake forecasting models. *Acta Geophys* 59:728–747. <https://doi.org/10.2478/s11600-011-0013-5>
- Robinson D, Dhu T, Schneider J (2006) SUA: a computer program to compute regolith site-response and estimate uncertainty for probabilistic seismic hazard analyses. *Comput Geosci* 32:109–123
- Rout MM, Das J, Kamal Das R (2015) Probabilistic seismic hazard assessment of NW and central Himalayas and the adjoining region. *J Earth Syst Sci* 124(3):577–586
- Scholz CH, Small C (1997) The effect of seamount subduction on seismic coupling. *Geology* 25:487–490
- Schorlemmer D, Gerstenberger MC (2007) RELM testing center. *Seismol Res Lett* 78:30–36
- Schorlemmer D, Wiemer S, Wyss M (2005) Variations in earthquake-size distribution across different stress regimes. *Nature* 437(22):539–542
- Schorlemmer D, Christophersen A, Rovida A, Mele F, Stucchi M, Marzocchi W (2010) Setting up an earthquake forecast experiment in Italy. *Ann Geophys Italy* 53:1–9
- Seeber L, Armbruster JG (1984) Some elements of continental subduction along the Himalayan front. *Tectonophysics* 105:263–278
- Shen ZK, Jackson DD, Kagan YY (2007) Implications of geodetic strain rate for future earthquakes, with a five-year forecast of M 5 earthquakes in southern California. *Seismol Res Lett* 78:116–120
- Smyth C, Mori J (2011) Statistical models for temporal variations of seismicity parameters to forecast seismicity rates in Japan. *Earth, Planets Space* 63:231–238
- Snedecor GW, William CG (1989) *Statistical methods*, 8th edn. Iowa State University Press, Ames
- Thingbaijam KKS, Nath SK, Yadav A (2008) Recent seismicity in northeast India and its adjoining region. *J Seismol* 12:107–123
- Thingbaijam KKS, Chingtham P, Nath SK (2009) Seismicity in the north-west frontier province at the Indian-Eurasian Plate Convergence. *Seismol Res Lett* 80:599–608
- Toda S, Enescu B (2011) Rate/state Coulomb stress transfer model for the CSEP Japan seismicity forecast. *Earth Planets Space* 63:171–185
- Valdiya KS (2003) Reactivation of Himalayan frontal fault: implication. *Curr Sci* 85(7):1031–1040
- Wessel P, Smith WHF (1998) New improved version of generic mapping tools released. *EOS Trans Am Geophys Union* 79:579
- Yadav RBS, Bayrak Y, Tripathi JN, Chopra S, Singh AP, Bayrak E (2011) A probabilistic assessment of earthquake hazard parameters in NW Himalaya and the adjoining regions. *Pure appl Geophys* 169:1619–1639
- Yadav RBS, Tsapanos TM, Bayrak Y, Koravos GCH (2012) Probabilistic appraisal of earthquake hazard parameters deduced from a Bayesian approach in the northwest frontier of the Himalayas. *Pure appl Geophys* 170:283–295
- Zechar JD, Gerstenberger MC, Rhoades DA (2010) A Likelihood based tests for evaluating space-rate magnitude earthquake forecasts. *Bull Seismol Soc Am* 100:1184–1195



Developing an XGBoost model to predict blast-induced peak particle velocity in an open-pit mine: a case study

Hoang Nguyen^{1,2} · Xuan-Nam Bui^{1,2} · Hoang-Bac Bui^{3,4} · Dao Trong Cuong⁵

Received: 26 February 2018 / Accepted: 19 February 2019 / Published online: 25 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

Ground vibration is one of the most undesirable effects induced by blasting operations in open-pit mines, and it can cause damage to surrounding structures. Therefore, predicting ground vibration is important to reduce the environmental effects of mine blasting. In this study, an eXtreme gradient boosting (XGBoost) model was developed to predict peak particle velocity (PPV) induced by blasting in Deo Nai open-pit coal mine in Vietnam. Three models, namely, support vector machine (SVM), random forest (RF), and *k*-nearest neighbor (KNN), were also applied for comparison with XGBoost. To employ these models, 146 datasets from 146 blasting events in Deo Nai mine were used. Performance of the predictive models was evaluated using root-mean-squared error (RMSE) and coefficient of determination (R^2). The results indicated that the developed XGBoost model with RMSE = 1.554, $R^2 = 0.955$ on training datasets, and RMSE = 1.742, $R^2 = 0.952$ on testing datasets exhibited higher performance than the SVM, RF, and KNN models. Thus, XGBoost is a robust algorithm for building a PPV predictive model. The proposed algorithm can be applied to other open-pit coal mines with conditions similar to those in Deo Nai.

Keywords eXtreme gradient boosting · XGBoost · Ground vibration · Peak particle velocity

Introduction

Blasting is one of the highly effective methods in open-cast mining when used to move rocks and overburden. However, only 20–30% of explosion energy is used for rock fragmentation (Chen and Huang 2001; Coursen 1995; Gad et al. 2005; Gao et al. 2018e). The remaining energy is wasted

and generates undesirable effects such as ground vibration, air-blast overpressure (AOp), fly rock, and back break (Ak and Konuk 2008; Bui et al. 2019; Chen and Huang 2001; Ghasemi et al. 2016; Hajihassani et al. 2014; Hasanipanah et al. 2017a; Monjezi et al. 2011a; Nguyen and Bui 2018b; Nguyen et al. 2018a). Among these effects, PPV is one of the most undesirable effects because it may be harmful to humans and structures. To reduce the adverse effects of blasting operations, many researchers have proposed empirical equations to predict PPV; among these researchers are the United States Bureau of Mines (Duvall and Fogelson 1962; Ambraseys and Hendron 1968; Davies et al. 1964; Standard 1973; Roy 1991). However, influencing parameters are numerous, and the relationship among them is complicated. Thus, the empirical methods may not be entirely suitable for predicting PPV in open-cast mines (Ghasemi et al. 2013; Hajihassani et al. 2015; Hasanipanah et al. 2015; Monjezi et al. 2011b, 2013; Nguyen and Bui 2018a; Nguyen et al. 2018b, 2019; Saadat et al. 2014).

Nowadays, artificial intelligence (AI) is well known as a robust tool for solving the real-life problems (Alnaqi et al. 2019; Gao et al. 2018a, c; Moayedi and Nazir 2018; Moayedi et al. 2019; Moayedi and Rezaei 2017). Many researchers

✉ Hoang Nguyen
nguyenhoang@humg.edu.vn

¹ Department of Surface Mining, Mining Faculty, Hanoi University of Mining and Geology, 18 Vien St., Duc Thang Ward, Bac Tu Liem Dist., Hanoi, Vietnam

² Center for Mining, Electro-Mechanical Research, Hanoi University of Mining and Geology, 18 Vien St., Duc Thang Ward, Bac Tu Liem Dist., Hanoi, Vietnam

³ Faculty of Geosciences and Geoenvironment, Hanoi University of Mining and Geology, 18 Vien St., Duc Thang Ward, Bac Tu Liem Dist., Hanoi, Vietnam

⁴ Center for Excellence in Analysis and Experiment, Hanoi University of Mining and Geology, 18 Vien St., Duc Thang Ward, Bac Tu Liem Dist., Hanoi, Vietnam

⁵ Ministry of Industry and Trade, Hanoi, Vietnam

have studied and applied AI in predicting blast-induced issues, especially blast-produced PPV. Longjun et al. (2011) applied two benchmark algorithms for estimating PPV, including support vector machine (SVM) and random forest (RF); two other parameters with 93 explosions were used as training datasets, and 15 observations among 93 views were selected as testing datasets. Their study indicated that the SVM and RF models performed well in estimating blast-induced PPV. The SVM model was introduced as a superior model in their study. Hasanipanah et al. (2017b) also developed a Classification and regression tree (CART) model to predict PPV at Miduk copper mine (Iran) using 86 blasting events. Multiple regression (MR) and various empirical techniques were also considered to predict PPV and compared with the CART model. As a result, the CART model was exhibited better performance than the other models with $RMSE=0.17$ and $R^2=0.95$ in their study. In another work, Chandar et al. (2017) estimated blast-induced PPV using ANN model; 168 blasting operations were collected in dolomite, coal mine, and limestone (Malaysia) for their aim. The results indicated that the ANN model, with $R^2=0.878$ for the three mines, is the best among the approaches used in their study. Metaheuristics algorithm was also considered and used to predict PPV by Faradonbeh and Monjezi (2017), i.e., gene expression programming (GEP); 115 blasting operations were used for their study. Accordingly, a formula based on the GEP was developed to estimate PPV as the first step in their study. Then, it was compared with several nonlinear and general equation models as the second step as well. Their results designated that the GEP model was better than the other models in forecasting blast-induced PPV. Similar works can be found at those references (Faradonbeh et al. 2016; Hasanipanah et al. 2017c; Sheykhi et al. 2018; Taheri et al. 2017).

In this study, an XGBoost model was developed to predict blast-induced PPV in Deo Nai open-pit coal mine (Vietnam). Three other models were also produced, including SVM, RF, and KNN for comparison with the constructed XGBoost model.

This paper is organized as follows. Section “two” describes the site study and the data used. Section “three” provides an overview of the algorithms used in this study. Section “four” reports the results and discussion. Section “five” shows the validation of the constructed models. Finally, Section “six” presents our conclusions.

Site study and data used

Study area

With the total area up to $\sim 6 \text{ Km}^2$, the Deo Nai open-pit coal mine was a large open-cast coal mine in Vietnam (Fig. 1).

It is located in Quang Ninh province, Vietnam, with the proven reserve is 42.5 Mt, and productivity is 2.5 Mt/year. The study area has a complex geological structure, includes many different phases and faults. Conglomerate, siltstone, sandstone, claystone, and argillic rock were included in the overburden of this mine (Vinacomin 2015). The hardness of these rocks (f) in the range of 11–12 according to Protodiakonov’s classification (Protodiakonov et al. 1964); specific weight (γ) in the range of 2.62–2.65 t/m^3 . Therefore, blasting operations for rock fragmentation in this mine is a high-performance method.

However, the Deo Nai open-pit coal mine is located near residential areas (Fig. 1), which have a distance of approximately 400 m from the blasting sites. Moreover, the capacity of burden must explode significantly in a blast of up to more 20 tons, and the adverse effects (especially PPV) of the blasting operation to the surrounding environment are substantial. Thus, we have selected this area as a case study to consider and predict PPV caused by blasting operations with the aim of controlling the undesirable effects on the environment and residential areas.

Data collection

To conduct this study, 146 blasting events were collected with nine parameters, such as the number of borehole rows per blast (N), charge per delay (Q), powder factor (q), length of stemming (T), burden (B), monitoring distance (D), spacing (S), bench height (H), and time interval between blasts (Δt) which were considered as nine input parameters to predict the outcome, i.e., PPV. Table 1 shows a brief of the datasets used in this study.

For monitoring PPV, the Blastmate III instrument (Instan-tel, Canada) was used with the specifications that are shown in Table 2. In this study, PPV values were recorded in the range of 2.140 to 33.600 mm/s. A GPS device was used to determine D . The remaining parameters were extracted from blast patterns.

Preview of XGBoost, SVM, RF, and KNN

eXtreme gradient boosting (XGBoost)

XGBoost is an improved algorithm based on the gradient boosting decision proposed by (Friedman et al. 2000, 2001; Friedman 2001, 2002). XGBoost, which was created and developed by Chen and He (2015), can construct boosted trees efficiently, operate in parallel, and solve both classification and regression problems. The core of the algorithm is the optimization of the value of the objective function. It implements machine learning algorithms under the gradient boosting framework. XGBoost can

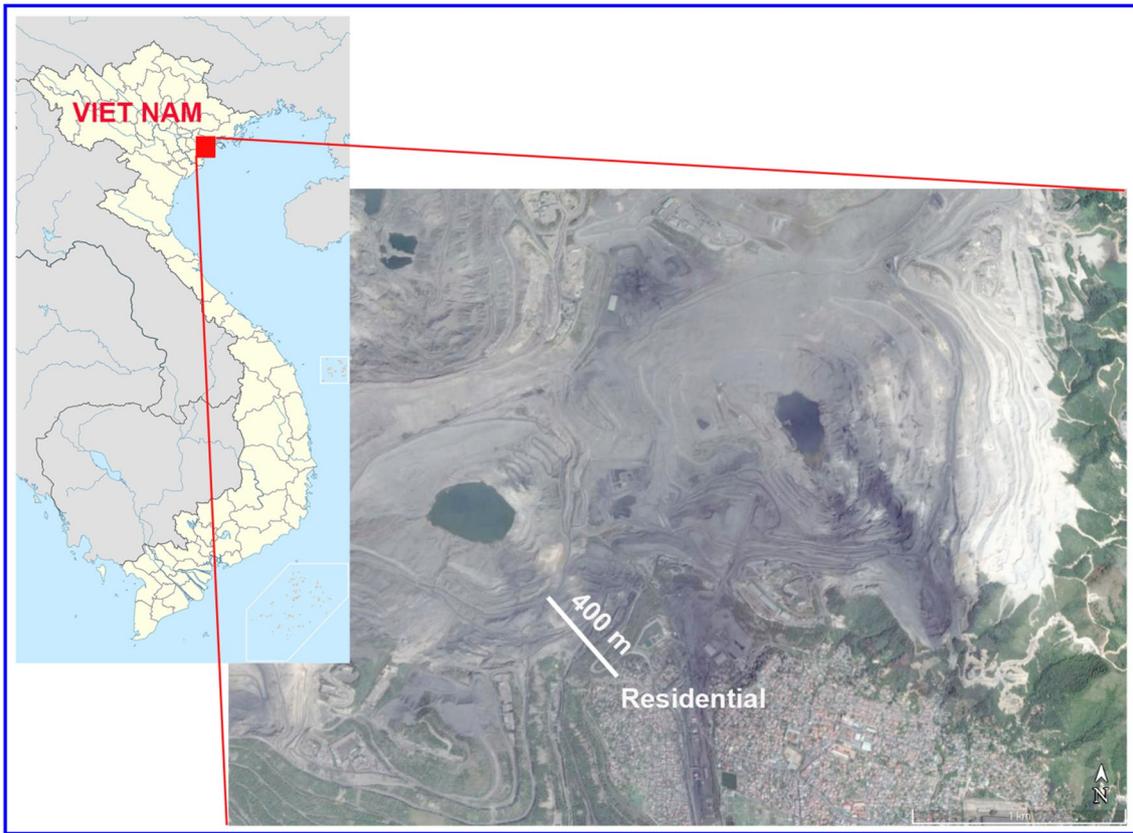


Fig. 1 Location of the study area

Table 1 Blasting events recorded for this study

Q (Kg)	H (m)	B (m)	S (m)	T (m)
Minimum: 3007	Minimum: 13.00	Minimum: 7.500	Minimum: 7.40	Minimum: 6.20
Mean: 13,324	Mean: 14.38	Mean: 8.071	Mean: 7.82	Mean: 6.87
Maximum: 24,171	Maximum: 16.0	Maximum: 8.50	Maximum: 8.20	Maximum: 7.50
q (kg/m ³)	N (borehole)	D (m)	Δt (ms)	PPV (mm/s)
Minimum: 0.3500	Minimum: 2.000	Minimum: 180.0	Minimum: 17.00	Minimum: 2.140
Mean: 0.4184	Mean: 3.525	Mean: 465.8	Mean: 27.64	Mean: 13.322
Maximum: 0.480	Maximum: 5.00	Maximum: 726	Maximum: 42.00	Maximum: 33.600

Table 2 Basic parameters of the PPV monitoring instrument Source: <http://www.instantel.com>

Features	Parameters
Range	0.127–254 mm/s
Accuracy	± 0.5 mm/s or 5% (0.02 in/s)
Transducer density	2.13 g/cc (133 lbs/ft ³)

solve many data science problems in a fast and accurate way with parallel tree boosting such as gradient boosting decision tree and gradient boosting machine.

An objective function usually consists of two parts (training loss and regularization):

$$\text{Obj}(\theta) = L(\theta) + \Omega(\theta), \tag{1}$$

where L is the training loss function and Ω is the regularization term. The training loss is used to measure the model performance on training data. The regularization term aims to control the complexity of the model such as overfitting (Gao et al. 2018d). Various ways are conducted to define complexity. However, the complexity of each tree is often computed as the following equation:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2, \tag{2}$$

where T is the number of leaves and ω is the vector of scores on leaves.

The structure score of XGBoost is the objective function defined as follows:

$$\text{Obj} = \sum_{j=1}^T \left[G_j \omega_j + \frac{1}{2} (H_j + \lambda) \omega_j^2 \right] + \gamma T, \tag{3}$$

where ω_j are independent of each other. The form $G_j \omega_j + \frac{1}{2} (H_j + \lambda) \omega_j^2$ is quadratic and the best ω_j for a given structure $q(x)$.

Support vector machines (SVM)

SVM is a machine learning method based on statistical theory and developed by (Cortes and Vapnik 1995). This method continues to be applied to high-performing algorithms with slight tuning. Similar to CART, SVM can also be used to solve classification and regression problems. According to Cortes and Vapnik (1995), SVM was used for classification analysis. SVR, a version of SVM for regression analysis, was proposed by Drucker et al. (1997).

In SVM, fitting data $\{x_i, y_i\}$, $(i = 1, 2, \dots, n)$, $x_i \in R^n$, $y_i \in R$ with a function $f(x) = w \cdot x + b$ is a problem. Thus, according to SVM theory, the fitting problem function is expressed as follows:

$$f(x) = w \cdot x + b = \sum_{i=1}^k (a_i - a_i^*) K(x, x_i) + b \tag{4}$$

where a_i , a_i^* , and b are obtained by solving subsequent second optimization problems. Usually, a small fraction of a_i , a_i^* is not zero; this fraction is called support vector.

Max:

$$w(a, a^*) = -\frac{1}{2} \sum_{i,j=1}^k (a_i - a_i^*)(a_j - a_j^*) K(x_i, x_j) + \sum_{i=1}^k y_i (a_i - a_i^*) - \epsilon \sum_{i=1}^k (a_i + a_i^*), \tag{5}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^k (a_i - a_i^*) = 0 \\ 0 \leq a_i, a_i^* \leq C, (i = 1, 2, \dots, k) \end{cases}, \tag{6}$$

where C is a penalty factor that shows the penalty degree to samples of excessive error ϵ ; $K(x_i, x_j)$ is kernel function,

which solves calculation problems of high dimension skillfully by introducing kernel functions. These functions are mainly of the following types:

1. Linear kernel

$$K(x, y) = x \cdot y, \tag{7}$$

2. Polynomial kernel

$$K(x, y) = [(x \cdot y) + 1]^d; \quad d = (1, 2, \dots), \tag{8}$$

3. Radial original kernel function

$$K(x, y) = \exp \left[\frac{-\|x - y\|^2}{\sigma^2} \right], \tag{9}$$

4. Two-layer neural kernel

$$K(x, y) = \tanh [a(x \cdot y) - \delta]. \tag{10}$$

In this study, the SVM method with a polynomial kernel function is used to develop the SVM model for anticipating PPV.

Random forest (RF)

RF is one of the decision tree algorithms and introduced by Breiman (2001) for the first time. It is well known as a robust non-parametric statistical technique for both regression and classification problems. On the other hand, RF was introduced as an ensemble method based on the results from different trees to achieve predictive accuracy (Vigneau et al. 2018). For each new observation, RF combines the predicted values from the individual tree in the forest to give the best result. In the forest, each tree roles as a voter for the final decision of the RF (Gao et al. 2018b). The core of the RF model for regression can be described as three steps follow:

Step 1 Create bootstrap samples as the number of the tree in the forest (n_{tree}) based on the dataset.

Step 2 Develop an unpruned regression tree for each bootstrap sample by random sampling of the predictors (m_{try}). Among those variables, select the best split.

Step 3 Predict new observation by ensemble the predicted values of the trees (n_{tree}). For the regression problem as well as predicting blast-induced PPV, the average value of the predicted values by the individual tree in the forest used.

Based on the training dataset, an estimate of the error rate can be obtained by the following:

- At each bootstrap iteration, predict the data not in the bootstrap sample using the tree grown with the bootstrap sample, called “out-of-bag” (OOB).
- Aggregate the OOB predictions and calculate the error rate.

The implementation of the RF algorithm for predicting blast-induced PPV in this study is shown in Fig. 2. More details of the RF algorithm can be found at those references (Breiman 2001; Bui et al. 2019; Nguyen and Bui 2018b).

k-nearest neighbor (KNN)

KNN is known as a favorite technique for solving regression and classification problems in machine learning and introduced by Altman (1992). Based on the closest neighbors (*k* neighbors), the KNN algorithm determines the testing point and classify them. On the other hand, the KNN algorithm does not learn anything from training data. It only remembers the weights of neighbors in the functional space. When it comes to forecasting a new observation, it searches similar results and calculates the distance to those neighbors. Therefore, KNN is classified as “lazy learning” algorithms (Fig. 3).

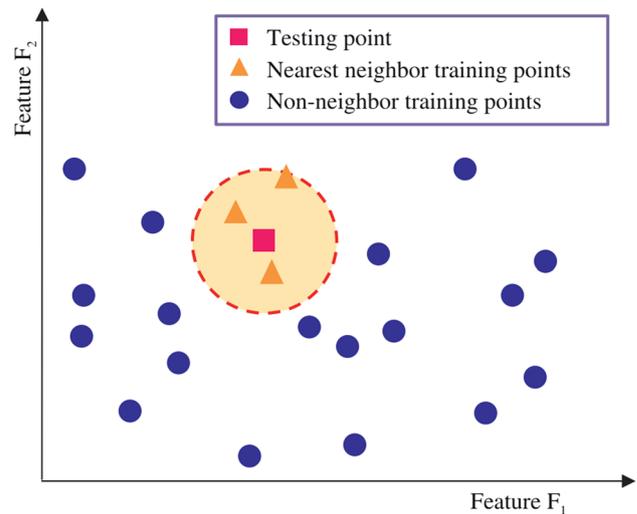


Fig. 3 Illustration of KNN algorithm for two-dimensional feature space (Hu et al. 2014)

For regression problems as well as predicting blast-induced PPV, the KNN algorithm uses a weighted average of the *k*-nearest neighbors, computed by their distance inversely. The KNN for regression can be worked as four steps follow:

Step 1 Determine the distance from the query sample to the labeled samples.

$$d(x_{tr}, x_t) = \sqrt{\sum_{n=1}^N w_n (x_{tr,n} - x_{t,n})^2} \tag{11}$$

where *N* is the number of features; *x_{tr,n}* and *x_{t,n}* denote the *n*th feature values of the training (*x_{tr}*) and testing (*x_t*) points, respectively; *w_n* is the weight of the *n*th feature and lies interval [0,1].

Step 2 Order the labeled examples by increasing distance.
Step 3 Based on RMSE (Eq. 12), define the optimal number of neighbors. Cross-validation can be used for this task.

Step 4 Calculate the average distance inversely with *k*-nearest neighbors.

Results and discussion

In this study, the datasets are divided into two sections: training and testing. Of the total datasets, 80% (approximately 118 blasting events) are used for the training process, and the rest (28 observations) are used for the testing process. The training dataset is used for the development of the mentioned

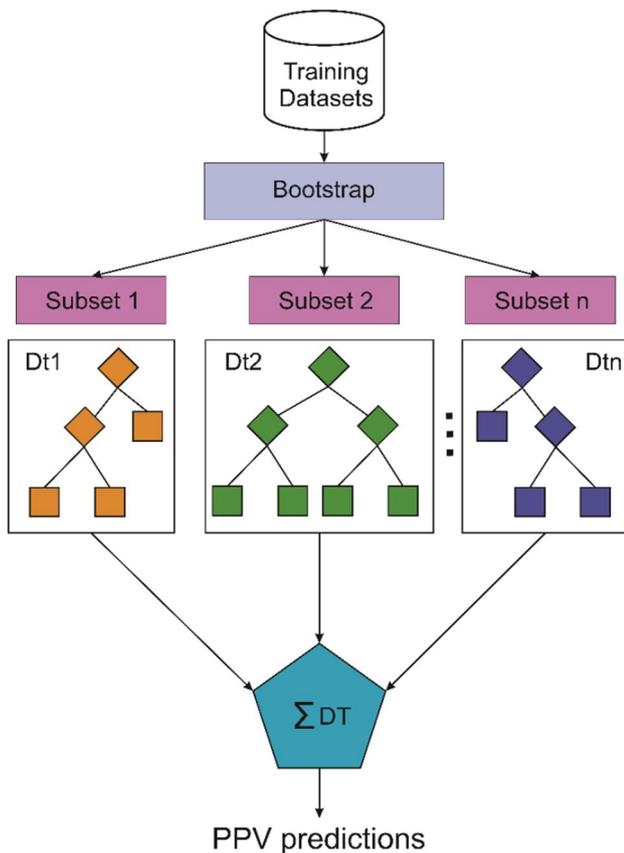


Fig. 2 Workflow of RF in predicting blast-induced PPV

models. The testing dataset is used to assess the performance of the constructed models.

To evaluate the performance of the constructed models, two criteria statistical include determination coefficient (R^2) and root-mean-square error (RMSE) are used with RMSE provide an idea of how wrong all predictions are (0 is perfect), and R^2 provides an idea of how well the model fits the data (1 is perfect, 0 is worst). In this study, RMSE and R^2 were computed using the following equations:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{12}$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{13}$$

where n denotes for the number of data, y_i and \hat{y}_i denotes the measured and predicted values, respectively; \bar{y} is the mean of the measured values.

Additionally, the Box–Cox transform and 10-fold cross-validation methods are used to avoid overfitting/underfitting.

XGBoost

In XGBoost, two stopping criteria, namely, *maximum tree depth* and *nrounds*, were considered to prevent complexity in modeling. Selecting the significant values for *maximum tree depth* and the *nrounds* causes excessive growth of the tree and an overfitting problem. Therefore, the *maximum tree depth* is set in the 1–3 range, and *nrounds* is set as 50, 100, and 150.

To achieve an optimum combination of these two parameters, a trial-and-error procedure was conducted with the range of two settings proposed. The performance indices, which include RMSE and R^2 , were calculated to evaluate the

XGBoost models on both the training and testing datasets (Table 3).

Based on Table 3, nine XGBoost models were developed and evaluated. The results of the XGBoost models in Table 3 are very close to each other, which causes difficulty in selecting the best model. Thus, a simple procedure with the ranking method proposed by Zorlu et al. (2008) is applied in Table 4. The XGBoost models in Table 4 are ranked and evaluated through ranking indicators. The results of the overall grade for XGBoost models 1–9 are summarized in Table 5.

According to Table 5, model 1 with the total rank value of 35 reached the highest value among all the constructed XGBoost models. On other words, the XGBoost model No. 1 performed better than the other XGBoost models in this study.

Support vector machine (SVM)

In SVM, the kernel function with polynomial kernel was used to develop the SVM models. Two stopping criteria, namely *degree* and *cost*, were considered to prevent complexity in modeling. Also, the *scale* parameter was held constant at a value of 0.1. In this study, we select the range of 1–3 for the *degree* and set the *cost* as 0.25, 0.5, and 1.

To achieve an optimum combination of these two parameters, a trial-and-error procedure was also conducted similarly to that for the XGBoost method with the range of the two SVM parameters. The performance indices, namely, RMSE and R^2 , were calculated to evaluate the SVM models on both the training and testing datasets (Table 6).

Table 6 shows some low-performance models such as nos. 1, 4, 7, 2, 5. However, some models exhibit high performances that are almost similar. Thus, a simple ranking method should be applied to determine the best SVM model among the developed ones, as shown in Table 7. Table 8 indicates the total rank of the SVM models 1–9.

According to Table 8, model 6 with a total rank of 32 achieved the best performance among all the developed

Table 3 Performance indicators of the XGBoost models

Technique	Model no.	nrounds	max_depth	Training process		Testing process	
				RMSE	R^2	RMSE	R^2
XGBoost	1	150	1	1.554	0.955	1.742	0.952
	2	150	2	1.636	0.951	1.783	0.951
	3	150	3	1.711	0.945	1.840	0.948
	4	100	1	1.568	0.954	1.745	0.951
	5	100	2	1.609	0.953	1.765	0.950
	6	100	3	1.662	0.949	1.853	0.947
	7	50	1	1.725	0.951	1.947	0.946
	8	50	2	1.618	0.953	1.748	0.952
	9	50	3	1.617	0.952	1.808	0.949

Table 4 The ranking of the XGBoost models based on their performance

Technique	Model	RMSE	R^2	Rank for RMSE	Rank for R^2	Total rank
XGBoost	Training 1	1.554	0.955	9	9	18
	Training 2	1.636	0.951	4	4	8
	Training 3	1.711	0.945	2	1	3
	Training 4	1.568	0.954	8	8	16
	Training 5	1.609	0.953	7	7	14
	Training 6	1.662	0.949	3	2	5
	Training 7	1.725	0.951	1	3	4
	Training 8	1.618	0.953	5	6	11
	Training 9	1.617	0.952	6	5	11
	Testing 1	1.742	0.952	9	8	17
	Testing 2	1.783	0.951	5	6	11
	Testing 3	1.840	0.948	3	3	6
	Testing 4	1.745	0.951	8	7	15
	Testing 5	1.765	0.950	6	5	11
	Testing 6	1.853	0.947	2	2	4
	Testing 7	1.947	0.946	1	1	2
	Testing 8	1.748	0.952	7	9	16
	Testing 9	1.808	0.949	4	4	8

Table 5 Total rank of XGBoost models

Technique	Model no.	Total rank
XGBoost	1	35
	2	19
	3	9
	4	31
	5	25
	6	9
	7	6
	8	27
	9	19

The best model was shown in bold

SVM models. Thus, we conclude that model 6 is the best SVM model with the SVM method. Note that, the same training and testing datasets were applied for the development of the SVM models as those used for the XGBoost models.

Random forest (RF)

With the RF technique, two stopping criteria called n_{tree} and m_{try} were considered to prevent complexity and reduce the running time of the model. A trial-and-error procedure with n_{tree} is discussed in the range of 50–150, whereas m_{try} set as 5, 7, and 9 is implemented in Table 9. Likewise to the development of the XGBoost and SVM models, the same training and testing datasets were applied for the development of the RF models in this study.

Table 6 Performance indices of SVM models

Method	Model no.	Degree	Cost	Training data		Testing data	
				RMSE	R^2	RMSE	R^2
SVM	1	1	0.25	2.763	0.877	3.875	0.805
	2	2	0.25	2.213	0.917	2.991	0.869
	3	3	0.25	2.074	0.923	2.580	0.926
	4	1	0.5	2.725	0.880	3.763	0.808
	5	2	0.5	2.185	0.918	2.753	0.890
	6	3	0.5	2.036	0.921	2.566	0.934
	7	1	1	2.664	0.883	3.908	0.783
	8	2	1	2.186	0.916	2.469	0.916
	9	3	1	2.152	0.911	2.440	0.944

Table 7 Performance indices of SVM models with the rank

Method	Model	RMSE	R^2	Rank for RMSE	Rank for R^2	Total rank
SVM	Training 1	2.763	0.877	1	1	2
	Training 2	2.213	0.917	4	6	10
	Training 3	2.074	0.923	8	9	17
	Training 4	2.725	0.880	2	2	4
	Training 5	2.185	0.918	6	7	13
	Training 6	2.036	0.921	9	8	17
	Training 7	2.664	0.883	3	3	6
	Training 8	2.186	0.916	5	5	10
	Training 9	2.152	0.911	7	4	11
	Testing 1	3.875	0.805	2	2	4
	Testing 2	2.991	0.869	4	4	8
	Testing 3	2.580	0.926	6	7	13
	Testing 4	3.763	0.808	3	3	6
	Testing 5	2.753	0.890	5	5	10
	Testing 6	2.566	0.934	7	8	15
	Testing 7	3.908	0.783	1	1	2
	Testing 8	2.469	0.916	8	6	14
	Testing 9	2.440	0.944	9	9	18

Table 8 Total rank of SVM models

Technique	Model no.	Total rank
SVM	1	6
	2	18
	3	30
	4	10
	5	23
	6	32
	7	8
	8	24
	9	29

The best model was shown in bold

Based on Table 9, all of the nine constructed RF models are suitable for estimating blast-produced PPV in this study. Some of the RF models, such as models 5–9, provide higher performance than others. However, the results of the models are nearly similar. Thus, concluding which model is the best for the RF technique is difficult. A ranking technique was used to identify the best model for the RF technique, as reported in Table 10. Additionally, a total ranking of the RF models is computed in Table 11.

According to Tables 10 and 11, RF model 7 with a total ranking value of 30 reached the highest value among all the developed RF models. Thus, we can conclude that RF model 7 with $n_{\text{tree}} = 150$ and $m_{\text{try}} = 9$ is the superior model in the RF technique for anticipating blast-produced PPV in this study.

Table 9 The RF models performance for predicting blast-induced PPV

Technique	Model no.	n_{tree}	m_{try}	Training process		Testing process	
				RMSE	R^2	RMSE	R^2
RF	1	150	5	1.816	0.944	2.837	0.885
	2	100	5	1.819	0.944	2.975	0.874
	3	50	5	1.803	0.944	3.126	0.859
	4	150	7	1.736	0.948	2.031	0.939
	5	100	7	1.737	0.948	1.945	0.943
	6	50	7	1.745	0.948	1.925	0.944
	7	150	9	1.719	0.946	1.811	0.947
	8	100	9	1.723	0.946	1.804	0.948
	9	50	9	1.717	0.946	1.886	0.944

Table 10 The RF models with their rank through performance indicators

Technique	Model	RMSE	R^2	Rank for RMSE	Rank for R^2	Total rank
RF	Training 1	1.816	0.944	2	1	3
	Training 2	1.819	0.944	1	3	4
	Training 3	1.803	0.944	3	2	5
	Training 4	1.736	0.948	6	8	14
	Training 5	1.737	0.948	5	9	14
	Training 6	1.745	0.948	4	7	11
	Training 7	1.719	0.946	8	6	14
	Training 8	1.723	0.946	7	4	11
	Training 9	1.717	0.946	9	5	14
	Testing 1	2.837	0.885	3	3	6
	Testing 2	2.975	0.874	2	2	4
	Testing 3	3.126	0.859	1	1	2
	Testing 4	2.031	0.939	4	4	8
	Testing 5	1.945	0.943	5	5	10
	Testing 6	1.925	0.944	6	7	13
	Testing 7	1.811	0.947	8	8	16
	Testing 8	1.804	0.948	9	9	18
	Testing 9	1.886	0.944	7	6	13

Table 11 Total ranking of RF models

Technique	Model no.	Total rank
RF	1	9
	2	8
	3	7
	4	22
	5	24
	6	24
	7	30
	8	29
	9	27

The best model was shown in bold

k-nearest neighbor (KNN)

In this study, nine KNN models were developed with the *k* neighbors set in a range of 3–11 through training datasets. The performance of the KNN models was evaluated using the testing dataset as the second step in the development of the KNN models. Note that the same datasets were used for the development of the KNN models as those used for the development of the models above. The performance indices of the KNN models are shown in Table 12.

As shown in Table 12, the results of the constructed KNN models are close to one another. Thus, determining which model is the most optimal among the built KNN models is difficult. A simple ranking method similar to the previous sections was applied to the KNN technique. The performance indices of the KNN models with their rank were

Table 12 The KNN models performance in this study

Technique	Model no.	<i>k</i>	Training process		Testing process	
			RMSE	R^2	RMSE	R^2
KNN	1	3	2.878	0.853	3.708	0.791
	2	4	2.864	0.861	3.581	0.809
	3	5	2.796	0.876	3.161	0.859
	4	6	2.857	0.878	3.326	0.851
	5	7	2.985	0.867	3.203	0.870
	6	8	3.008	0.867	3.278	0.874
	7	9	3.006	0.866	3.368	0.861
	8	10	3.061	0.859	3.466	0.850
	9	11	3.017	0.864	3.434	0.863

Table 13 Performance of the KNN models with the rank

Technique	Model	RMSE	R^2	Rank for RMSE	Rank for R^2	Total rank
KNN	Training 1	2.878	0.853	6	1	7
	Training 2	2.864	0.861	7	3	10
	Training 3	2.796	0.876	9	8	17
	Training 4	2.857	0.878	8	9	17
	Training 5	2.985	0.867	5	7	12
	Training 6	3.008	0.867	3	6	9
	Training 7	3.006	0.866	4	5	9
	Training 8	3.061	0.859	1	2	3
	Training 9	3.017	0.864	2	4	6
	Testing 1	3.708	0.791	1	1	2
	Testing 2	3.581	0.809	2	2	4
	Testing 3	3.161	0.859	9	5	14
	Testing 4	3.326	0.851	6	4	10
	Testing 5	3.203	0.870	8	8	16
	Testing 6	3.278	0.874	7	9	16
	Testing 7	3.368	0.861	5	6	11
	Testing 8	3.466	0.850	3	3	6
	Testing 9	3.434	0.863	4	7	11

Table 14 Total rank of KNN models

Technique	Model no.	Total rank
KNN	1	9
	2	14
	3	31
	4	27
	5	28
	6	25
	7	20
	8	9
	9	17

The best model was shown in bold

calculated and the results are presented in Table 13. Additionally, Table 14 shows the total rank of KNN models.

According to Tables 13 and 14, nine KNN models were ranked with the value of total rank in the range of 9–31. As shown in the tables, KNN model 3 with an entire rank value of 31 achieved the highest value among the developed KNN models.

Validation performance of models

In this study, two statistical criteria, namely, R^2 and RMSE, were employed to measure the performance of the selected predictive models and computed using Eqs. (12–13). After the optimal models for each technique

Table 15 Statistical values for selected predictive models

Model	Training data		Testing data	
	RMSE	R^2	RMSE	R^2
XGBoost	1.554	0.955	1.742	0.952
SVM	2.036	0.921	2.566	0.934
RF	1.719	0.946	1.811	0.947
KNN	2.796	0.876	3.161	0.859

were selected, the values of the aforementioned statistical criteria for all models were calculated for both the training and testing datasets, as indicated in Table 15. According to these results, the accuracy level of the XGBoost technique is better than those of the SVM, RF, and KNN models. Figure 4 demonstrates the performance of the models in forecasting blast-induced PPV on the testing dataset.

Figure 5 presents a useful way to consider the spread of the estimated accuracies for the various methods and how they relate among the XGBoost, SVM, RF, and KNN techniques. According to Fig. 5, the KNN technique has the lowest accuracy level with several outliers, whereas the XGBoost technique exhibits the highest accuracy level without outliers. The RF technique can also provide an approximation of the XGBoost performance. However, a closer look shows that the developed XGBoost model offers higher performance than the RF model. Furthermore, the RF technique appears to have outliers, whereas the established XGBoost model has none. Additionally, the accuracy of the selected PPV predictive models was also

Fig. 4 Measured versus predicted values on the testing dataset

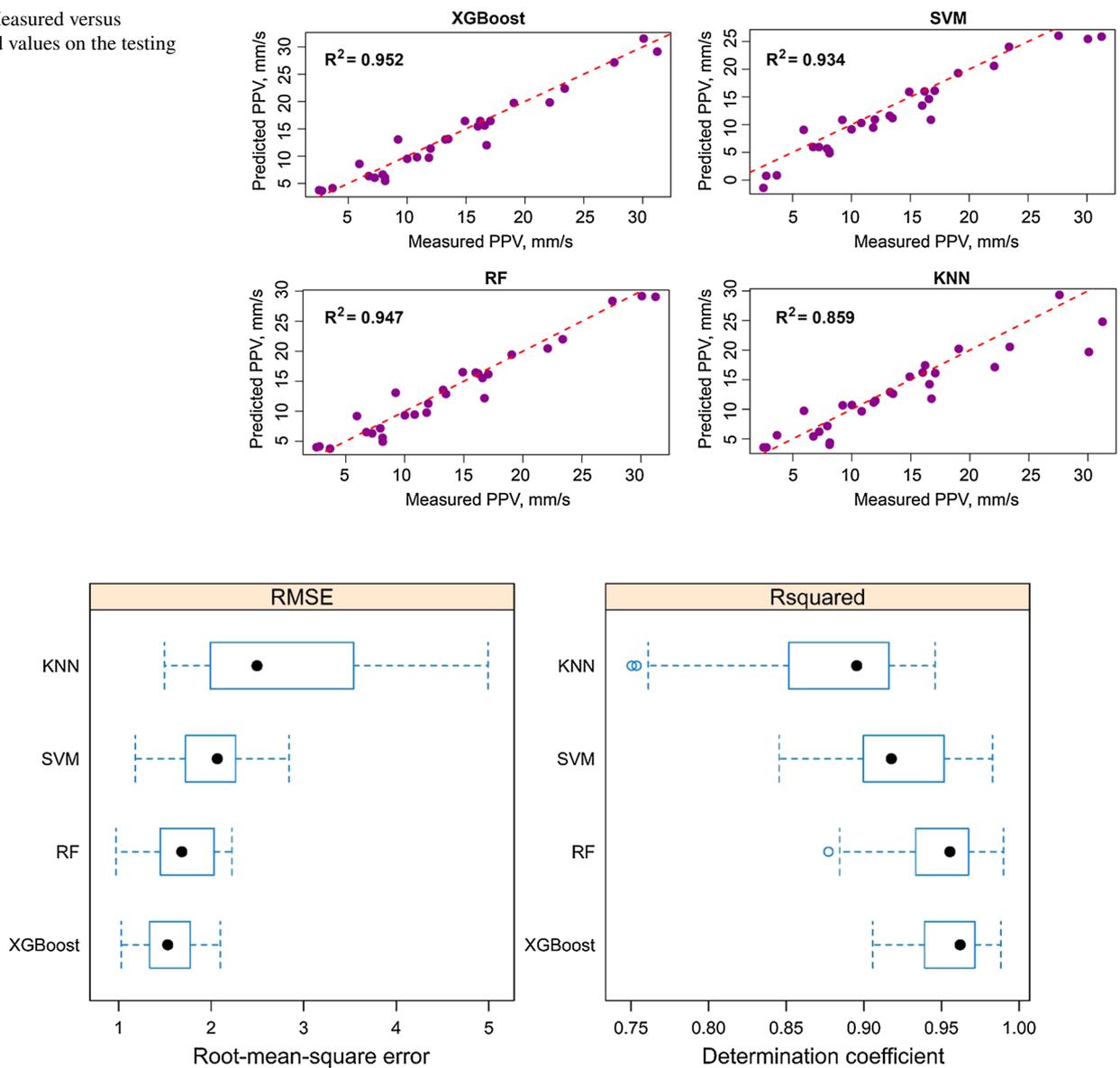


Fig. 5 Comparison of machine learning algorithms in box and whisker plots

compared and shown in Fig. 6. According to Fig. 6, among the developed models, the XGBoost technique yields the most reliable results in forecasting blast-produced PPV.

Considering the input variables in this study, it shows that the number of input variables is high (9 input variables). Therefore, an analysis procedure of sensitivity was performed to find out which input variable(s) is/are the most influential parameters on blast-induced PPV as shown in Fig. 7. As a result, Q (charge) and D (distance) are the most influential factors on blast-induced PPV in this study. They should be used in practical engineering to control blast-induced PPV. The other input parameters were also effected on blast-induced PPV but not much.

Conclusions and recommendations

In practice, an accurate and efficient estimation of PPV is essential to reduce the environmental effects of blasting operations, especially near residential areas. This study developed the XGBoost, SVM, RF, and KNN models to predict PPV caused by blasting operations in the Deo Nai open-pit coal mine in Vietnam. Nine input parameters (Q , H , B , S , T , q , N , D , and Δt) were used to predict PPV from 146 blasting events at the mine. For modeling purposes, all datasets were divided into training and testing sets, with 80% (118 observations) of the entire dataset used for training and 20% (28 representations) for testing. The performance

Fig. 6 Prediction values of selected predictive models on testing datasets

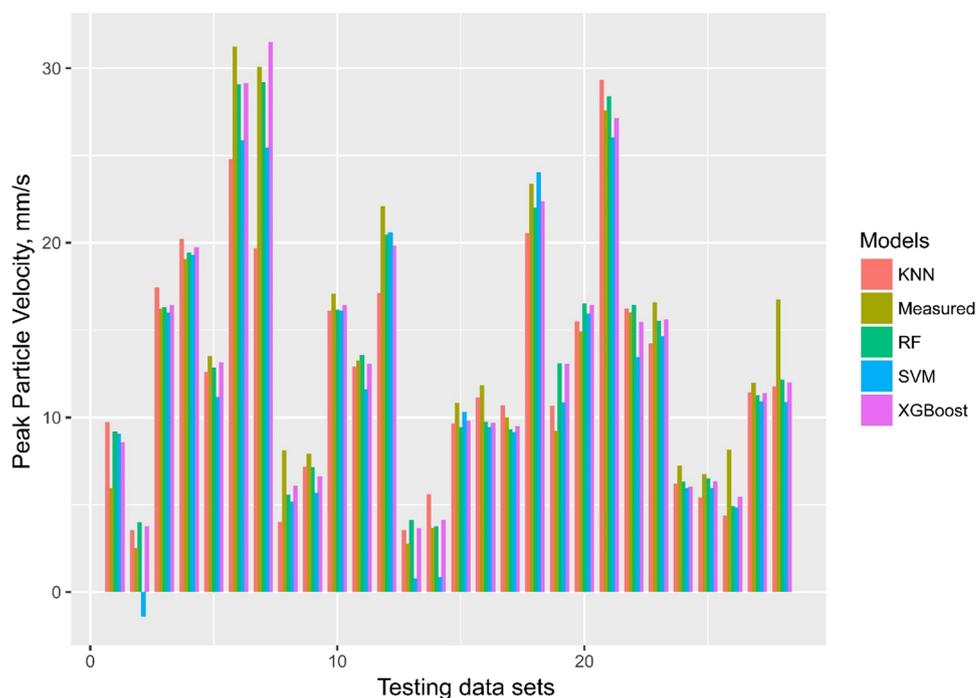
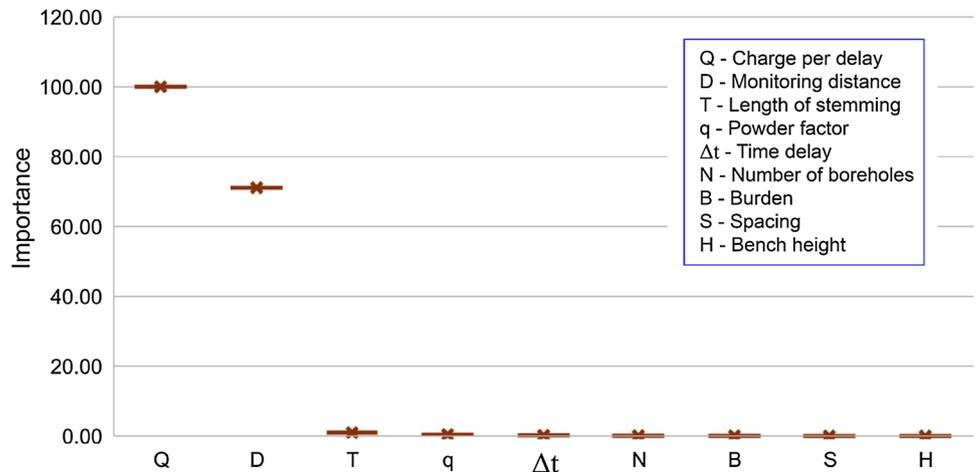


Fig. 7 Sensitivity analysis of independent variables for the PPV predictive model



of the predictive models was evaluated based on two criteria, namely, R^2 and RMSE, using the training and testing datasets. Based on the results of this study, RMSE values of 1.554 and 1.742 were obtained for the XGBoost model on the training and testing datasets, respectively. These values are the smallest among the RMSE values of the constructed models, which shows that the XGBoost model can be introduced as a new approach to solve environmental problems caused by blasting. Furthermore, R^2 values of 0.955 and 0.952, respectively, for the training and testing datasets of the XGBoost technique indicate that the capability of the proposed technique is slightly higher than that of the other developed models for PPV prediction.

Although XGBoost was a robust model for predicting blast-induced PPV in this study, it is still needed to be further studied for improving the accuracy level as well as the computational time. Also, a hybrid model based on XGBoost and another algorithm are also a good idea for future works.

Acknowledgements We would like to thank the Hanoi University of Mining and Geology (HUMG), Vietnam; Ministry of Education and Training of Vietnam (MOET); The Center for Mining, Electro-Mechanical research of HUMG.

Compliance with ethical standards

Conflict of interest On behalf of all authors, I hereby attest that no conflict of interest exists in financial relationships, intellectual property, or any point related to publishing ethics.

References

- Ak H, Konuk A (2008) The effect of discontinuity frequency on ground vibrations produced from bench blasting: a case study. *Soil Dynamics and Earthquake Engineering* 28:686–694
- Alnaqi AA, Moayedi H, Shahsavari A, Nguyen TK (2019) Prediction of energetic performance of a building integrated photovoltaic/thermal system thorough artificial neural network and hybrid particle swarm optimization models. *Energy Convers Manag* 183:137–148
- Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 46:175–185
- Ambraseys NR, Hendron AJ (1968) Dynamic behavior of rock masses. In: Stagg KG, Zienkiewicz OC (eds) *Rock mechanics in engineering practices*. Wiley, New York, pp 203–207
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Bui X-N, Nguyen H, Le H-A, Bui H-B, Do N-H (2019) Prediction of blast-induced air over-pressure in open-pit mine: assessment of different artificial intelligence. *Tech Natl Resour Res*. <https://doi.org/10.1007/s11053-019-09461-0>
- Chandar KR, Sastry V, Hegde C (2017) A critical comparison of regression models and artificial neural networks to predict ground vibrations. *Geotech Geol Eng* 35:573–583
- Chen T, He T (2015) Xgboost: extreme gradient boosting R package version 04-2
- Chen G, Huang SL (2001) Analysis of ground vibrations caused by open pit production blasts—a case study. *Fragblast* 5(1–2):91–107
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
- Coursen DL (1995) Method of reducing ground vibration from delay blasting. Google Patents
- Davies B, Farmer I, Attewell P (1964) Ground vibration from shallow sub-surface blasts. *Engineer* 217(5644):553–559
- Drucker H, Burges CJ, Kaufman L, Smola AJ, Vapnik V (1997) Support vector regression machines. In: *Advances in neural information processing systems*, pp 155–161
- Duvall WI, Fogelson DE (1962) Review of criteria for estimating damage to residences from blasting vibrations. US Department of the Interior, Bureau of Mines
- Faradonbeh RS, Monjezi M (2017) Prediction and minimization of blast-induced ground vibration using two robust meta-heuristic algorithms. *Eng Comput* 33(4):835–851. <https://doi.org/10.1007/s00366-017-0501-6>
- Faradonbeh RS, Armaghani DJ, Majid MA, Tahir MM, Murlidhar BR, Monjezi M, Wong H (2016) Prediction of ground vibration due to quarry blasting based on gene expression programming: a new model for peak particle velocity prediction. *Int J Environ Sci Technol* 13:1453–1464
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29(5):1189–1232
- Friedman JH (2002) Stochastic gradient boosting. *Comput Stat Data Anal* 38:367–378
- Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *Ann Stat* 28:337–407
- Friedman J, Hastie T, Tibshirani R (2001) *The elements of statistical learning* vol 1. vol 10. Springer series in statistics New York, NY, USA
- Gad EF, Wilson JL, Moore AJ, Richards AB (2005) Effects of mine blasting on residential structures *Journal of performance of constructed facilities* 19:222–228
- Gao W, Dimitrov D, Abdo H (2018a) Tight independent set neighborhood union condition for fractional critical deleted graphs and ID deleted graphs. *Discrete Continuous Dyn Syst S* 123–144. <https://doi.org/10.3934/dcdss.2019045>
- Gao W, Guirao JL, Basavanagoud B, Wu J (2018b) Partial multi-dividing ontology learning algorithm. *Inf Sci* 467:35–58
- Gao W, Guirao JLG, Abdel-Aty M, Xi W (2018c) An independent set degree condition for fractional critical deleted graphs. *Discrete Continuous Dyn Syst S* 12:877–886. <https://doi.org/10.3934/dcdss.2019058>
- Gao W, Wang W, Dimitrov D, Wang Y (2018d) Nano properties analysis via fourth multiplicative ABC indicator calculating. *Arab J Chem* 11:793–801
- Gao W, Wu H, Siddiqui MK, Baig AQ (2018e) Study of biological networks using graph theory. *Saudi J Biol Sci* 25:1212–1219
- Ghasemi E, Ataei M, Hashemolhosseini H (2013) Development of a fuzzy model for predicting ground vibration caused by rock blasting in surface mining. *J Vib Control* 19:755–770
- Ghasemi E, Kalhori H, Bagherpour R (2016) A new hybrid ANFIS–PSO model for prediction of peak particle velocity due to bench blasting. *Eng Comput* 32:607–614
- Hajihassani M, Armaghani DJ, Sohaei H, Mohamad ET, Marto A (2014) Prediction of airblast-overpressure induced by blasting using a hybrid artificial neural network and particle swarm optimization. *Appl Acoust* 80:57–67
- Hajihassani M, Armaghani DJ, Marto A, Mohamad ET (2015) Ground vibration prediction in quarry blasting through an artificial neural network optimized by imperialist competitive algorithm. *Bull Eng Geol Environ* 74:873–886
- Hasanipanah M, Monjezi M, Shahnazar A, Armaghani DJ, Farazmand A (2015) Feasibility of indirect determination of blast induced ground vibration based on support vector machine. *Measurement* 75:289–297
- Hasanipanah M, Armaghani DJ, Amnieh HB, Majid MZA, Tahir MM (2017a) Application of PSO to develop a powerful equation for prediction of flyrock due to blasting. *Neural Comput Appl* 28:1043–1050
- Hasanipanah M, Faradonbeh RS, Amnieh HB, Armaghani DJ, Monjezi M (2017b) Forecasting blast-induced ground vibration developing a CART model. *Eng Comput* 33:307–316
- Hasanipanah M, Golzar SB, Larki IA, Maryaki MY, Ghahremanians T (2017c) Estimation of blast-induced ground vibration through a soft computing framework. *Eng Comput* 33(4):951–959. <https://doi.org/10.1007/s00366-017-0508-z>
- Hu C, Jain G, Zhang P, Schmidt C, Gomadam P, Gorka T (2014) Data-driven method based on particle swarm optimization and k -nearest neighbor regression for estimating capacity of lithium-ion battery. *Appl Energy* 129:49–55
- Longjun D, Xibing L, Ming X, Qiyue L (2011) Comparisons of random forest and support vector machine for predicting blasting vibration characteristic parameters. *Procedia Eng* 26:1772–1781
- Moayedi H, Nazir R (2018) Malaysian experiences of peat stabilization, state of the art. *Geotech Geol Eng* 36:1–11
- Moayedi H, Rezaei A (2017) An artificial neural network approach for under-reamed piles subjected to uplift forces in dry sand. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-017-2990-z>
- Moayedi H, Raftari M, Sharifi A, Jusoh WAW, Rashid ASA (2019) Optimization of ANFIS with GA and PSO estimating α ratio in

- driven piles. *Eng Comput.* <https://doi.org/10.1007/s00366-018-00694-w>
- Monjezi M, Bahrami A, Varjani AY, Sayadi AR (2011a) Prediction and controlling of flyrock in blasting operation using artificial neural network. *Arab J Geosci* 4:421–425
- Monjezi M, Ghafurikalajahi M, Bahrami A (2011b) Prediction of blast-induced ground vibration using artificial neural networks. *Tunn Undergr Space Technol* 26:46–50
- Monjezi M, Hasanipanah M, Khandelwal M (2013) Evaluation and prediction of blast-induced ground vibration at Shur River Dam, Iran, by artificial neural network. *Neural Comput Appl* 22:1637–1643
- Nguyen H, Bui X-N (2018a) A comparison of artificial neural network and empirical technique for predicting blast-induced ground vibration in open-pit mine. In: *Mining sciences and technology—XXVI*, Mong Cai, Hanoi, Vietnam. Industry and trade of the socialist Republic of Vietnam, pp 177–182
- Nguyen H, Bui X-N (2018b) Predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest. *Natl Resour Res.* <https://doi.org/10.1007/s11053-018-9424-1>
- Nguyen H, Bui X-N, Bui H-B, Mai N-L (2018a) A comparative study of artificial neural networks in predicting blast-induced air-blast overpressure at Deo Nai open-pit coal mine, Vietnam. *Neural Comput Appl.* <https://doi.org/10.1007/s00521-018-3717-5>
- Nguyen H, Bui X-N, Tran Q-H, Le T-Q, Do N-H, Hoa LTT (2018b) Evaluating and predicting blast-induced ground vibration in open-pit mine using ANN: a case study in Vietnam SN. *Appl Sci* 1:125. <https://doi.org/10.1007/s42452-018-0136-2>
- Nguyen H, Bui X-N, Tran Q-H, Mai N-L (2019) A new soft computing model for estimating and controlling blast-produced ground vibration based on hierarchical *k*-means clustering and cubist algorithms. *Appl Soft Comput.* <https://doi.org/10.1016/j.asoc.2019.01.042>
- Protodiakonov M, Koifman M, Chirkov S, Kuntish M, Tedder R (1964) *Rock strength passports and methods for their determination.* Nauka, Moscow
- Roy PP (1991) Prediction and control of ground vibration due to blasting. *Colliery Guard* 239:215–219
- Saadat M, Khandelwal M, Monjezi M (2014) An ANN-based approach to predict blast-induced ground vibration of Gol-E-Gohar iron ore mine. *Iran J Rock Mech Geotech Eng* 6:67–76
- Sheykhi H, Bagherpour R, Ghasemi E, Kalhori H (2018) Forecasting ground vibration due to rock blasting: a hybrid intelligent approach using support vector regression and fuzzy C-means clustering. *Eng Comput* 34(2):357–365. <https://doi.org/10.1007/s00366-017-0546-6>
- Standard I (1973) *Criteria for safety and design of structures subjected to under ground blast ISI, IS-6922*
- Taheri K, Hasanipanah M, Golzar SB, Majid MZA (2017) A hybrid artificial bee colony algorithm-artificial neural network for forecasting the blast-produced ground vibration. *Eng Comput* 33:689–700
- Vigneau E, Courcoux P, Symoneaux R, Guérin L, Villière A (2018) Random forests: a machine learning methodology to highlight the volatile organic compounds involved in olfactory perception. *Food Qual Prefer* 68:135–145
- Vinacomin (2015) *Report on geological exploration of Coc Sau open pit coal mine, Quang Ninh, Vietnam* (in Vietnamese-unpublished). VINACOMIN, Vietnam
- Zorlu K, Gokceoglu C, Ocakoglu F, Nefeslioglu H, Acikalin S (2008) Prediction of uniaxial compressive strength of sandstones using petrography-based models. *Eng Geol* 96:141–158



Subdividing the tectonic elements of Aegean and Eastern Mediterranean with gravity and GPS data

Muzaffer Kahveci¹ · Ayça Çırmık² · Fikret Doğru³ · Oya Pamukçu² · Tolga Gönenç²

Received: 30 October 2018 / Accepted: 11 February 2019 / Published online: 23 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

Western Anatolia has been formed by the motions of the African plate, Arabian plate and Hellenic Subduction zone. The Hellenic Subduction zone, which has high seismicity, is the main tectonic feature of the eastern Mediterranean Sea related to the subduction of the African Plate beneath the Aegean Sea Plate. The Hellenic Subduction zone has a complex lithospheric structure and shows complex differences in the Aegean Sea in terms of continental crust thickness and mantle velocity. In the study area, the directions of Global Positioning System (GPS) velocity vectors which are towards SE change towards S from North of Western Anatolia to Hellenic Subduction zone. It is thought that the factor which controls this mechanism is the shear force or subduction zone located in Aegean Sea. Western Anatolia region, which is located in Western Anatolia Extensional province, includes several morphologically significant N–S trending active normal faults. Besides, the NE–SW and NW–SE trending faults, which their kinematic features change from north to south, are very effective on the tectonic regime of the region. Additionally, for determining the boundaries of these tectonic elements, the Complete Spherical Bouguer (CSB) gravity anomaly of study area was calculated by using World Gravity Map (WGM2012) model. Moreover, in historical and instrumental studies, the high seismicity of the study area is remarkable. It is thought that this case is also related with the mechanism which oriented the GPS velocity vectors to southward. Consequently, the dominant kinematic structure of the region was classified by combining the GPS velocity vectors computed for Izmir and its surroundings bounded by Western Anatolia, Aegean Sea and Eastern Mediterranean and the CSB gravity anomaly. Finally, the results were interpreted together with focal depth distributions of earthquakes and Bouguer gravity data.

Keywords GPS · Tectonic elements · WGM2012 model · Gravity · Western Anatolia · Eastern Mediterranean · Aegean Sea

Introduction

The tectonic movements of Arabian, African and Eurasian plates have formed the Eastern Mediterranean tectonism by compression in Eastern Anatolia, extension in Western Anatolia and faulting along North Anatolian Fault and East Anatolian Fault zones. Therefore, the Anatolian block has moved to W–SW, and additionally, due to the movement

of African plate towards Eurasian plate, these oceanic subductions are driven throughout Hellenic Subduction zone (Fig. 1) (McKenzie 1972; Le Pichon and Angelier 1979; McClusky et al. 2000; Mart and Ryan 2003; Pamukçu 2016). The roll-back system of the Mediterranean slab subduction under the Aegean Sea causes the great subduction throughout Hellenic Subduction zone. Additionally, the slab pull force in the Mediterranean creates N–S extension in the Western Anatolia and Aegean region (Le Pichon and Angelier 1979; Le Pichon 1982; Mercier et al. 1989; Sorel et al. 1988).

The Western Anatolia is one of the most active extensional regions in the world (McKenzie 1978; Le Pichon and Angelier 1979; Dewey et al. 1986; Jackson and McKenzie 1988; Taymaz et al. 1990; Ambraseys and Jackson 1990; Goldsworthy et al. 2002; Nyst and Thatcher 2004). For clarifying the reason of extension, different models were suggested by Dewey (1988), Seyitoğlu and Scott (1991),

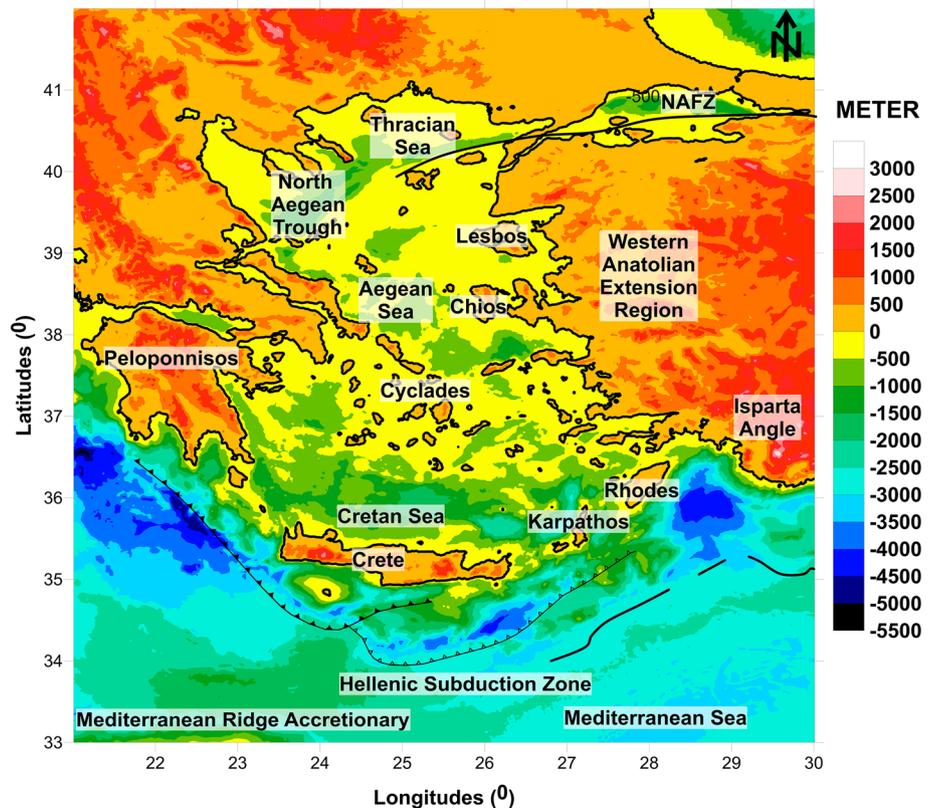
✉ Ayça Çırmık
ayca.cirmik@deu.edu.tr

¹ Department of Geomatics Engineering, Engineering and Natural Sciences Faculty, Konya Technical University, Konya, Turkey

² Department of Geophysical Engineering, Engineering Faculty, Dokuz Eylul University, Buca, Izmir, Turkey

³ Department of Geophysical Engineering, Oltu Earth Sciences Faculty, Atatürk University, Oltu, Erzurum, Turkey

Fig. 1 The main tectonic structures in the study area (Gessner et al. 2013, the topographic data were obtained from TOPEX; Smith and Sandwell 1997)



Dewey and Şengör (1979) and Le Pichon and Angelier (1979). These models were presented as four type models: the first is Anatolian plate escape, the second is orogenic collapse, the third is back-arc extension, and the last one is the combined model of the three models. In According to Yılmaz et al. (2000) and Bozkurt and Sözbilir (2004) the N-S trending extensional system of Western Anatolia has been still active since Mio-Pliocene.

Reilinger et al. (2010) pointed out that there was no coherency between the recent motions of the Aegean region and the previous geophysical and geological findings of previous studies (Makris 1978; Armijo et al. 1996). The GPS studies were carried out in Aegean region (McClusky et al. 2000; Kahle et al. 2000; Reilinger et al. 2006, 2010) indicating that the southward of Aegean Sea and Peloponnisos transform towards SW with minor internal deformation and according to Reilinger et al. (2010) new fault system causes these transformations.

Western Anatolia and Aegean Sea, which are chosen as the study region in this study, were evaluated in the previous studies (McClusky et al. 2000; Reilinger et al. 2006; Nyst and Thatcher 2004; Floyd et al. 2010) as blocks due to the relative motions. In the study of Nyst and Thatcher (2004), it was indicated that the GPS velocity vectors in and around Izmir were distinct from Western Anatolia.

Hence, GPS measurements were carried out in 2009, 2010, 2011 for observing the behaviours of main tectonic structures located in Izmir and its surroundings by Kahveci et al. (2013), Pamukçu et al. (2015) and Çırmık et al. (2017).

In this study, the Eurasia-fixed-frame solutions of McClusky et al. (2000), Hollenstein et al. (2008), Çırmık et al. (2017) and Çırmık and Pamukçu (2017) were evaluated together and consequently, new microplate models were proposed for Western Anatolian region, Aegean Sea and eastward of Hellenic Subduction zone by examining the Eurasia-fixed-frame solutions with Complete Spherical Bouguer (CSB) gravity anomalies, earthquake focal depth distributions. In this study, the study region was divided into seven microplates and three regions due to the differences on the amplitudes and directions of GPS velocities, topographic/bathymetric and CSB gravity anomaly values. Additionally, when seismicity, gravity anomaly values and the microplate borders were evaluated it was found that the transitions in low and high seismicity zones present the borders of the microplates; therefore, it can be said that the separation of the study area into microplates was related with the density changes in the tectonic elements.

The whole steps of the study are given briefly with a flow-chart for following the applications clearly.

Analysis of the GPS solutions

In the scope of this study, the GPS velocities of four different studies (McClusky et al. 2000; Hollenstein et al. 2008; Çirmik et al. 2017; Çirmik and Pamukçu 2017) were evaluated together for examining the velocity change related with the kinematic structure in Western Anatolia, Aegean Sea and Eastern Mediterranean (Fig. 2). The GPS data of McClusky et al. (2000) and Hollenstein et al. (2008) belong to the years between 1988–1997 and 1993–2003, respectively. The GPS data of Çirmik et al. (2017) and Çirmik and Pamukçu (2017) belong to 2009, 2010 and 2011.

Before evaluating all GPS velocities, a reference frame transformation is needed since the GPS velocities of McClusky et al. (2000) were calculated with ITRF96, Hollenstein et al. (2008) were calculated with ITRF00 and the GPS velocities of Çirmik et al. (2017) and Çirmik and Pamukçu (2017) were calculated with ITRF08. By calculating the rotation between common sites of two velocity fields, the reference frame effect was removed. In this study, ITRF08 was chosen as the reference frame and the frames of McClusky et al. (2000) and Hollenstein et al. (2008) were transformed to ITRF08. Initially, the

GPS data of Çirmik et al. (2017) and Çirmik and Pamukçu (2017) were processed together with ITRF08 with Gamit/Globk (Herring et al. 2015) software. During processing, for increasing the common sites with McClusky et al. (2000) four IGS stations (NYAL, KIT3, METS, JOZE) were added. In this way, by calculating the rotation between Çirmik et al. (2017), Çirmik and Pamukçu (2017) and McClusky et al. (2000) at six common sites (KRPT, WTZR, NYAL, KIT3, METS and JOZE) and between Hollenstein et al. (2008) at nine common sites (KRPT, BUCU, GLSV, ISTA, MATE, NICO, TUBI, WTZR and ZECK), the reference frames effects were removed by using Gamit/Globk software.

Following this transformation, the plate-centric GPS velocities were inverted to estimate the angular velocities on the common sites velocities. Generally, in plate motion models, the motions of plates on the Earth's surface can be described by Euler rotations. In other words, the relative motions between any two plates can be described as a rotation about an Euler Pole (Euler theorem). Euler pole is the intersection of the rotation axis with the Earth's surface. Therefore, relative motions of plates can be defined by the rotations with the assumption of rigid plates on a spherical Earth (DeMets et al. 1990; Ahadov and Jin 2017; Stein and Wysession 2005). Application of Euler theorem in geophysics states that the displacement of one tectonic plate relative to other plates takes place as a rotation about the Euler pole of relative rotation between the plates (Goudarzi et al. 2014). In this study, calculation of Euler pole parameters (latitude, longitude, ω) was performed using GPS linear velocity vectors. Then, the magnitude of linear velocity (v) on the station on Plate B, with respect to Plate A due to rotation around the North Pole, can be obtained using Eq. 1:

$$v = \omega_{AB} \cdot r \quad (1)$$

where ω is the magnitude (Euclidian length) of angular velocity (Euler plate rotation) vector and r is the position vector of the station of interest (DeMets et al. 1990). One can compute this vector cross-product by converting station latitudes and longitudes (and thus converting velocity vectors) either in global Cartesian coordinate system (X, Y, Z) or in local geodetic (north, east, up) system.

As a result, the estimated Euler pole parameters are given in Table 1, where ω_x , ω_y and ω_z are called as the components of Euler vector and ω is called as the magnitude of the angular velocity calculated from Euler components.

It is seen from Table 1 that Gamit/Globk estimates of Euler parameters are smaller than their uncertainties, meaning that Euler pole in this study could be basically anywhere on Earth and it would not make a significant difference to the results of the velocity solution combination because the required rotation rate is so small (Floyd and Herring 2018). Differences at the common sites after the transformation are

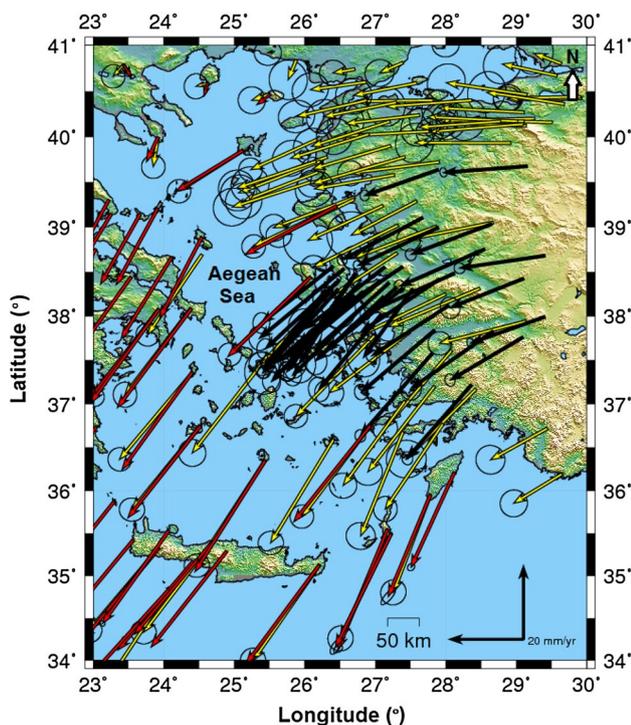


Fig. 2 GPS velocities and their 95% confidence ellipses in a Eurasia-fixed reference frame (yellow vectors: McClusky et al. 2000; red vectors: Hollenstein et al. 2008, black vectors: Çirmik et al. 2017; Çirmik and Pamukçu 2017) (This figure is created by using Generic Mapping Tools (GMT); Wessel and Smith 1998)

given in Table 2 for Hollenstein et al. (2008) and Table 3 for McClusky et al. (2000).

Determining the Complete spherical Bouguer (CSB) gravity anomaly

The satellite-based gravity models, Earth Gravity Model (EGM2008) (Pavlis et al. 2008), World Gravity Model (WGM2012) (Bonvalot et al. 2012), European Improved Gravity model of the Earth by New techniques (EIGEN6c-4) (Förste et al. 2014), Gravity field and steady-state Ocean Circulation Explorer (GOCE) and EGM2008 combined model (GECO) (Gilardoni et al. 2016) etc. have been used for decades. WGM2012 model is the first release of a high-resolution grid of the Earth's gravity anomalies (Bouguer, isostatic and surface free-air), calculated at global scale in spherical geometry by the Bureau Gravimétrique International (BGI) in the frame of collaborations with international

organizations such as Commission for the Geological Map of the World (CGMW), UNESCO, International Association of Geodesy (IAG), International Union of Geodesy and Geophysics (IUGG), International Union of Geological Sciences (IUGS) and with various scientific institutions. WGM2012 gravity anomalies are derived from EGM2008 and DTU10 (Global Ocean Tide model of Technical University of Denmark) and include “1 × 1” resolution terrain corrections derived from ETOPO1 model (Bonvalot et al. 2012). The background removal that is based on Least Squares approximation was applied on CSB gravity anomaly to remove local anomalies or possible geological trend from data. In this step, WGM2012 model was used for the calculation of CSB gravity anomaly of Western Anatolia, Aegean Sea and Eastern Mediterranean. In this purpose, trend statistics of CSB gravity anomalies were calculated with the equations given in Table 4. The best fit between WGM2012 and trend statistic was the linear trend of CSB gravity anomaly (Fig. 3).

Table 1 Euler pole parameters with respect to combined processing of Çırmık et al. (2017) and Çırmık and Pamukçu (2017) on Eurasia-fixed plate

	$\omega_x \pm \sigma$ (°/Myr)	$\omega_y \pm \sigma$ (°/Myr)	$\omega_z \pm \sigma$ (°/Myr)	$\omega \pm \sigma$ (°/Myr)	Latitude (°N)	Longitude (°E)
Hollenstein et al. (2008)	-0.00061 ± 0.0303	-0.01056 ± 0.0167	-0.02564 ± 0.0307	0.0277 ± 0.046	67.58	-93.31
McClusky et al. (2000)	-0.00086 ± 0.02606	-0.02035 ± 0.01519	-0.03103 ± 0.03464	0.0389 ± 0.045	56.72	-92.10

Table 2 Differences at the common sites for Hollenstein et al. (2008)

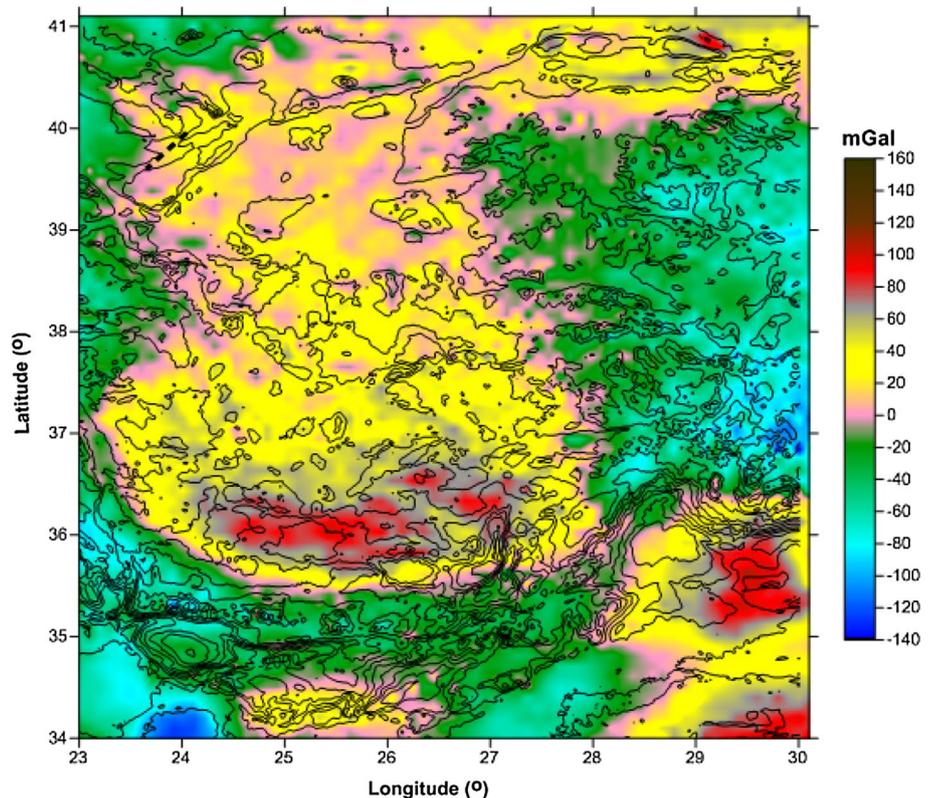
Station name	dN (mm) ± sN (mm)	dE (mm) ± sE (mm)	dU (mm) ± sU (mm)
KRPT	8.08 ± 0.98	-4.47 ± 0.84	0.99 ± 1.41
BUCU	-1.78 ± 0.55	0.94 ± 0.47	-0.00 ± 1.41
GLSV	0.17 ± 1.42	1.70 ± 1.11	-0.00 ± 1.41
ISTA	1.08 ± 0.59	-1.90 ± 0.52	0.00 ± 1.41
MATE	-1.68 ± 0.55	0.69 ± 0.60	-0.00 ± 1.41
NICO	-1.40 ± 0.72	2.62 ± 0.59	0.00 ± 1.41
TUBI	2.02 ± 0.65	-2.91 ± 0.57	-0.00 ± 1.41
WTZR	-0.45 ± 0.62	2.06 ± 0.75	-0.00 ± 1.41
ZECK	-0.25 ± 0.51	2.72 ± 0.79	-0.00 ± 1.41

Table 3 Differences at the common sites for McClusky et al. (2000)

Station name	dN (mm) ± sN (mm)	dE (mm) ± sE (mm)	dU (mm) ± sU (mm)
KRPT	5.71 ± 1.30	-5.87 ± 1.43	0.99 ± 1.41
WTZR	-2.44 ± 1.01	2.20 ± 1.09	-0.00 ± 1.41
NYAL	0.53 ± 2.44	-3.03 ± 2.04	0.00 ± 1.41
KIT3	-0.81 ± 1.91	2.26 ± 2.01	-0.00 ± 1.41
METS	-2.44 ± 1.95	1.51 ± 1.48	-0.00 ± 1.41
JOZE	-0.37 ± 1.11	0.62 ± 0.88	0.00 ± 1.41

Table 4 Trend statistics of CSB gravity anomaly

Model type	Formula	Model fit (%)
Linear	$m1 + m2 * x + m3 * y$	42.2775
Bilinear	$m1 + m2 * x + m3 * y + m4 * x * y$	41.2363
5th term	$m1 + m2 * x + m3 * y + m4 * x * y + m5 * x * x$	39.1628
6th term	$m1 + m2 * x + m3 * y + m4 * x * y + m5 * x * x + m6 * y * y$	39.1345
7th term	$m1 + m2 * x + m3 * y + m4 * x * y + m5 * x * x + m6 * y * y + m7 * x * x * x$	37.8922

Fig. 3 Linear trend removal map of WGM2012 gravity anomaly of the study area (The topographic data were obtained from TOPEX; Smith and Sandwell 1997)

Discussion and conclusion

In this study, the study region was defined as seven microplates (Fig. 4) by combining the GPS velocities of previous studies (McClusky et al. 2000; Hollenstein et al. 2008; Çırmık et al. 2017; Çırmık and Pamukçu 2017). These microplates were separated in detail by taking into account the amplitudes and the directions of the GPS velocities by benefitting from the blocks proposed by Nyst and Thatcher (2004) for Izmir and its surroundings. Eurasia-fixed reference frame solutions of these studies were examined due to the amplitudes and the directions of GPS vectors. For examining the relation between the microplates, their boundaries and topography/bathymetry, the topographic and bathymetric map (ETOPO2; NOAA, N. 2006) of the study region was created (Fig. 5a). Therefore, due to the

topographic and bathymetric differences (Fig. 5a), the study area was divided into three regions as regions A, B and C (Fig. 5b).

In Fig. 4, it is seen that the velocity vectors start to include the south component at the microplate No. 1. The movement mechanism of the microplate No.2 is exactly coherent with the tectonic escape model of McKenzie (1978). The microplate No. 3 does not move consistently with the movement mechanism of the westward tectonic escape model (McKenzie, 1978) with the effect of NAFZ mechanism. The topographic boundary between the microplate Nos. 2 and 3 (Fig. 5a, b) is the NAFZ, and NAFZ causes a boundary also at GPS velocity vector directions (Fig. 4). The microplate No. 4 is more stable with respect to Western Anatolia and Aegean Sea and moves like a different plate from the others. The movement mechanism of microplate No. 5, which is related with the movements

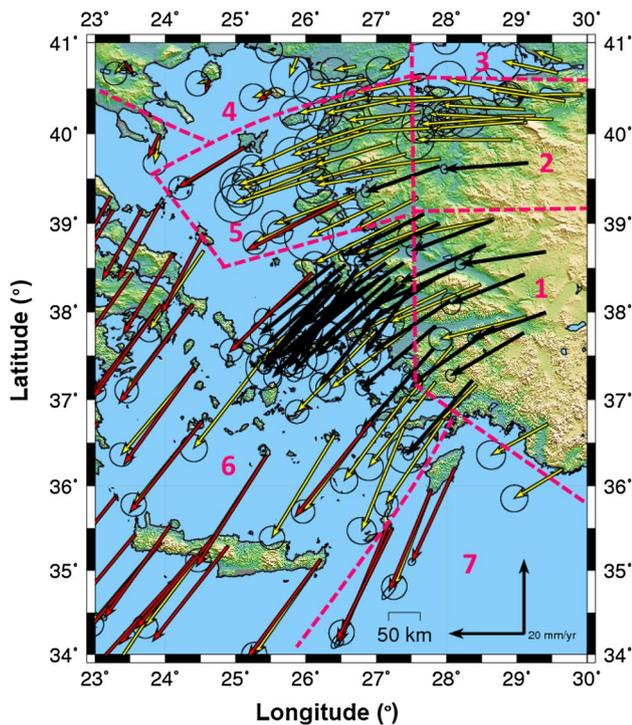


Fig. 4 The views of the microplates created by the differences on the amplitudes and the directions of the GPS velocities given in Fig. 2

of microplate No. 2, gains south component in this area. Probably, the microplate No. 5, when come across a stable mechanism of microplate No. 4, is forced to rotate southward. The main boundary of microplate No. 5 can be seen clearly at the topography and bathymetry (Fig. 5a, b). The N–S trending boundary between the microplate Nos. 2 and 5 is also monitored clearly at the topography (Figs. 4, 5a, b).

The elongation of topographic structures is NE–SW trending in the region which locates at the westward of the N–S trending border (which locates at approximately between 27° and 28° longitudes) was called as region A in this study (Fig. 5b). This NE–SW trending topography of region A may be explained by encountering the movement of the body which comes from eastward with a stable plate and having south component velocity. The microplate No. 6 (Fig. 4) has velocity towards SW, but the south component is more dominant than other microplates and moves like a different plate. The region which locates north of Crete (Fig. 1) representing different bathymetries with respect to its surroundings was called as region C (Fig. 5b), and it includes the effect of the Hellenic Subduction zone. Additionally, it is seen clearly at GPS velocities (Fig. 4) that Hellenic Subduction zone composes a great effect between microplate Nos. 1 and 7. The directions of GPS velocities change approximately in 100 km between Anatolian plate and Rhodes Island (Fig. 4). Besides, the seismic activity (Fig. 6) is very high at microplate Nos. 1, 5 and 7 where the

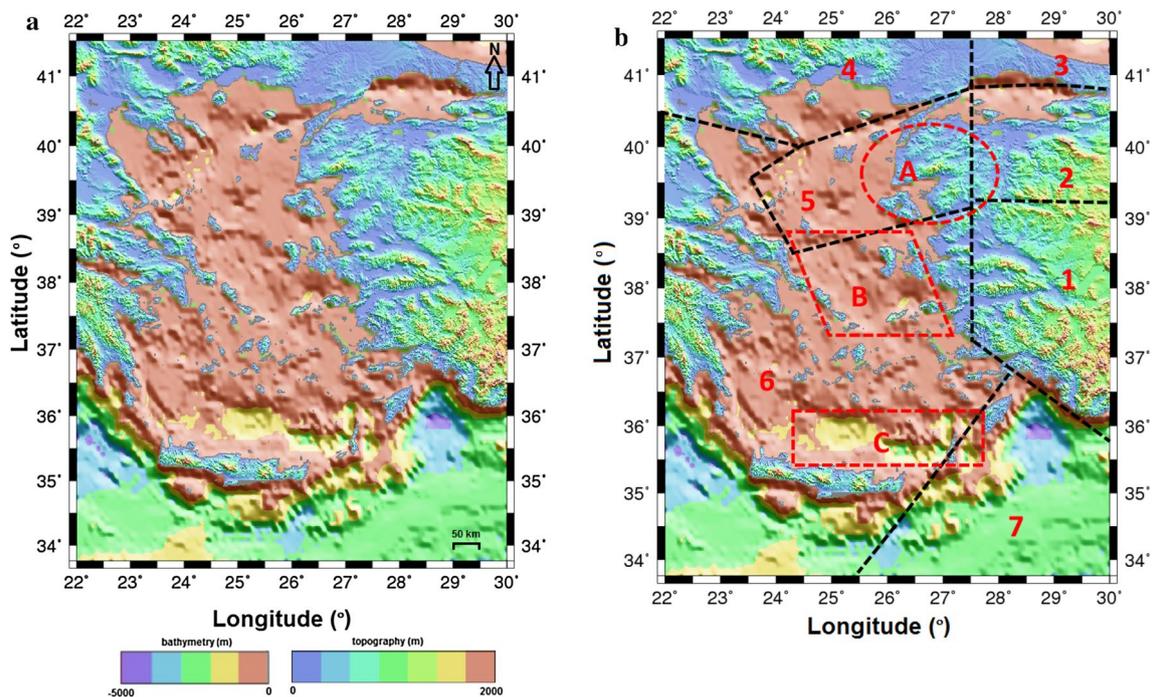
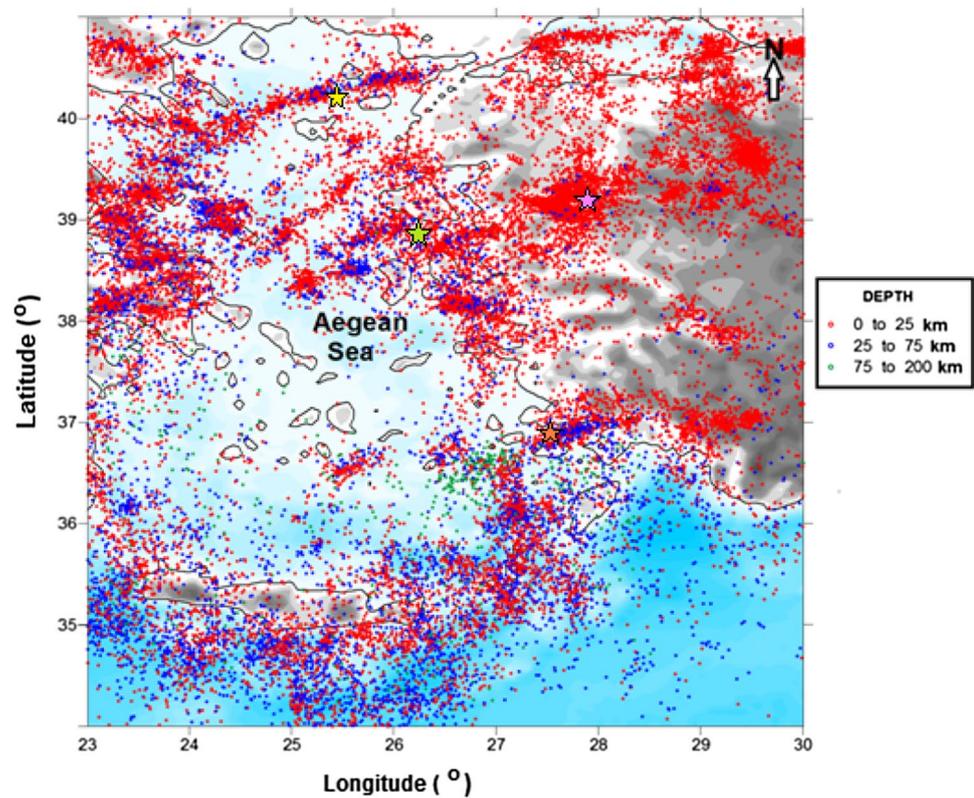


Fig. 5 **a** The topographic and bathymetric map (ETOPO2; NOAA, N. 2006) of the study region. **b** The views of the microplates created by the differences on topography and bathymetry (**a**) related with the

differences on the amplitudes and the directions of the GPS velocities given in Fig. 2. This figure is created by using Generic Mapping Tools (GMT) (Wessel and Smith 1998)

Fig. 6 Topographic map of study area and the earthquake focal depth (km) distributions between the years 1973 and 2017 (from USGS). (Yellow star: Gokceada earthquake, $M_w = 6.5$; green star: Lesvos earthquake, $M_w = 6.2$; orange star: Gokova earthquake, $M_w = 6.5$; pink star = Manisa-Golmarmara earthquake, $M = 5.1$; AFAD) The topographic data were obtained from TOPEX; Smith and Sandwell (1997)



velocity directions and the transition of the mass movements change rapidly.

For interpreting the similarities between the seismicity and the microplate borders of GPS velocities, the earthquakes that occurred between the years 1973 and 2017, with magnitude (M_w) range between 2.5 and 9.0, were taken from US Geological Service (USGS) and the focal depth distribution map is shown in Fig. 6. For instance, the Gokceada earthquake (24 May 2014, $M_w = 6.5$; Republic of Turkey Prime Ministry Disaster and Emergency Management Authority (AFAD)) and Lesvos earthquake (12 June 2017, $M_w = 6.2$; (AFAD)) which occurred in microplate No. 5 (Fig. 4), Gokova earthquake (21 July 2017, $M_w = 6.5$; (AFAD)) which occurred in microplate No. 6 and Manisa-Golmarmara earthquake (27 May 2017, $M = 5.1$; (AFAD)) which occurred at the intersection of microplate Nos. 1, 2 and 5 represent the most seismically active structures of the region (Fig. 6) in the present.

In Fig. 6, the high seismicity represents the borders of the microplates, and also low and high seismicity transitions represent the border between the microplates (Fig. 4). For instance, the GPS vector directions and earthquake distributions of the microplate Nos. 6 and 7 present different characters. Additionally, the earthquakes appear in concentrated along the border between the microplate Nos. 4 and 5. While the borders obtained from GPS velocities (Fig. 4) and earthquakes distributions (Fig. 6)

are evaluated together, it can be said that the intersection area of the microplate Nos. 1–2–5–6 represents high complex structure by noticing the high seismicity. Due to the obtained general views, the vertical motions are also effective in the region. It is known that the earthquakes occur at the plate boundaries in general. It is noticed from GNSS velocities (Fig. 4) that the mass movements are increasing and gaining south components while passing from microplate No. 1 to microplate No. 6. This case causes high deformation in microplate No. 6 as seen clearly in bathymetric changes (Fig. 5b) particularly in a region which was called as region B. Therefore, a new deformation zone (or zones) may be observed in that region in following geologic time scales with the effect of this high deformation.

In the study of Tirel et al. (2004), the regions between North Anatolian Trough and Cretan Sea (Fig. 1) were defined as “deformed areas” and “zones of recent thinning”. The boundary between microplate Nos. 4 and 5 and region C are consistent with the regions which were defined as “zones of recent thinning” by Tirel et al. (2004). The directions of GPS velocities rotate to south, and the amplitudes of the velocities increase in the region located at the south of “zones of recent thinning” (Fig. 4) due to the slab retreat effect of Hellenic Subduction zone. This effect stops rapidly in the boundary between the microplate Nos. 1 and 7 (Fig. 4).

In the study of Bohnhoff et al. (2005) which was achieved with the focal mechanism results in Hellenic Subduction zone, it was explained that the plate boundary mechanism of Hellenic Arc included different angles. The boundary which has this angle differences coincides with the border between the microplate Nos. 6 and 7.

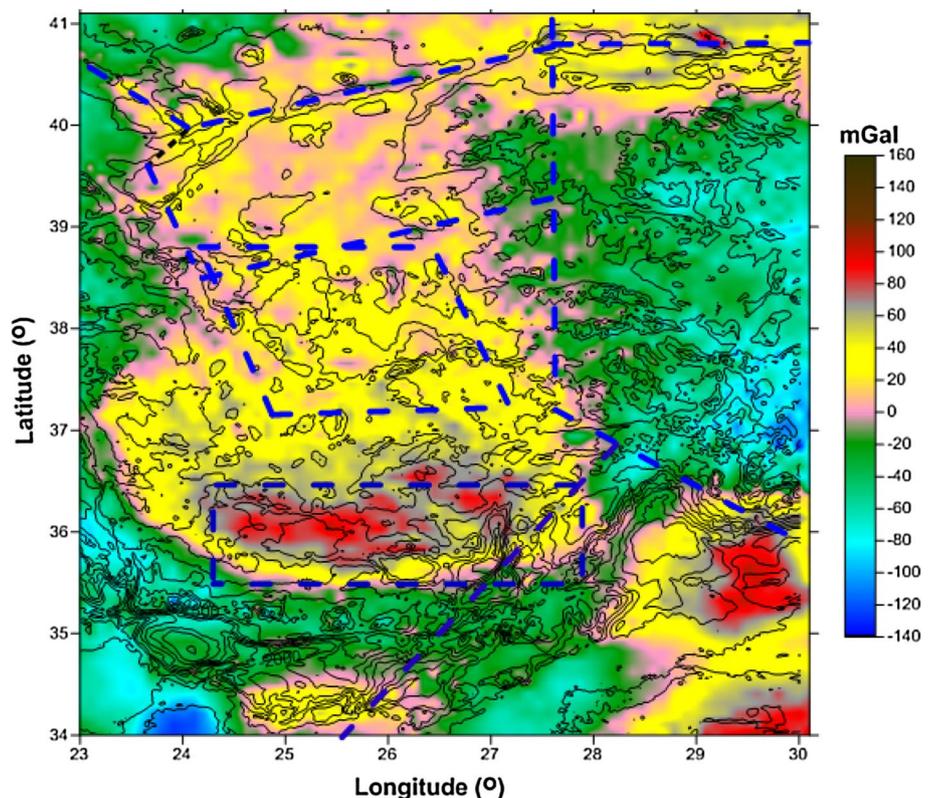
For evaluating the relationship between GPS velocities and the gravity values, the gravity anomaly map (Fig. 3) was divided into microplates (Fig. 7) due to the differences on the amplitudes of the gravity values. If Figs. 4 and 7 were evaluated together, it was determined that the boundaries of kinematic movements and the changes in the mass structures were almost same. Therefore, the directions of the GPS velocity vectors represent differences at the transitions between the tectonic elements which have low and high densities.

In the study region, the lowest amplitude gravity anomalies are seen in the Western Anatolia graben system and its southern part (Fig. 7). There is a transition zone between the longitudes 27° and 28° (Fig. 7) same as GPS velocities (Fig. 4) and earthquake focal depth distributions (Fig. 6). Besides, this transition zone between these longitudes was called as West Anatolia Transfer Zone (WATZ) by Gessner et al. (2013) and this zone was corrected by the Bouguer gravity anomaly study of Dogru et al. (2017). Besides, WATZ forms a boundary which separates west of Western Anatolia from Western Anatolian graben system. If

the Bouguer gravity anomaly map (Fig. 7) is examined, the region containing the shear zone is characterized with negative gravity anomalies on graben system. The, respectively, lower amplitude gravity anomalies are related with the graben systems which contain high sediment thickness or the regions which have high heat flows (Dolmaz et al. 2005) and may be associated with the asthenospheric upwelling and rising temperature (Fig. 7). When Figs. 4, 5b, 6 and 7 were evaluated together, both the boundaries which were obtained from GPS velocities, bathymetry–topography and gravity, represent high seismic activities (Fig. 6) and this knowledge verifies that these boundaries are also the boundaries of the microplates.

In this study, unlike previous studies, the microplates which represented Western Anatolia were distinguished from each other in more detail by using the results of GPS study realized in and around Izmir (Çırmık et al. 2017). In the study of Çırmık et al. (2017), in addition to Eurasia-fixed-frame calculation, the GPS velocities were computed with respect to Anatolian-block fixed frame, and therefore, a boundary which separated the kinematic structure of Izmir from Western Anatolia graben tectonic structure was proposed. That boundary is consistent with the N–S directional border located between Western Anatolian microplate Nos. 1 and 6 (Fig. 4). The tectonic mechanism related with the microplate No. 7 (Fig. 4) is thought to be affected by the interplate geodynamic structure around Isparta angle (Fig. 1)

Fig. 7 The views of the microplates created by the differences on the amplitudes of the gravity values given in Fig. 3



which was defined in the study of Över et al. (2016). Therefore, it is suggested that the continuity of Isparta angle where Mediterranean Sea, Aegean Sea and Western Anatolia intersect needs to be studied.

References

- Ahadov B, Jin S (2017) Present-day kinematics in the Eastern Mediterranean and Caucasus from dense GPS observations. *Phys Earth Planet Int* 268:54–64
- Ambraseys N, Jackson JA (1990) Seismicity and associated strain of central Greece between 1890 and 1988. *Geophys J Int* 101(3):663–708
- Armijo R, Meyer B, King G, Rigo A, Papanastassiou D (1996) Quaternary evolution of the Gulf of Corinth rift and its implications for the Late Cenozoic evolution of the Aegean. *Geophys JR Astron* 126:11–53
- Bohnhoff M, Harjes HP, Meier T (2005) Deformation and stress regimes in the Hellenic subduction zone from focal mechanisms. *J Seismo* 9(3):341–366
- Bonvalot S, Balmino G, Briais A, Kuhn M, Peyrefitte A, Vales N, Biancale R, Gabalda G, Reinquin F, Sarraillh M (2012) World gravity map. Commission for the Geological Map of the World. Eds. BGI-CGMW-CNES-IRD, Paris
- Bozkurt E, Sözbilir H (2004) Geology of the Gediz Graben: newfield evidence and its tectonic significance. *Geol Mag* 141:63–79
- Çırmık A, Pamukçu O (2017) Clarifying the interplate main tectonic elements of Western Anatolia, Turkey by using GNSS velocities and Bouguer gravity anomalies. *J Asian Earth Sci* 148:294–304
- Çırmık A, Pamukçu O, Gönenç T, Kahveci M, Şalk M, Herring T (2017) Examination of the kinematic structures in Izmir (Western Anatolia) with repeated GPS observations (2009, 2010 and 2011). *J African Earth Sci* 126:1–12
- DeMets C, Gordon RG, Argus DF, Stein S (1990) Current plate motions. *Geophys J Int* 101(2):425–478
- Dewey JF (1988) Extensional collapse of orogens. *Tectonics* 7(6):1123–1139
- Dewey JF, Şengör AMC (1979) Aegean and surrounding regions: complex multi-plate and continuum tectonics in a convergent zone. *Geol Soc Am Bull* 90(1):89–92
- Dewey JF, Hempton MR, Kidd WSF, Şaroglu F, Şengör AMC (1986) Shortening of continental lithosphere: the neotectonics of eastern Anatolia—a young collision zone. In: Coward MP, Ries AC (eds) *Collisional tectonics*, vol 19. Geological Society of London, London, pp 3–36
- Dogru F, Pamukçu O, Ozsoz I (2017) Application of tilt angle method to the Bouguer gravity data of Western Anatolia. *Bull Miner Res Explor* 155:45–55
- Dolmaz MN, Hisarlı ZM, Ustaomer T, Orbay N (2005) Curie point depths based on spectrum analysis of aeromagnetic data, West Anatolian extensional province, Turkey. *Pure Appl Geophys* 162:571–590
- Floyd M, Herring TA (2018) e-mail communications with Massachusetts Institute of Technology (MIT), Department of Earth Atmospheric and Planetary Sciences (EAPS)
- Floyd MA, Billiris H, Paradissis D, Veis G, Avallone A, Briole P, McClusky S, Nocquet JM, Palamartchouk K, Parsons B, England PC (2010) A new velocity field for Greece: implications for the kinematics and dynamics of the Aegean. *J Geophys Res* 115:B10403
- Förste C, Bruinsma SL, Abrikosov O, Lemoine JM, Marty JC, Flechtner F, Balmino G, Barthelmes F, Biancale R (2014) EIGEN-6C4 The latest combined global gravity field model including GOCE data up to degree and order 2190 of GFZ Potsdam and GRGS Toulouse. *GFZ Data Serv* 12:15. <https://doi.org/10.5880/ICGEM.2015.1>
- Gessner K, Gallardo LA, Markwitz V, Ring U, Thomson SN (2013) What caused the denudation of the Menderes Massif: review of crustal evolution, lithosphere structure, and dynamic topography in southwest Turkey. *Gondwana Res* 24(1):243–274
- Gilardoni M, Reguzzoni M, Sampietro D (2016) GECO: a global gravity model by locally combining GOCE data and EGM2008. *Stud Geophys Geod* 60:228–247
- Goldsworthy M, Jackson J, Haines J (2002) The continuity of active fault systems in Greece. *Geophys J Int* 148(3):596–618
- Goudarzi MA, Cocard M, Santerre R (2014) EPC: Matlab software to estimate Euler pole parameters. *GPS Solut* 18:153–162
- Herring TA, King RW, Floyd MA, McClusky SC (2015) Introduction to GAMIT/GLOBK, Release 10.6. Massachusetts Institute of Technology, Cambridge
- Hollenstein C, Müller MD, Geiger A, Kahle HG (2008) Crustal motion and deformation in Greece from a decade of GPS measurements, 1993–2003. *Tectonophysics* 449(1):17–40 <https://www.afad.gov.tr/>. Republic of Turkey Prime Ministry Disaster and Emergency Management Authority (AFAD)
- Jackson J, McKenzie DP (1988) The relationship between plate motion and seismic moment tensors, and the rates of active deformation in the Mediterranean and Middle-East. *Geophys J R Astron Soc* 93:45–73
- Kahle HG, Cocard M, Peter Y, Geiger A, Reilinger R, Barka A, Veis G (2000) GPS derived strain field within the boundary zone of the Eurasian, African, and Arabian plates. *J Geophys Res* 105(B10):23353–23370
- Kahveci M, Pamukçu O, Cirmik AY, Gonenc T (2013) Using GPS data together with geophysical data: a case study from a seismically active region, Izmir. In: 6th international conference on recent advances in space technologies (RAST). IEEE, pp 231–236
- Le Pichon X (1982) Land-locked oceanic basins and continental collision: eastern Mediterranean as a case example. In: Hsü K (ed) *Mountain building processes*. Academic, New York, pp 201–211
- Le Pichon X, Angelier J (1979) The Hellenic arc and trench system: a key to the evolution of eastern Mediterranean area. *Tectonophysics* 60:1–42
- Makris J (1978) The crust and upper mantle of the Aegean region from deep seismic soundings. *Tectonophysics* 46:269–284
- Mart Y, Ryan WBF (2003) The tectonics of Cyprus Arc: a model of complex continental collision. *Eur Geophys Soc Geophys Res Abstracts* 5:02282
- McClusky S et al (2000) Global Positioning System constraints on plate kinematics and dynamics in the eastern Mediterranean and Caucasus. *J Geophys Res* 105(B3):5695–5719
- McKenzie DP (1972) Active tectonics of the Mediterranean region. *Geophys J R Astron Soc* 55:217–254
- McKenzie DP (1978) Active tectonics of the Alpine-Himalayan belt: the Aegean sea and surrounding regions (tectonic of Aegean region). *Geophys J R Astron Soc* 55:217–254
- Mercier JL, Sorel D, Vergely P (1989) Extensional tectonic regimes in the Aegean basins during the Cenozoic. *Basin Res* 2:49–71
- NOAA, N (2006). 2-minute Gridded Global Relief Data (ETOPO2) v2
- Nyst M, Thatcher W (2004) New constraints on the active tectonic deformation of the Aegean. *J Geophys Res* 109:B11406
- Över S, Özden S, Kamacı Z, Yılmaz H, Ünlügenç UC, Pınar A (2016) Upper crust response to geodynamic processes beneath Isparta Angle, SW Turkey: revealed by CMT solutions of earthquakes. *Tectonophysics* 687:94–104
- Pamukçu O (2016) Geodynamic assessment of Eastern Mediterranean region: a joint gravity and seismic b value approach. *Arab J Geosci* 9(5):1–13

- Pamukçu O, Gönenç T, Çırmık YA, Kahveci M (2015) Investigation of the Sığacık Bay's displacement characteristic by using GPS and gravity data in Western Anatolia. *J Asian Earth Sci* 99:72–84
- Pavlis NK, Holmes SA, Kenyon SC, Factor JK (2008) An earth gravitational model to degree 2160: EGM2008. *EGU Gen Assem* 10:13–18
- Reilinger R et al (2006) GPS constraints on continental deformation in the Africa, Arabia, Eurasia continental collision zone and implications for the dynamics of plate interactions. *J Geophys Res* 111:B05411
- Reilinger R, McClusky S, Paradissis D, Ergintav S, Vernant P (2010) Geodetic constraints on the tectonic evolution of the Aegean region and strain accumulation along the Hellenic subduction zone. *Tectonophysics* 488:22–30
- Seyitoğlu G, Scott BC (1991) Late Cenozoic crustal extension and basin formation in west Turkey. *Geol Mag* 128:155–216
- Smith WHF, Sandwell DT (1997) Global seafloor topography from satellite altimetry and ship depth soundings. *Science* 277:1957–1962
- Sorel D, Mercier JL, Keraudren B, Cushing M (1988) Le rôle de la traction de la Lithosphère Subductée Dans L'évolution Géodynamique Plio-Pleistocene De l'Arc Egeen: Mouvement Verticaux Alternés Et Variations Du Régime Tectonique. *C R Geosci* 307:1981–1986
- Stein S, Wysession M (2005) An introduction to seismology, earthquakes, and earth structure. Blackwell Publishing, Oxford, pp 290–296
- Taymaz T, Jackson J, Westaway R (1990) Earthquake mechanisms in the Hellenic Trench near Crete. *Geophys J Int* 102(3):695–731
- Tirel C, Gueydan F, Tiberi C, Brun JP (2004) Aegean crustal thickness inferred from gravity inversion, Geodynamical implications. *Earth Planet Sci Lett* 228(3–4):267–280
- Wessel P, Smith WHF (1998) New, improved version of the generic mapping tools released. *Am Geosci Union* 79:579
- Yılmaz Y, Genç C, Gürer F, Bozcu M, Yılmaz K, Karacık Z, Altunkaynak Ş, Elmas A (2000) When did the western Anatolian grabens begin to develop? In: Bozkurt E, Winchester JA, Piper JDA (eds) *Tectonics and magmatism in Turkey and surrounding area*, vol 173. Geological Society of London, London, pp 353–384



Multi-frequencies GPR measurements for delineating the shallow subsurface features of the Yushu strike slip fault

Di Zhang¹ · Jiacun Li² · Shaotang Liu¹ · Guo Wang¹

Received: 19 July 2018 / Accepted: 20 February 2019 / Published online: 25 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

In order to assess Ground penetrating radar (GPR) for imaging the shallow subsurface geometry and characteristics of the fault in Yushu area, details GPR measurements with 25 MHz, 100 MHz and 250 MHz frequencies antenna were firstly conducted in four sites along the Yushu fault after geomorphologic and geological investigation. The 25 MHz profiles delineated an excellent general view of deformation zones at a much wider area and greater depth. While the 100 MHz and 250 MHz data provided more detailed analysis of the shallow subsurface deformation about the geologic structure and the fault, including the stratigraphic structures, the dip angle and direction of the fault plane. The remarkable variation in the pattern and relative amplitude of electromagnetic waveform on the two-dimensional GPR profiles are all obvious and it is considered as the main fault zone with a nearly vertical fault with the dip angle of 70°–85°. High frequency GPR profiles show a good consistency with the trench sections at three sites. The geometry of the main fault zone can be depicted and deduced up to ~12 m deep or even deeper on the 25 MHz GPR profile in Yushu area and it is considered to be the result of the movement of active faults. What's more, the geophysical features on GPR profile associated to the strike-slip fault are further summarized in different geological and geomorphological environment in Yushu area, the study also provides further evidence that GPR is valuable for fault investigation and palaeoseismic study in the Qinghai-Tibet Plateau area.

Keywords GPR · The Yushu fault · Subsurface geometry · Trench

Introduction

Paleoseismology defines the study of identifying the history and pre-historic earthquakes in tectonically active regions all over the world. Paleoseismology study requires detail investigations and the main intention is to offer the crucial physical features of active fault associate to past earthquakes, including precise location, geometries, the orientation, dip angle, fault throw and the ranges of deformation zones (Mccalpin 1996; Reicherter et al. 2012; Wu et al. 2014a). The most general adopted method of palaeoseismic investigation is trenching. However, it becomes difficult to determine the appropriate site when the surface geological and geomorphological records are partly preserved due to

high degree of sedimentation and erosion. In addition, when the geological environment is severe, especially in the Qinghai-Tibet Plateau area, it is expensive and time-wasting for trench investigation with damaging the surrounding environment (Shi et al. 2016; Sun et al. 2017; Yan and Lin 2017). In these cases, several geophysical techniques (integrated with photogrammetric analysis, field surveys) had been carried out to map the shallow subsurface geometry in the vicinity of the fault. Among these geophysical techniques, GPR has been proved to supply a more reliable high-resolution internal geometry of fine structures for paleoseismology research in a non-destructive and cost-effective fashion. In recent years, the GPR has been used to explore and reveal the active fault for different purposes in the literature: (1) locating the optimal trench sites where the geomorphic markers are unobservable or employing complementary information to previous palaeoseismic trench work (Anderson et al. 2003; Grützner et al. 2012; Cahit et al. 2013; Anchuela et al. 2016); (2) imaging shallow subsurface geometry and extracting qualitative and quantitative features of the fault, such as precise location, the orientation, dip angle and fault throw

✉ Di Zhang
zhangdi1987228@163.com

¹ College of Civil Engineering, Henan University of Engineering, Zhengzhou 451159, China

² College of Resource Environment and Tourism, Capital Normal University, Beijing 100048, China

of the fault (Grasmueck 1996; Demanet et al. 2001; Rashed et al. 2003; Reiss et al. 2003; Malik et al. 2007; Pauselli et al. 2010; Ercoli et al. 2013; Maurya et al. 2013; Dujardin et al. 2014; Ercoli et al. 2014; Bubeck et al. 2015; McBride et al. 2015; Lunina 2016; Maurya et al. 2017; Lunina et al. 2018); (3) combining analysis of GPR and trench data (Gross et al. 2002; McClymont et al. 2009; Malik et al. 2010; Salvi et al. 2010; Forte et al. 2015; Lorenzo et al. 2016; Schneiderwind et al. 2016).

Although GPR method have been widely applied to investigate the active fault in different geological environment all over the world, there is scarcely any GPR surveys have been performed on active fault zones with sinistral-lateral strike slip component. It is the first time that we use multi-frequencies GPR antennas to explore and reveal the shallow subsurface geometry and characteristics of active fault in Yushu area. The shallow subsurface geometry of the faults was depicted mostly using the trench based on surface expression and geomorphic markers. In addition, the Yushu area locates in Qinghai-Tibet Plateau area and has a severe natural environment with water-saturated clay in near-surface layer, where the electrical conductivity constitutes a challenge for radar wave propagation. The main objective of the study is to demonstrate of GPR to characterize the shallow subsurface deformation about geologic structure and behaviour of the Yushu fault. After preliminary geomorphologic investigation, different frequencies GPR profiles have been performed at four sites along the Yushu fault to delineate shallow subsurface nature for evaluating the seismic activity in the Yushu area. Several two-dimensional GPR profiles were recorded perpendicular to the fault using the 25 MHz antenna to provide a general skeleton map of subsurface features at a much wider area and greater depth. After an in-site processing, we remarked the distortion and deformation zones with remarkable change in the amplitude and pattern of radar reflections in the 25 MHz radargraphs. The deformation zones in 25 MHz profiles were scanned again with the 100 MHz and 250 MHz antenna to obtain a better lateral and vertical resolution and more detailed analysis for depicting subsurface geometry. At last, in order to assess the interpreted multi-frequencies GPR results, the GPR profiles were compared with trench sections at three study sites, which were excavated by the previous researches in the Yushu area.

Geological setting

Since the late Miocene or so (50–55 M_a), the collision between the Indian and Eurasian plates caused the tectonic movements and ground uplifts in the interior of the Qinghai-Tibet Plateau (Tapponnier et al. 2001; Gan et al. 2007; Xu et al. 2015). With the northward accumulated squeezing by Indian plate and rigid

obstructed in north, west and north-east along the Qinghai-Tibet Plateau, the tectonic blocks are squeezed towards the eastern and slipped along some border fault zones in the Qinghai-Tibet Plateau, such as the East Kunlun fault zone, the Altyn fault zone, the Yushu-Garzê-Xianshuihe fault zone and the Jiali fault zone, which are responsible for the occurrence of earthquakes above magnitude 7.0 in the Qinghai-Tibetan Plateau (Wang et al. 2013) (Fig. 1a).

The Yushu-Garzê fault is a sinistral-lateral strike slip fault that has been high activity in the late Quaternary and is one of the most active segments in the east and north boundary of the Sichuan-Yunnan Rhombic Block. It extends for at least 500 km long that start from the Garzê City in Sichuan Province, traversing the northeast of the Yushu City, and ends at Zhiduo City in Qinghai Province (Fig. 1a). It has a large scales and the bedrock fracture ranges from 10 m to a few hundred meters, and it is composed of three main fault zones from northwest to southeast, which are the Dangyang-Duocai fault zone, the Yushu fault zone and the Garzê fault zone (Zhou et al. 1997a; Wen et al. 2003; Wu et al. 2014b; Shi et al. 2016).

The Yushu fault zone, situated in middle portion of the Yushu-Garzê fault zone, has the length of approximately 150 km and shows a NW 120°–130° direction (Fig. 1b). There are a few offset stream and gullies, pull-apart basin, fault facets, river terrace, surface ruptures along the fault zone (Fig. 2). It is stated that the tension effect is weakening while the pushing effect is gradually increasing from west to east along the fault based on the variation of the pull-apart basins scales (Busby and Merritt 1999; Tobita et al. 2011; Pei et al. 2013; Zhang et al. 2016). It lies in the plate marge of the Sichuan-Yunnan Rhombic Block and has high levels of seismic activity when compare with the center areas in this plate, which is responsible for the occurrence of large and moderate magnitude earthquake (Fig. 1b). Based on the rich historical records, at least three earthquakes of 6.5 magnitude or greater have been occurred in the Yushu area including 1738 magnitude 6.5 earthquake hit in northwest Yushu area (Zhou et al. 1997b), a strong magnitude 7.0 Shiqu earthquake happened in 1896 and the latest earthquake with an intensity of 7.1 occurred in 2010 causing serious damage (Chen et al. 2010a, b; Ma et al. 2010). Thus, it is crucial to conduct the fault investigation to obtain a further understanding of the seismic potential and seismic hazards in Yushu area.

Methodology

GPR data collection

The resolution and the depth range of EM-wave pulses normally are dominated by the electrical impedance of the

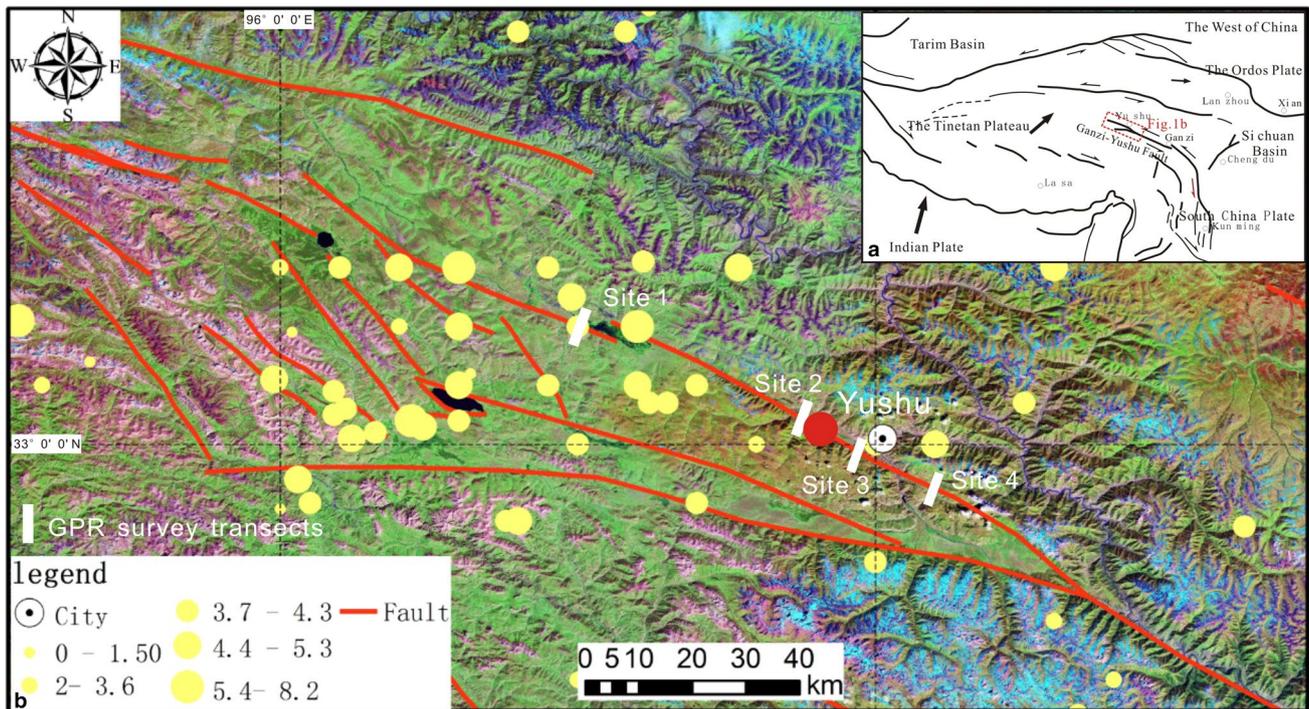


Fig. 1 **a** Simplified geological map of western china and the location of study area (Li et al. 2013; Zhou et al. 2014). The red rectangle area represents the Yushu fault zone, which situated in middle portion of the Yushu-Garzê fault zone. **b** Satellite image of study area and GPR survey sites. Red lines show the Yushu fault with the sinistral-lateral

strike slip component in Qinghai-Tibet Plateau (Li et al. 2014). The yellow dots show the historical earthquakes happened in Yushu area. The red dot is the macroscopic epicenter of 7.1 magnitude earthquakes occurred in 2010. White rectangular boxes are the locations of the GPR survey transects

subsurface materials (such as grain size distribution, porosity and water content) and the frequency of radar antenna (Jol 2009). High frequency GPR signals tend to provide better resolution at the expense of reducing the probing depth, electromagnetic energy attenuation will be increased at depth. In this study, the GPR data were recorded using a RAMAC GPR system equipped with 25 MHz, 100 MHz Rough Terrain Antennas (RTA) and 250 MHz shielded antenna oriented perpendicular to the fault (Fig. 3). The 25 MHz RTA was chosen to optimize the penetration depth and obtain general view of subsurface deformation in the vicinity of the fault, while the 100 MHz RTA and 250 MHz shielded antennas provided higher lateral resolution and more detailed analysis for identifying shallow subsurface structure of the fault.

The mode of the deployment acquisition was the common-offset configuration using a single transmitting and receiving antenna moved over at a fixed distance. The 25 MHz and 100 MHz profiles were released along the survey lines using the keyboards mode (Fig. 3a, b), while the 250 MHz data was performed in two-dimensional continuous reflection images using a calibrated odometer fixed on the survey-wheel (Fig. 3c). The acquisition parameters are listed in Table 1.

Along with the GPR image was obtained, GPS data of each trace in two-dimensional profile were acquired to keep track of the accurate position in the Real-Time Kinematic (RTK) mode (Tanajewski and Bakuła 2016). The Trimble R8 GPS unit can be intercommunicated to the GPR unit and laptop by the serial com-port or USB port. When working with the 250 MHz shielded antennas, the GPS antenna is mounted directly on the middle position between the transmitter and receiver antennas. The measurement points of the GPR are in agreement with GPS position (Fig. 3c). However, the measurement points are different when the GPS and GPR antenna is separated, especially is the RTA. In this case, the GPR measurement point is the mid-position of the transmitter and receiver antennas while the GPS antenna is situated on the top of the backpack carried by the gatherer (Fig. 3a, b), so it is crucial for correction of initial position between GPR and GPS data. The combination of GPR and GPS will give the chance to easily handle accurate spatial position for data processing, visualization, and also integrate GPR data with seismogeologic and neotectonic studies. What's more, the elevation data along the GPR survey lines were collected for topographic correction.



Fig. 2 Photomosaic showing the general geomorphology of the Yushu fault. **a** Surface subsidence. **b** Surface rupture. **c** Fault facets. **d** River terrace

GPR data processing

For a first analysis of the GPR images, radar signals were processed and interpreted using Reflexw 7.2 software. Radar signals processing was performed to reduce the noise in the raw images, overcome energy decay and improve the recognition of the useful information from GPR profiles. The sequence was a relatively standard procedure. After decreasing the initial direct current component, we used time-zero correction to remove high amplitude reflections at the uppermost, then automatic gain control (AGC) was employed to compensate for the amplitude losses due to the effect of signal attenuation and geometrical spreading losses. Bandpass filters were applied to cut off exceedingly high and low frequencies noise signals. The signal noise was suppressed and reduced after running average filter. Before the final layout, topographic correction was applied to correct the two-way travel time of the traces to a flat datum level in a vertical sense using the elevation changes recorded by DGPS. To obtain more realistic shallow subsurface structures, the velocity of the electromagnetic wave should be evaluated in the study sites. The average velocity of 0.07 m/ns has been calculated at different study sites by analyzing diffraction hyperbolae in the GPR data, which employed for time-depth conversion. In view of the principle of F-K migration, we could get the most optimum velocity by

analyzing the diffraction hyperbola with a range of different velocities, but this value is representative only of one stratigraphic media (Lehmann and Green 2000; Dujardin and Bano 2013). Figure 4 presents a portion of the 500 MHz GPR profile with F-K migration ranging different velocities from 0.04 to 0.08 m/ns with a 0.01 m/ns increment. The diffraction hyperbola was not completely collapsed in the two left figures, while it is over-migrated and defocusing in the right two ones (red rectangular box). The feature of the diffraction hyperbola is correctly focused on a single point with the velocity of 0.06 m/ns and the value is considered to be the most appropriate subsurface velocity at the site 1.

GPR results

Site 1

Site 1 is located to the west of Longbao Lake about 70 km away from Yushu city as showed in Fig. 1b (N33.21°, E96.48°). The surface expression and geomorphic markers show a prominent pressure ridge exists in the loose sediments and the evidence of the active fault in Fig. 5a. This provides a good chance to evaluate the reliability of GPR method to depict shallow geologic structure about the fault zone in the depositional setting. The SW–NE GPR profile



Fig. 3 The composition of the GPR acquisition system and a differential GPS receiver mounted on the top of GPR antenna or the backpack. The GPR system is mainly composed of the center control

unit, the transmitting and receiving antenna, the distance measuring unit and a laptop. **a** The 25 MHz RTA. **b** The 100 MHz RTA. **c** The 250 MHz shielded antenna. **d** The base station of differential GPS

Table 1 Details of the acquisition parameters

Frequency (MHz)	25	100	250
Trace interval (m)	0.3	0.1	0.05
Samples	500	512	488
Sampling frequency (MHz)	285	925	3061
Stacks (times)	128	128	8
Time window (ns)	800	300	160

was acquired perpendicular to the pressure ridge using the 25 MHz RTA in the continuous mode. On the processed radargraphs, the left and right portions of the profile are all characterized by the high-amplitude radar reflections which indicate the sedimentary deposition. On the contrary, the middle part of the profile (a distance of 45–60 m and 0 to ~9 m in depth) shows the low-amplitude reflections where in the brown rectangular box (Fig. 5b). The waveform and the drastic variation of radar reflections are pretty obvious when crossing the middle portion. What's more, the pattern of radar reflections is dominated by diffraction hyperbolas at the uppermost part of the horizontal distances 52–54 m (red

arrows in Fig. 5b), which are on account of surface ruptures. In consequence, the prominent change of high-amplitude and low-amplitude radar reflections is considered as the fault zone with a nearly vertical fault plane. In order to show more detailed structures of shallow subsurface, the deformation zone was scanned again with the 100 MHz and 250 MHz antenna to obtain a better lateral and vertical resolution and more detailed analysis.

Figure 5c, d show the GPR profiles taken with the 100 MHz and 250 MHz antenna. The interrupted continuous reflections are interpreted as the fault zone between ~19 and 21 m on the 250 MHz GPR profile, which almost the same features as revealed for the 25 MHz dataset. These signatures have the more apparent features of weak amplitudes reflections because of a drastic dielectric permittivity contrast. Due to differences in vertical exaggeration between two frequency images, the fault zone appears to have different geometry structure. Compared with 100 MHz and 250 MHz data, the more detail discernment of the radar reflections allows increased the demarcation of the fault zone on 250 MHz GPR profile, including stratigraphic units and the dip angle of the fault plane. In particular, the reflectors

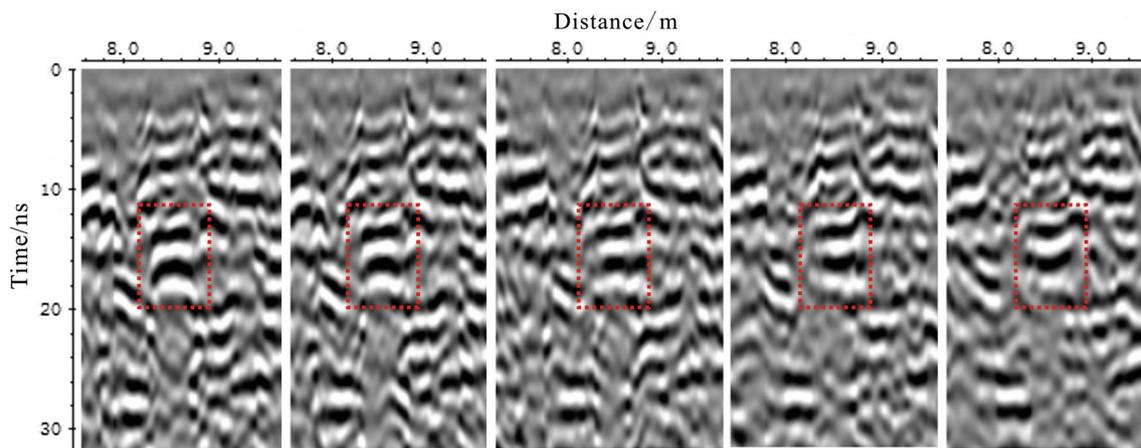


Fig. 4 Detail area of the diffraction hyperbola (at 8–9 m horizontal distance) for different velocities ranging from 0.04 m/ns (on the left) to 0.08 m/ns (on the right) with a 0.01 m/ns increment. The middle

figure shows the correct velocity of F-K migration with $v=0.06$ m/ns at the site 1

of stratigraphic units on the 250 MHz data are more visible than the 25 MHz and 100 MHz data, there are four different continuous radar reflections (blue, red, yellow and green dotted lines in Fig. 5d) on the 250 MHz GPR data. These continuous reflections represent the interface of the different stratigraphic units attributed to lithological distinction.

To have a proper understanding of the GPR profiles, an 8 m long and about 2 m deep trench T1 was excavated in 2014 (Wu et al. 2014b), which location in Fig. 5a. Detailed logging of the western trench wall shows four different stratigraphic units and a main deformation zone. Compared with the trench wall, the GPR profiles are in well matching the trench wall exposure, especially the 250 MHz data. The rather flat radar reflections on the upper of the GPR data (blue and yellow dotted lines in Fig. 5d) are almost perfectly aligned with the location where the grey-yellow and orange-grey units interface. The middle part of the trench wall (black) is related with the low-amplitude radar reflections on the GPR profiles. However, subject to the depth of the trench, it is not possible to link the deeper units and ruptures with the GPR data.

Generally, a main fault zone with the NE–SW dipping is clearly identified by the drastic variation of radar reflections on the different frequencies GPR data at site 1. The horizontal reflections are bent when crossing the fault zone. The geometry of the main fault zone can be depicted up to ~9 m deep or even deeper on the 25 MHz GPR profile and it is considered to be the result of the movement of active faults. The flank of the main fault zone corresponds to the location of the pressure ridge at site 1 (Fig. 5a). The 100 MHz and 250 MHz antennas provide more detailed geometry of the main fault zone, which has strike-slip geometry and a NE–SW dip at nearly 90°, especially in the 250 MHz image. Furthermore, the four flat radar reflections on the 250 MHz

GPR data also give the vertical offset in different stratigraphic units.

Site 2

Site 2 is located the northwest of Jiegu town (N33.06°, E96.85°), where in the vicinity of the macroscopic epicenter of the 7.1 magnitude earthquake occurred in 2010. In the field, the surface evidences of the fault are clear and a series of “small tent” can be easily traced along the fault (Fig. 6a). An 85 m long GPR profile was recorded by the 25 MHz RTA in the common-offset reflection mode along NE–SW directing measurement line. Between the horizontal distances 30–60 m, a sharp variation in the pattern of radar reflections is recognized from top to bottom on the processed GPR profile (brown rectangular box in Fig. 6b). This presence of irregular wavy radar reflections is in consistent with the location of geological and geomorphological markers on the surface. In addition, the right part of the GPR profile is characterized by chaotic strong reflectors from upper to ~10 m deep and the left part is dominated by the rather flat reflections at the depth of 2–5 m. The drastic variation in the radar reflection patterns at the distance of 48 m is considered as the fault plane. For revealing more detailed features of shallow subsurface deformation, the 100 MHz RTA and 250 MHz bow-tie antenna were chose to collected again in the transformation and deformation area in 25 MHz GPR data.

In the GPR images obtained by 100 MHz and 250 MHz antenna, there is an abrupt variation in the waveform and relative amplitude of electromagnetic reflection depicting at a distance of ~21 m (Fig. 6c, d), especially in the 100 MHz GPR image. Trench T2 was dug in 2014 on the edge of the alluvial fan by other researchers, it is around 8 m long

Fig. 5 **a** View of the site 1 located west of Longbao Lake. White, yellow and red lines are 25 MHz, 100 MHz and 250 MHz GPR survey lines, red arrows are surface rupture and white arrow is trench T1. **b** The processed 25 MHz GPR profile. Highlighted area (brown rectangular box) indicates the anomalous zone at the distance of 45–60 m. **c** The processed 100 MHz GPR profile. The continuous horizontal reflection is flexure and terminal at the horizontal distance ~19–21 m. Different colours dotted lines show different continuous radar reflections related to a drastic dielectric permittivity contrast and red arrow indicates the surface rupture. **d** The processed 250 MHz GPR profile. Different colours lines represent different rather horizontal radar reflections and red arrow shows the surface rupture. **e** Photomosaic of trench T1 (the western wall). Trench T1 is about 2 m depth at the site 1, so some units cannot be seen in the trench as deep as the GPR profiles. Four main different stratigraphic units are identified in the trench wall by highlighted here by brown, grey, yellow and orange colours. The brown is a thin recent soil at the top of the trench, the grey is a greyish unit with fine material, the yellow is the yellowish coarse deposits unit (centimetric sized) and the orange is Holocene sediments unit. The black zone, in the middle of trench wall, is considered to be the result of the movement of active faults

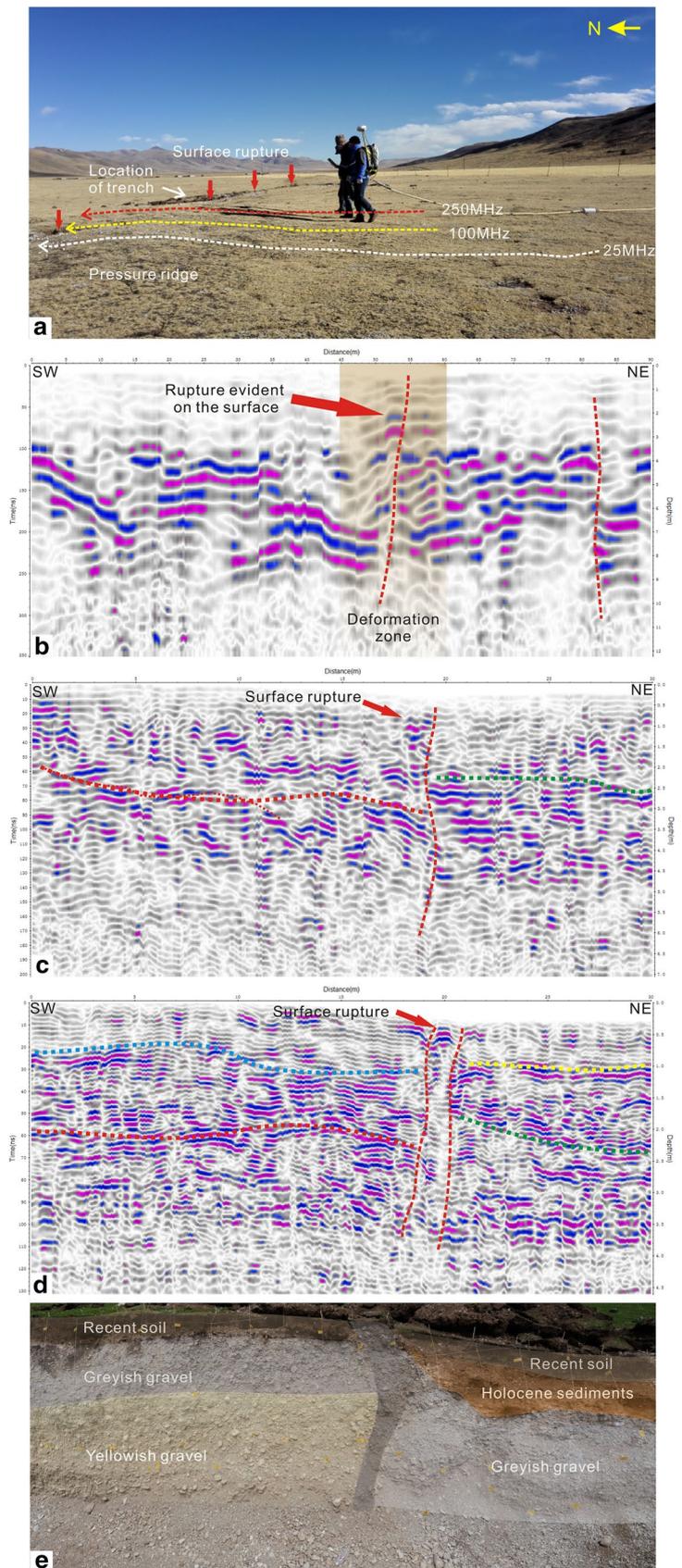
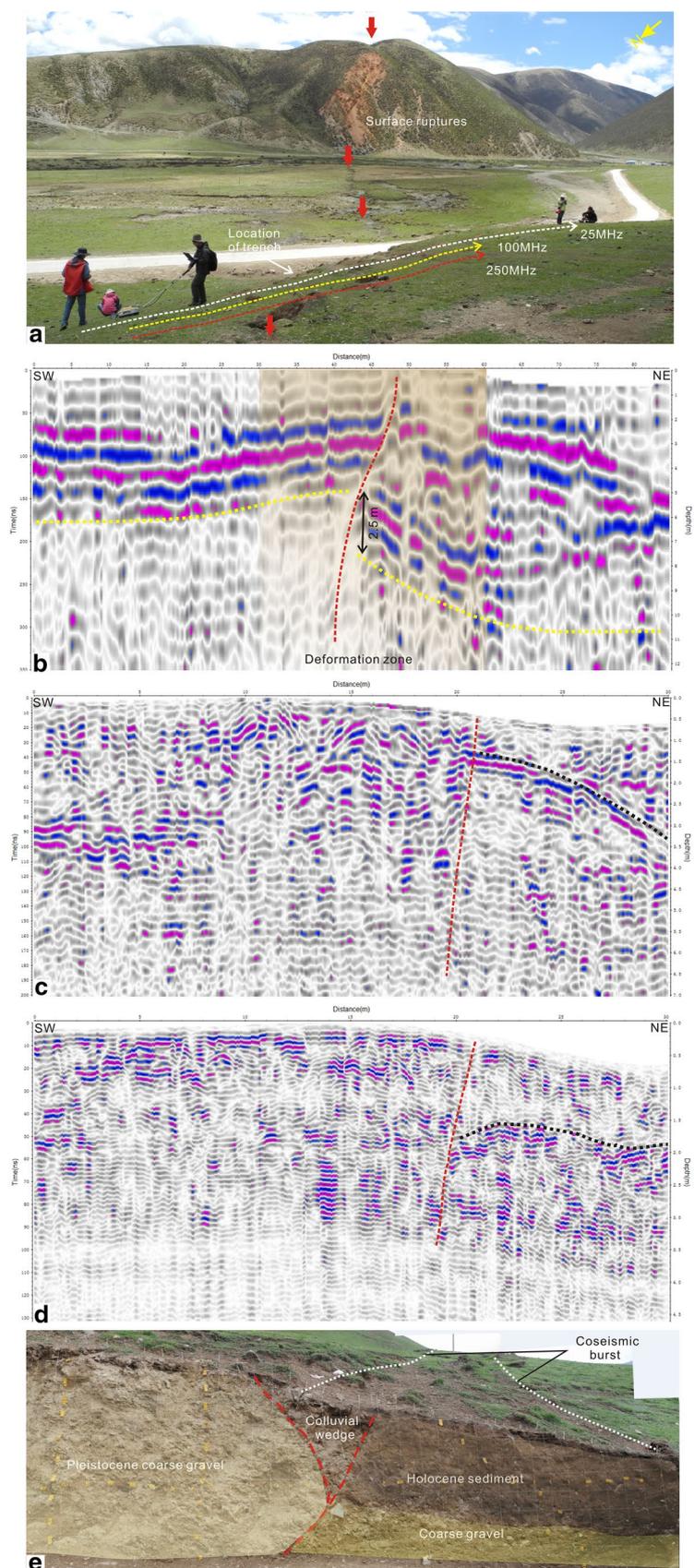


Fig. 6 **a** View of the site 2 located the northwest of Jiegu town. White, yellow and red lines are 25 MHz, 100 MHz and 250 MHz GPR profiles, red arrows are surface ruptures and white arrow is trench T2. **b** The processed 25 MHz GPR profile, highlighted area (brown rectangular box) indicates the deformation zone at the horizontal distance 30–60 m and 0–10 m in depth, there is a sharp variation in the pattern of radar reflections. **c** The processed 100 MHz GPR profile. There is an abrupt variation in the waveform and relative amplitude of electromagnetic reflection at the horizontal distance ~21 m. The black line represent the continuous radar reflections and red line indicates the fault plane. **d** The processed 250 MHz GPR profile. The same interpreted results to the 100 MHz data. **e** Photomosaic of the trench T2 (the western wall). White dashed lines represent coseismic burst of the 7.1 magnitude earthquakes occurred in 2010. Three main different stratigraphic units are identified in the trench wall by highlighted here light yellow, brown and yellow colours. The light yellow unit is the coarse gravel and coarse sand material with some centimetric gravel inside it. The brown unit is the Holocene sediment with fine material and the yellow unit is the yellowish coarse deposits with many scatter stones. The middle portion of trench wall is the colluvial wedge (red lines)



and 2 m deep (Wu et al. 2014b; Zhang et al. 2014). Three stratigraphic units are distinguished in trench wall and highlighted by the light yellow, brown and yellow colours shown in Fig. 6e. Though some units cannot be seen in the trench as deep as the GPR profiles, the 100 MHz and 250 MHz GPR profiles exhibit well consistency with the trench wall. The radargraphs show relatively high-amplitude reflection at a distance of 0–21 m and 0–5 m in depth, which is consistent with in the Pleistocene coarse gravel unit (light yellow colours) on the trench wall. At a distance of 21–30 m, the upper of the radar reflections are characterized by low-amplitude corresponding to the Holocene fine-grained sediments (brown colours), while the bottom part of high-amplitude radar reflections is well matching coarse gravel deposits with many scatter stones (yellow colour). The interface between the brown and yellow units is similar to the rather flat reflection indicated by black dashed lines in the 100 MHz and 250 MHz GPR profiles. The boundary of the high-amplitude signal and low-amplitude signals in the middle part of GPR profiles is regarded as the fault plane (see red dashed lines).

To sum up, two different types of radar reflections are observed on GPR profiles at site 2 (high-amplitude reflections and low-amplitude reflections), and the border is interpreted as the fault plane with the dip in NE–SW directing. The radar reflections, on the left part of the 25 MHz GPR data, are clear lateral continuity with a slightly upward bend when meeting the fault plane. While the radar reflections tend to the downward bend on the right part. This is also well certified by the exposed trench wall section. What's more, the boundaries of the radar reflection (yellow lines in Fig. 6b) illustrate a ~2.5 m vertical offset and it is indicated that the fault has strike-slip component.

Site 3

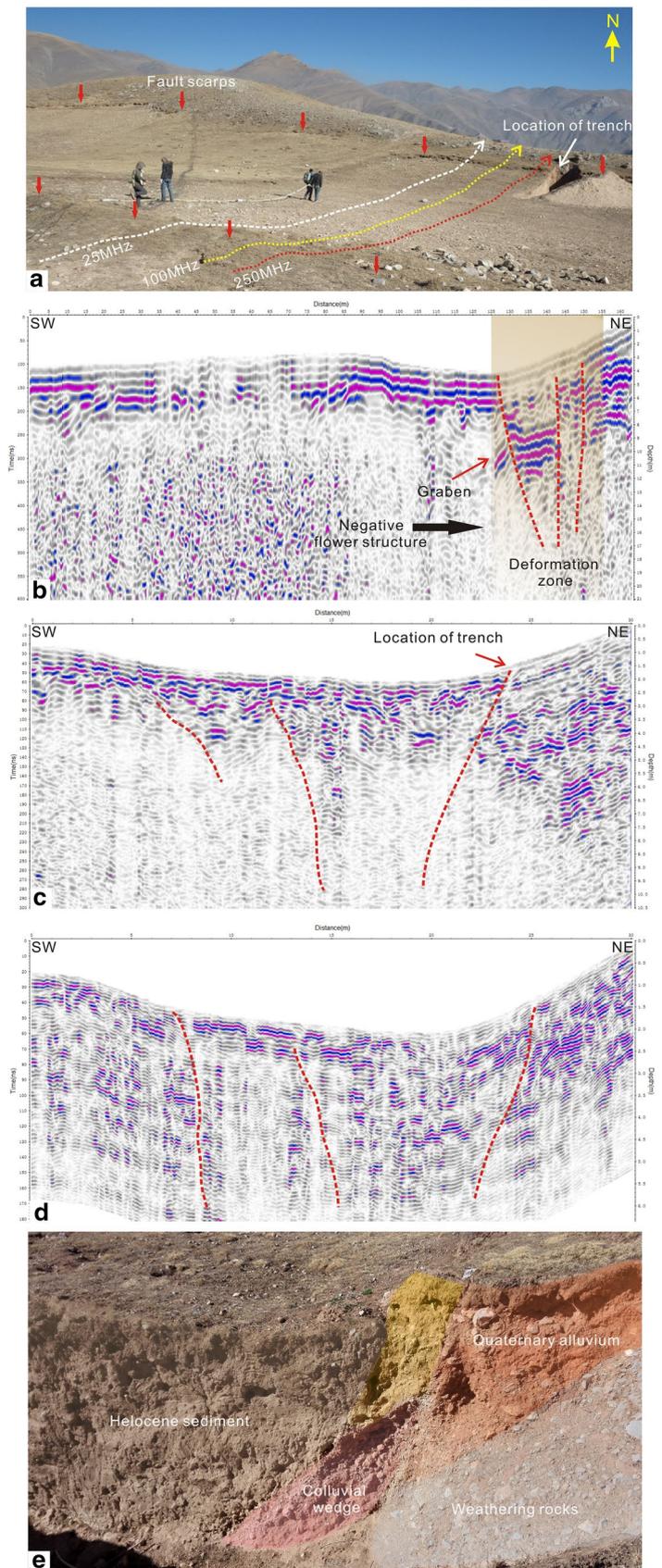
Site 3 is located to the west of Minzhu village (N32.96°, E96.99°), about 2 km away from Yushu city. The transect was selected as the study site for delineating the shallow subsurface geometry of the Yushu fault based on the surface subsidence and small outcrops with the height of ~25–30 cm (Fig. 7a). The SW–NE trending profile was performed crossing the deformation zone with the help of 25 MHz RTA antenna. On the processed data, we remarked a distortion and deformation zone between 125 to 155 m showed in Fig. 7b (brown rectangular box), and the boundaries of the deformation zone is well aligned with the surface subsidence. The portion of the profile at a distance of 0–125 m and 0 to ~4 m in depth is characterized by the high-amplitude continuous reflections relation to the homogeneous material in shallow subsurface. There is an abrupt change in the pattern of radar reflections illustrated at a distance of ~125 m and ~143 m. The continuous horizontal reflections are truncated when passing the deformation zone associated to the

waveform and relative amplitude of these signatures has a drastic variation on account of sharp lithological contrast. Moreover, the top part of the deformation zone is dominated by the relative moderate amplitude reflections. Below it, strong chaotic reflections are obviously noticed at the depth of ~8–12 m. Considering the field setting of the region and the pattern of the radar reflections, the deformation zone is interpreted as the negative flower structure at a distance of 125–145 m and 0–12 m in depth attributed to the strike-slip fault with normal movement. The prominent contacts of high-amplitude and low-amplitude reflections at a distance of ~125 m, ~143 m and 150 m are inferred as the fault plane which is expressed as the roughly vertical.

The 2D 100 MHz and 250 MHz profiles, realized perpendicular to the surface subsidence zone, show distinctly change at a distance of ~13 and ~25 m (Fig. 7c, d), especially in the 100 MHz GPR profile. At the distance of ~25 m, the upmost part of 0 to ~2.5 m in depth shows high-amplitude continuous horizontal reflections in the south portion of the GPR profile, while the strong chaotic reflection is seen at the depth of ~2.5–8 m in the north portion. As a result, the remarkable variation in the pattern and relative amplitude of electromagnetic waveform is considered as the fault plane (see red lines). The trench was excavated by other researcher and the GPR results show a good spatial correlation with the trench section (Fig. 7e). Three main stratigraphic units are identified in the trench wall by highlighted by brown, orange and grey colours. The brown unit is homogeneous sediments with fine material matching with high-amplitude continuous horizontal reflections at a distance of 0–25 m and 0–2.5 m in depth on the GPR profile. The orange unit is the Quaternary alluvium with some centimetric gravel inside it related to the relative moderate amplitude reflections at the top of the north portion. The weathering rocks (grey colour) are easily identified by the strong irregular or chaotic radar reflection at a distance of 25–30 m and ~2.5–8 m in depth. The middle part of the trench wall is colluvial wedge which correspond to intense change at a distance of ~25 m on the GPR profiles. However, duo to the nature of GPR, the fine structure of colluvial wedge is difficult identified on the profile. In addition, the exposed trench wall only shows the right boundary of the deformation zone controlled by the length of trench, which is well aligned with the 25 MHz, 100 MHz and 250 MHz GPR profiles in this site.

Thus, the geometry of a negative flower structure can be clearly observed on the 25 MHz GPR profile based on the pattern of radar reflections at a distance of 125–155 m and 12 m in depth, which is considered to be the movement of the main fault zone. The flanks of the negative flower structure are related to the width of the surface subsidence and the fault strikes can be inferred at deeper than 12 m in the main fault zone. In addition, there is a buried graben in the negative flower structure, which varies in width from ~12 m

Fig. 7 **a** View of the site 3 located to the west of Minzhu village. White, yellow and red lines are 25 MHz, 100 MHz and 250 MHz GPR profiles, red arrows are the fault scarps. The trench T3 is shown by white arrow. **b** The processed 25 MHz GPR profile, highlighted areas (brown rectangular box) indicates the location of deformation zone at the horizontal distance of 125–155 m. **c** The processed 100 MHz GPR profile. There is an abrupt variation in the waveform and relative amplitude of electromagnetic reflection at the horizontal distance of ~13 m and ~25 m. Red line indicates the fault plane. **d** The processed 250 MHz GPR profile. The same interpreted results to the 100 MHz data. **e** Photomosaic of the trench T3 (the western wall). Three main different stratigraphic units are identified in the trench wall by highlighted here brown, orange and grey colours. The brown unit is Holocene sediments with fine material, the orange unit is the Quaternary alluvium with some centimetric gravel inside it and the gray unit is the weathering rocks with many scatter stones. The middle portion of trench wall is colluvial wedge indicated by red colour



at the 8 m depth to ~5 m at 12 m depth. This provides evidence that the fault has strike-slip component and the graben may be associated with the main strike-slip fault plane.

Site 4

Site 4 is located in the east of Changu Temple (N32.93°, E97.06°). The offset ridges and the rivers are unambiguous with left-lateral dislocation through field observations. In addition, the fine geomorphic marks are easily traced in the study site show in Fig. 8a (red arrows). The SW–NE GPR data employing the 25 MHz RTA was realized in point mode to expose a general skeleton map of subsurface features. A high-amplitude reflections zone is showed at the distance of 60–140 m and 0–25 m in depth on the profile in Fig. 8b (brown rectangular box), which is in contrast to the nature structure of the radar reflections at each side of the profile. What's more, the inclined radar reflections are identified from the top to bottom in deformation zone showed by red arrows in Fig. 8b. The flanks of the deformation zone can be clearly defined by the dipping continuous reflections (red line at the right part) and the variation of high-amplitude and low amplitude (red line at the left and right part). Furthermore, the dipping radar reflections indicate an unambiguous increase in dip from northeast to southwest across the deformation zone at the distance of 110–140 m, providing further evidence that the dip of the fault plane is NE–SW directing.

In the processed GPR images obtained by 100 MHz and 250 MHz antenna, a significantly distortion and deformation zone of high-amplitude radar reflections is observed whose thickness increases southward (Fig. 8c, d). The drastic variation of high-amplitude and low-amplitude is the signature of the interface of different media (red lines in Fig. 8c, d). The left portion of the profile is characterized by the high-amplitude continuous and wavy radar reflections consistent with Quaternary alluvium at the site. While the right portion of the profile is mainly dominated by low-amplitude reflections relate to basement soil in the shallow subsurface. These high-amplitude reflections are from the Quaternary cover over basement soil from 3 to 4 m. Moreover, there is a remarkable elevation change on the 100 MHz and 250 MHz profiles when the antennas crossing the surface mark (red arrows in Fig. 8c, d).

Finally, combining the geomorphological and geological setting at this site, the pattern of the reflections is identified as a negative flower structure by the strike-slip fault with normal movement on the 25 MHz GPR profile. The dipping continuous reflection is one flank of the negative flower structure in Fig. 8b, which correspond to the recovered result from the interface between the high-amplitude and low-amplitude on the 100 MHz and 250 MHz GPR sections. There are also some high-amplitude reflections in the

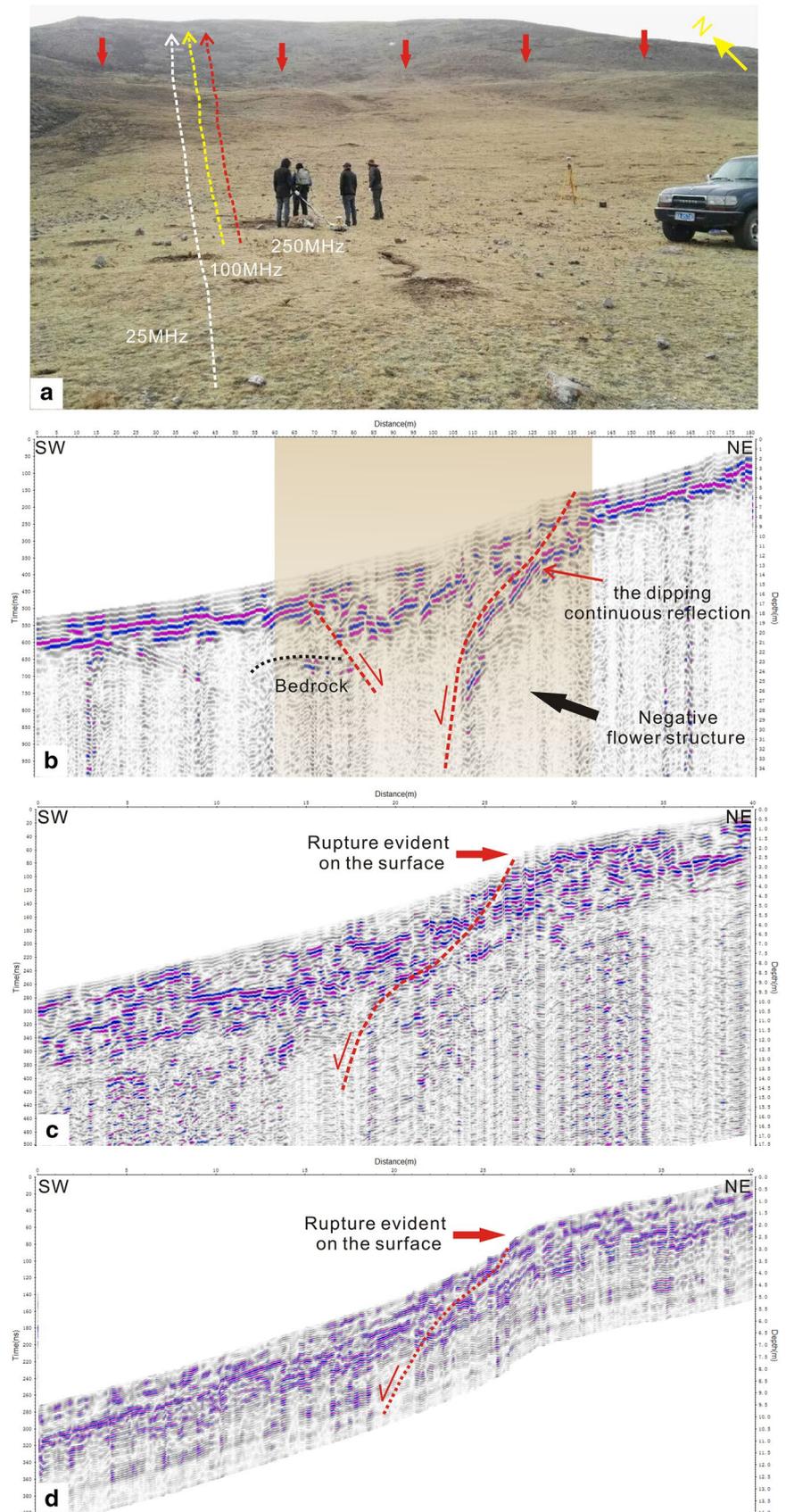
negative flower structure and it is may be the disorganized filling materials by the movement of the main fault.

Discussion

In general, the geophysical feature on GPR profile associated to the well-layer sedimentary deposition is the continuous horizontal reflections. The continuous reflections would be flexure or terminal when encountering to the fault zone. The fault and associated structures are mainly identified by three criteria in the literature: (1) interruption or off-fault in the continuity reflecting layers; (2) diffraction hyperbolas at each side of the fault planes; (3) the drastic variations in the waveform and relative amplitude with the surrounding medium (Busby and Merritt 1999; Christie et al. 2009; Cahit et al. 2013; Maurya et al. 2013; Lunina et al. 2016). However, the geophysical features on GPR profile associated to the strike-slip fault were scarcely summarized.

The Yushu fault is a sinistral-lateral strike-slip fault with high activity in the late Quaternary, which is responsible for the occurrence of large and moderate magnitude earthquakes in the Yushu area, especially the 7.1 magnitude earthquake happened in 2010. The geomorphologic and trench investigations are mostly used to obtain the important information about Yushu fault (Zhang et al. 2014; Huang et al. 2015; Wu et al. 2017). In this study, multi-frequencies GPR studies were firstly implemented to give us important information about shallow subsurface geometry along the Yushu fault, especially in the ~4 m to ~12 m deep. Although the main fault zones are all obvious and identified at four sites, the pattern of the radar reflections are different in terms of the surface geological and geomorphological records. At site 1, the radar signatures have the more apparent features of weak amplitudes reflections and the strong horizontal reflections are bent when crossing the main fault zone (Fig. 5b), which related to a prominent pressure ridge exists in the loose sediments. On the other hand, diffraction hyperbolas reflections are observed and considered to be the surface ruptures at the uppermost part of the three frequencies data (red arrows in Fig. 5b–d). Because the lithological contrast on either side of the fault zone is sharp and distinct at site 2, the different electrical properties will result in the drastic variations in the waveform and relative amplitude on the GPR data. Thus, two different types of radar reflections are easily revealed on GPR profiles at site 2 (high-amplitude reflections and low-amplitude reflections) in the alluvial fan environment (Fig. 6b–d). A negative flower structure can be explained by the pattern of radar reflections at site 3 and site 4. The flanks of the negative flower structure coincide with the width of the surface subsidence, which interpreted as the main deformation zone. The chaotic reflections are considered to be the disorganized filling materials by the movement of the main

Fig. 8 **a** View of the site 4 in the east of the Changu Temple. White, yellow and red lines are location of 25 MHz, 100 MHz and 250 MHz GPR profiles, red arrows are the fault scarps. **b** The processed 25 MHz GPR profile, high-lighted areas (brown rectangular box) indicates the location of deformation zone at the distance of 60–140 m and 0–25 m in deep. **c** The processed 100 MHz GPR profile at the horizontal distance 100–140 m in 25 MHz profile. There is a significantly distortion and deformation zone of high-amplitude radar reflections whose thickness increases southward. Red arrows show surface marks and red dashed line represents the fault plane. **d** The processed 250 MHz GPR profile. The same interpreted results to the 100 MHz data



fault. To sum up, the main strike-slip fault zone and associated structures are easily identified on the GPR profile by the pattern of the radar reflections (mainly amplitudes reflection) at the surface subsidence environment in Yushu area. The interfaces between the high-amplitude and low-amplitude reflections are generally the flanks of the main fault zone, but there is an ambiguous radar reflection in the deformation zone, which filled with dielectrically different materials. In other geological environment, the geophysical features on GPR profile associated to the main strike-slip fault zone are decided by the drastic variations in the waveform and relative amplitude. What's more, the radar reflections will be bent upward or downward when crossing the main fault zone and it gives rise to the movement of the strike-slip fault.

It is not surprising that 100 MHz and 250 MHz antenna provide more detailed analysis for imaging shallow subsurface features than the 25 MHz antenna because high-frequency antenna offers better lateral and vertical resolution. Nevertheless, the 25 MHz RTA show more remarkable variation in the pattern and amplitude of radar reflections compared with the 100 MHz and 250 MHz antenna in Yushu area. The general view of the deformation zone is clearly displayed in the 25 MHz profiles at four different sites, especially in site 1, site 3 and site 4. Moreover, the 25 MHz data well yield the shallow subsurface geometry and important behavior of the fault at greater depth, including the location and flanks of the deformation zones, the dipping angle and direction of the fault. This seems to be attribute to the high frequency antenna may provide too much details to discern the extensive clutter in the shallow subsurface, such as coarse gravel or scatter stones. We also focus that the propagate depth of GPR signals with 25 MHz, 100 MHz and 250 MHz antenna ranges from about 4 m to 15 m and has a more energy losing during the radar wave propagation. It probably indicates that the shallow subsurface has the high water-saturated clay in Yushu area. Thus, it is more appropriate that a combination of different antenna can be employed in the Yushu fault.

Although the trench wall and GPR profile have the different scales, the features of the fault seen in the GPR images are well correlated with trench walls. The GPR profiles show that the Yushu fault is nearly vertical fault with the angle of 70°–85°. The dip angle of the fault at site1, site 2 and sites 3 is almost steeper than the dip angle at site 4 from top to bottom in the shallow subsurface. These angle variations of the fault plane illustrate the northwest portion of Yushu fault mainly is the left-lateral strike slip with the normal fault components. On the contrast, the tendency of the Yushu fault has the thrust components in the southeast segment indicated by the dip angle of the fault plane and the negative flower structure on GPR profile at site 3 and site 4. What's more, the GPR data also show that the stress perturbations of the 7.1 magnitude earthquake occurred in 2010 are mainly

released in northwest portion of Yushu fault, which is consistent with the geomorphologic rupture indications. It also has been inferred that the fault in northwest segments has higher-intensity seismic activity than southeast segment.

Conclusion

Detailed multi-frequencies GPR measurements were firstly performed for exploring and revealing shallow subsurface geometry about the geologic structure at the four sites along the Yushu fault. It is a challenge for GPR to image shallow subsurface geometry in Yushu area with water-saturated clay in near-surface layer that cause more energy losing during the radar wave propagation. Despite the low probing depth of the GPR (up ~4 m to ~12 m), all GPR data distinctly show shallow subsurface geometry about the geologic structure and important physical characteristics of the fault. The GPR data suggest that the Yushu fault is in general almost vertical fault with the dipping angle ranges from 70° to 85°. What's more, the geophysical features on GPR profile associated to the strike-slip fault are further summarized in different geological and geomorphological environment in Yushu area.

The GPR images are in well agreement with the trench wall sections in the three study sites and it further improve our understanding of the geologic structure and active tectonic setting in the Yushu fault. This is also demonstrated that the ability of GPR method for providing the detailed information of geologic structure and faulting geometry in the Yushu fault, especially in regions where surface expression and geomorphic markers are unobservable. Moreover, GPR measurement will play an increasingly important role in fault study in the Qinghai-Tibet Plateau area.

Acknowledgements This work was partially funded by the work of Chinese Geological Survey (Nos. 12120114002101, DD20160268), the Science and Technique Foundation of Henan Province (No. 182102310001, 192102310001), the Key scientific research Foundation of the university in Henan Province (No. 19A420007), Doctoral Fund of the Henan University of Engineering (No. 2016004), the National Natural Science Foundation of China (No. U1704124).

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Anchuela ÓP, Lafuente P, Arlegui L, Liesa CL, Simón JL (2016) Geophysical characterization of buried active faults: the Conclud Fault (Iberian Chain, NE Spain). *Int J Earth Sci* 105:2221–2239

- Anderson KB, Spotila JA, Hole JA (2003) Application of geomorphic analysis and ground-penetrating radar to characterization of paleoseismic sites in dynamic alluvial environments: an example from southern California. *Tectonophysics* 368:25–32
- Bubeck A, Wilkinson M, Roberts GP, Cowie PA, Mccaffrey KJW, Phillips R, Sammonds P (2015) The tectonic geomorphology of bedrock scarps on active normal faults in the Italian Apennines mapped using combined ground penetrating radar and terrestrial laser scanning. *Geomorphology* 237:38–51
- Busby JP, Merritt JW (1999) Quaternary deformation mapping with ground penetrating radar. *J Appl Geophys* 41:75–91
- Cahit CY, Erhan A, Maksim B, Mustapha M, Volkan K, Serdar Akyüz H (2013) Application of GPR to normal faults in the Büyük Menderes Graben, western Turkey. *J Geodyn* 65:218–227
- Chen LC, Wang H, Ran YK (2010a) The Ms7.1 Yushu earthquake surface ruptures and historical earthquakes. *Chin Sci Bull* 55:1200–1205
- Chen LC et al (2010b) The Ms 7.1 Yushu earthquake surface rupture and large historical earthquakes on the Garze-Yushu Fault. *Chin Sci Bull* 55:3504–3509
- Christie M, Tsollias GP, Stockli DF, Black R (2009) Assessing fault displacement and off-fault deformation in an extensional tectonic setting using 3-D ground-penetrating radar imaging. *J Appl Geophys* 68:9–16
- Demanet D, Renardy F, Vanneste K, Jongmans D, Camelbeeck T, Meghraoui M (2001) The use of geophysical prospecting for imaging active faults in the Roer Graben, Belgium. *Geophysics* 66:78–89
- Dujardin J-R, Bano M (2013) Topographic migration of GPR data: Examples from Chad and Mongolia. *CR Geosci* 345:73–80. <https://doi.org/10.1016/j.crte.2013.01.003>
- Dujardin J-R, Bano M, Schlupp A, Ferry M, Munkhuu U, Tsendayush N, Enkhee B (2014) GPR measurements to assess the Emeelt active fault's characteristics in a highly smooth topographic context, Mongolia. *Geophys J Int* 198:174–186
- Ercoli M, Pauselli C, Frigeri A, Forte E, Federico C (2013) “Geophysical paleoseismology” through high resolution GPR data: A case of shallow faulting imaging in Central Italy. *J Appl Geophys* 90:27–40
- Ercoli M, Pauselli C, Frigeri A, Forte E, Federico C (2014) 3-D GPR data analysis for high-resolution imaging of shallow subsurface faults: the Mt Vettore case study (Central Apennines, Italy). *Geophys J Int* 198:90–99
- Forte E, Ercoli M, Cinti FR, Pauselli C, Volpe R (2015) Imaging of an active fault: Comparison between 3D GPR data and outcrops at the Castrovillari fault, Calabria, Italy. *Interpretation* 3:SY57–SY66
- Gan WJ et al (2007) Present-day crustal motion within the Tibetan Plateau inferred from GPS measurements. *J Geophys Res Solid Earth* 112:1–14. <https://doi.org/10.1029/2005JB004120>
- Grasmueck M (1996) 3-D ground-penetrating radar applied to fracture imaging in gneiss. *Geophysics* 61:1050. <https://doi.org/10.1190/1.1444026>
- Gross R, Green A, Holliger K, Horstmeyer H, Baldwin J (2002) Shallow geometry and displacements on the San Andreas Fault near Point Arena based on trenching and 3-D georadar surveying. *Geophys Res Lett* 29:34–31–34–34. <https://doi.org/10.1029/2002GL015534>
- Grützner C, Reicherter K, Hübscher C, Silva PG (2012) Active faulting and neotectonics in the Baelo Claudia area, Campo de Gibraltar (southern Spain). *Tectonophysics* 554–557:127–142
- Huang XM, Tian K, Du Y, He ZT, Lei JH, Ma BQ, Xie FR (2015) Late quaternary slip rate of the Yushu Batang fault and its tectonic significance. *Acta Scientiarum Naturalium Universitatis Pekinensis* 51:65–78
- Jol HM (2009) Ground penetrating radar theory and applications. Elsevier, Amsterdam
- Lehmann F, Green AG (2000) Topographic migration of georadar data. *Proc SPIE Int Soc Opt Eng* 4084:163–167
- Li YH et al (2013) The geomorphologic and geological marks of the active left-lateral strike-slip fault and the characteristics of geometry and kinematics along the Yushu fault zone in southeastern Tibet. *Geol Bull China* 32:1410–1422
- Li JC, Wu ZH, Zhang D, Liu XD (2014) Remote sensing image interpretation and tectonic activity study of the main active faults in Yushu area, Qinghai Province. *Geol Bull China* 33:535–550
- Lorenzo LS, et al (2016) Active faulting Vs other surface displacing complex geomorphic phenomena. Case studies from a tectonically active area, Abruzzi Region, central Apennines, Italy. In: EGU general assembly conference, Vienna, Austria
- Lunina O (2016) Style of deformation in the Mondy active fault zone based on ground-penetrating radar and structural observations (southern East Siberia). In: International Inqua meeting on paleoseismology, active tectonics and archeoseismology, Pescina, Fucino Basin, Italy
- Lunina OV, Gladkov AS, Afonkin AM, Serebryakov EV (2016) Deformation style in the damage zone of the Mondy fault: GPR evidence (Tunka basin, southern East Siberia). *Russ Geol Geophys* 57:1269–1282
- Lunina OV, Gladkov AS, Gladkov AA (2018) Surface and shallow subsurface structure of the Middle Kedrovaya paleoseismic rupture zone in the Baikal Mountains from geomorphological and ground-penetrating radar investigations. *Geomorphology*. <https://doi.org/10.1016/j.geomorph.2018.03.009>
- Ma YS et al (2010) The surface ruptures and the macroscopical epicenter of Yushu MS 7.1 earthquake. *J Geomech* 16:115–128
- Malik JN, Sahoo AK, Shah AA (2007) Ground-penetrating radar investigation along Pinjore Garden Fault: implication toward identification of shallow subsurface deformation along active fault, NW Himalaya. *Curr Sci* 93:1422–1427
- Malik JN, Sahoo AK, Shah AA (2010) Paleoseismic evidence from trench investigation along Hajipur fault, Himalayan Frontal Thrust, NW Himalaya: Implications of the faulting pattern on landscape evolution and seismic hazard. *J Struct Geol* 32:350–361
- Maurya DM, Chouksey V, Joshi PN, Chamyal LS (2013) Application of GPR for delineating the neotectonic setting and shallow subsurface nature of the seismically active Gedi fault, Kachchh, western India. *J Geophys Eng* 10:034006. <https://doi.org/10.1088/1742-2132/10/3/034006>
- Maurya DM, Chowksey V, Tiwari P, Chamyal LS (2017) Tectonic geomorphology and neotectonic setting of the seismically active South Wagad Fault (SWF), Western India, using field and GPR data. *Acta Geophys* 65:1167–1184
- Mcbride JH, Nelson ST, Heiner BD, Tingey DG, Morris TH, Rey KA (2015) Neotectonics of the Sevier Desert basin, Utah as seen through the lens of multi-scale geophysical investigations. *Tectonophysics* 654:131–155
- Mccalpin J (1996) Paleoseismology. Academic Press, Cambridge
- Mcclymont AF, Villamor P, Green AG (2009) Assessing the contribution of off-fault deformation to slip-rate estimates within the Taupo Rift, New Zealand, using 3-D ground-penetrating radar surveying and trenching. *Terra Nova* 21:446–451. <https://doi.org/10.1111/j.1365-3121.2009.00901.x>
- Pauselli C, Federico C, Frigeri A, Orosei R, Barchi MR, Basile G (2010) Ground penetrating radar investigations to study active faults in the Norcia Basin (central Italy). *J Appl Geophys* 72:39–45
- Pei SP, Chen YS, Feng B, Gao X, Su JR (2013) High-resolution seismic velocity structure and azimuthal anisotropy around the 2010 Ms = 7.1 Yushu earthquake, Qinghai, China from

- 2D tomography. *Tectonophysics* 584:144–151. <https://doi.org/10.1016/j.tecto.2012.08.020>
- Rashed M, Kawamura D, Nemoto H, Miyata T, Nakagawa K (2003) Ground penetrating radar investigations across the Uemachi fault, Osaka, Japan. *J Appl Geophys* 53:63–75. [https://doi.org/10.1016/S0926-9851\(03\)00028-4](https://doi.org/10.1016/S0926-9851(03)00028-4)
- Reicherter K, Michetti AM, Barroso PGS (2012) Palaeoseismology: historical and prehistorical records of earthquake ground effects for seismic hazard assessment. *Environ Eng Geosci* 316:1–10. <https://doi.org/10.2113/gseegeosci.18.3.309>
- Reiss S, Reicherter KR, Reuther CD (2003) Visualization and characterization of active normal faults and associated sediments by high-resolution GPR. *Geol Soc Lond Spec Publ* 211:247–255
- Salvi S, Cinti FR, Colini L, D'Addezio G, Doumaz F, Pettinelli E (2010) Investigation of the active Celano–L'Aquila fault system, Abruzzi (central Apennines, Italy) with combined ground-penetrating radar and palaeoseismic trenching. *Geophys J Roy Astron Soc* 155:805–818
- Schneiderwind S, Mason J, Wiatr T, Papanikolaou I, Reicherter K (2016) 3-D visualisation of palaeoseismic trench stratigraphy and trench logging using terrestrial remote sensing and GPR—combining techniques towards an objective multiparametric interpretation. *Solid Earth* 7:323–340. <https://doi.org/10.5194/se-7-323-2016>
- Shi F, He H, Densmore AL, Li A, Yang X, Xu X (2016) Active tectonics of the Ganzi-Yushu fault in the southeastern Tibetan Plateau. *Tectonophysics* 676:112–124. <https://doi.org/10.1016/j.tecto.2016.03.036>
- Sun H, He H, Wei Z, Shi F, Gao W (2017) Late Quaternary paleoearthquakes along the northern segment of the Nantinghe fault on the southeastern margin of the Tibetan Plateau. *J Asian Earth Sci* 138:258–271. <https://doi.org/10.1016/j.jseaes.2017.02.023>
- Tanajewski D, Bakula M (2016) Application of ground penetrating radar surveys and GPS surveys for monitoring the condition of Levees and Dykes. *Acta Geophys* 64:1093–1111
- Tapponnier P, Xu ZQ, Roger F, Meyer B, Arnaud N, Wittlinger G, Yang JS (2001) Oblique stepwise rise and growth of the Tibet Plateau. *Science* 294:1671–1677. <https://doi.org/10.1126/science.105978>
- Tobita M, Nishimura T, Kobayashi T, Hao KX, Shindo Y (2011) Estimation of coseismic deformation and a fault model of the 2010 Yushu earthquake using PALSAR interferometry data. *Earth Planet Sci Lett* 307:430–438. <https://doi.org/10.1016/j.epsl.2011.05.017>
- Wang WL, Wu JP, Fang LH, Wang CZ (2013) Relocation of the Yushu M_s7.1 earthquake and its aftershocks in 2010 from HypoDD. *Sci Sin (Terrae)* 56:182–191
- Wen XZ, Xu XW, Zheng RZ, Xie YQ, Wan C (2003) Average slip-rate and recent large earthquake ruptures along the Garzê-Yushu fault. *Sci China Earth Sci* 46:276–288
- Wu ZH, Zhang YQ, Hu DG (2014a) Neotectonics, active tectonics and earthquake geology. *Geol Bull China* 33:391–402
- Wu ZH et al (2014b) Active faults and earthquake around Yushu in eastern Tibetan Plateau. *Geol Bull China* 33:419–469
- Wu JW, Huang XM, Xie FR (2017) Late quaternary slip rate of the Garze-Yushu fault zone (Dangjiang Segment). *Chin J Geophys* 60:3872–3888
- Xu C, Xu X, Shyu JB (2015) Database and spatial distribution of landslides triggered by the Lushan, China Mw 6.6 earthquake of 20 April 2013. *Geomorphology* 248:77–92. <https://doi.org/10.1016/j.geomorph.2015.07.002>
- Yan B, Lin A (2017) Holocene activity and paleoseismicity of the Selaha Fault, southeastern segment of the strike-slip Xianshuihe Fault Zone, Tibetan Plateau. *Tectonophysics* 694:302–318. <https://doi.org/10.1016/j.tecto.2016.11.014>
- Zhang KQ et al (2014) A preliminary study of the climate and environment of Yushu area in Qinghai Province based on the carbon-bearing strata in trenches and profiles. *Geol Bull China* 33:485–496
- Zhang GH, Shan XJ, Feng GC (2016) The 3-D surface deformation, coseismic fault slip and after-slip of the 2010 M_w 6.9 Yushu earthquake, Tibet, China. *J Asian Earth Sci* 124:260–268. <https://doi.org/10.1016/j.jseaes.2016.05.011>
- Zhou RJ, Ma SH, Chang X (1997a) Active feature of the Ganzi-Yushu fault zone in the late quaternary. *Earthq Res China* 11:51–63
- Zhou RJ, Wen XZ, Chai CX, Ma SH (1997b) Recent earthquakes and assessment of seismic tendency on the Ganzi-Yushu Fault zone. *Seismol Geol* 19:115–124
- Zhou CJ, Wu ZH, Nima CR, Li JC, Jiang Y, Liu YH (2014) Structural analysis of the co-seismic surface ruptures associated with the Yushu Ms7.1 earthquake, Qinghai Province. *Geol Bull China* 33:551–566



Relationship between mineral magnetic properties and soil textural parameters

M. O. Kanu¹ · O. C. Meludu² · N. Basavaiah³ · A. S. Oniku²

Received: 2 January 2018 / Accepted: 11 January 2019 / Published online: 13 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

The study analyzed the relationship between mineral magnetic properties and particle size in order to determine the suitability of magnetic measurements to serve as particle size proxy for urban top soils samples collected from Jalingo, NE Nigeria. Pearson correlation analyses between mineral magnetic parameters (χ_{lf} , $\chi_{fd\%}$, χ_{ARM} , SIRM, Soft IRM, ARM/SIRM) and particle size properties (PM_{2.5}, PM₁₀, clay, silt and sand) are reported. Results indicate that each particle size class shows different strength of correlation with magnetic parameters, implying that each particle class contains a proportion of magnetic minerals. For the whole data set ($n = 154$), $\chi_{fd\%}$, χ_{ARM} and ARM/SIRM show significant positive correlation with the fine size fractions (PM_{2.5}, PM₁₀, clay and silt) but negatively correlated with the sand proportion, while χ_{lf} , SIRM and Soft IRM exhibit contrasting relationship. Of all the magnetic parameters, the magnetic parameters indicating fine-grained ferrimagnetic minerals (ARM/SIRM and χ_{ARM}) have the strongest and most significant correlations. The nature of the magneto-particle size relationship is different when examined in terms of land use. The fact that magnetic properties and textural parameters have strong relationship indicates that magnetic methods could be considered a potential particle size proxy in Jalingo. However, since the relationship does not follow predictable and consistent patterns of other studies in top soils and sedimentary settings, it brings to fore the non-universality of the method. Hence, the type of relationship existing between magnetic and particle size properties in an environmental setting must be determined before applying the magnetic method as particle size proxy.

Keywords Mineral magnetic · Particle size · Particulate matter · Environmental · Soil

Introduction

The effect of particulate matter (PM) is diverse and has been of great concern to scientist and environmental managers. Particulate matter can threaten human health (Dockery et al. 1993), affects soils, crops (Luo et al. 2011) and visibility (Chen and Xie 2012). Out of these effects, more emphasis is on the health-associated effects since according to WHO (2006) it represents a scientific and social issue. Strong

associations have been found to exist between PM and respiratory symptoms and cardiovascular mortality (Burnett et al. 1997; Lippmann et al. 2003; Lippmann 2011). Particulate matter is primarily derived from natural sources derived from soil (wind-blown soils) and particles derived from human activities such as materials used for constructing roads (e.g., paints, cements, asphalt); emissions from vehicles' exhaust pipes, tire and brake linings; inputs from industrial wastes (e.g., coal and oil combustions); biomass burning (natural fires) and long range atmospheric depositions (Adachi and Tainosho 2005; Meza-Figuero et al. 2007; Jordanova et al. 2012).

Particulate matter is ordinarily known as dust. The major classification of particulate matter are PM₁₀ (particles having diameter less than 10 micron, i.e., $< 10 \mu\text{m}$), PM_{2.5} (describing concentration of particles having diameter less than 2.5 micron, $< 2.5 \mu\text{m}$) and PM₁ (describing particles smaller than 1 μm). Usually, due to the concentrated traffic density in urban areas, PM_{2.5} and PM₁₀ concentrations

✉ M. O. Kanu
maxwell.kanu@tsuniversity.edu.ng

¹ Department of Physics, Taraba State University, P.M.B. 1167, Jalingo, Taraba State, Nigeria

² Department of Physics, Modibbo Adama University of Technology, P.M.B. 2076, Yola, Adamawa State, Nigeria

³ Environmental Magnetism Laboratory, Indian Institute of Geomagnetism, Kalamboli Highway, Navi-Mumbai, Mumbai 410218, India

increase in urban centers than rural areas (Chen et al. 2007). As a result of the contribution of most pollution by anthropogenic sources, urban centers most often experience spatial and daily fluctuations in the concentration of air pollution (Petrovský et al. 2000; Maher et al. 2008; Brimblecombe 2011; Crosby et al. 2014).

Natural soils are made up of soil particles that vary in size. Soil texture gives information about the relative amount of the different particle sizes such as clay, silt and sand in the soil. It influences the mechanism of flow of water in the soil and the ease at which soil can be worked. Besides its use in agriculture, particle size is also used in measuring the pedogenic and detrital proportions (Zan et al. 2015) and sediment provenance studies (Thompson and Morton 1979; Booth et al. 2005). It has been demonstrated that particle size can greatly affect analytical data related to soils (Booth et al. 2005; Crosby et al. 2014). Pollutants like concentrations of trace metal, radionuclide content and quantity of polychlorinated biphenyls are affected by particle sizes: the concentration increases with decreasing sediment size. Fine-grained sediments usually have larger specific surface areas, cation exchange capacities and surface charges. These increase considerably the extent to which they are adsorbed preferentially (Booth et al. 2005).

Environmental magnetic methods have gained popularity recently as an established proxy method for determining soil heavy metal pollution (Hoffmann et al. 1999; Matzka and Maher 1999; Petrovský et al. 2000; Maher et al. 2008; Basavaiah et al. 2012; Crosby 2012; Jordanova et al. 2012; Kanu et al. 2017), climatic studies (Maxbauer et al. 2016), archaeological studies (Hill et al. 2007), sediment source tracing (Beckwith et al. 1986) and radionuclides concentration (Clifton et al. 1999). These studies utilized the relationships between mineral magnetic parameters and physical and chemical properties of soil, sediments, dusts, etc. The application of environmental magnetic parameters as proxy for particle size is not as established as its use in pollution studies. However, some studies have achieved considerable success in establishing mineral magnetism as particle size proxy (Thompson and Morton 1979; Zheng et al. 1991; Chen et al. 1995; Clifton et al. 1999; Booth et al. 2005, 2007; Crosby et al. 2009, 2014; Oldfield et al. 2009). Most of these studies focused on sedimentary environments, except Booth et al. (2008) and Crosby et al. (2009, 2013, 2014) who focused on roadside soils and Manx top soils from British Isle, respectively. They found that the relationship between magnetic and particle size properties may not be applicable universally; hence, caution must be exercised in its usage. Furthermore, the studies are concentrated in Europe, America and Asia, none so far coming from Africa with tropical climate. Therefore, more case studies are required to understand better and in detail the magnetic proxy method, especially on how it works and under which conditions the technique can

be applied in different environments, climates and ecosystems. It must be noted that it is a necessary task to identify the process that drives the magnetic response as magnetic grain size, concentration and mineralogy can respond to changes in geology, size of particle, transport, delivery and flux, anthropogenic pollution and diagenetic and authigenic changes (Hatfield 2014). Hence, magnetic proxies cannot be generalized without prior information on their origin in the samples.

In this study, we intend to achieve the following: (1) examine the level of application of magnetic parameters as reliable proxy for particle size, (2) determine if the relationship between mineral magnetic and textural parameters can be applied universally, that is follow the pattern of other studies and (3) examine if human activities/land use affects the way in which particle size and magnetic properties relate.

Materials and methods

Study area and sample collection strategy

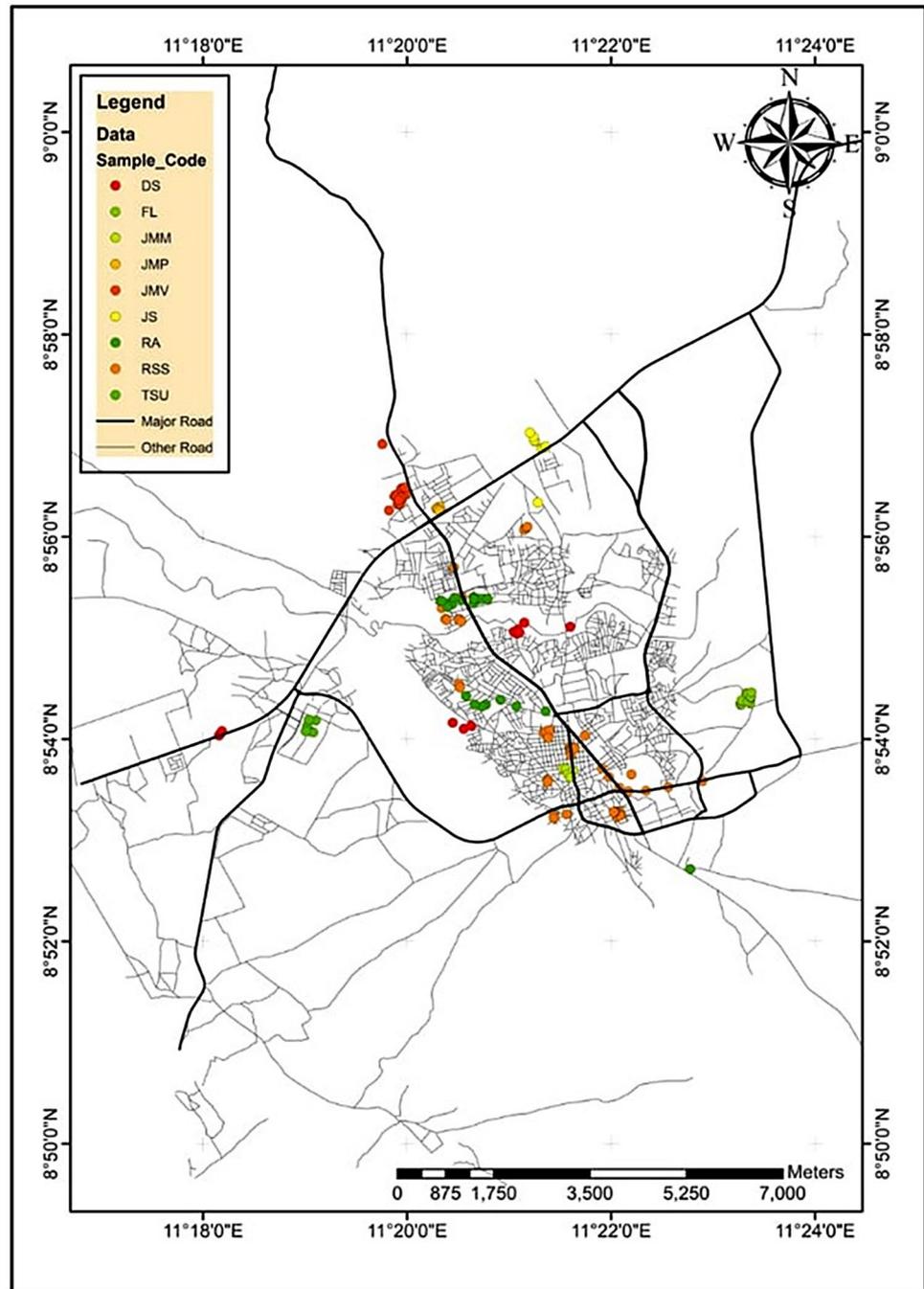
This study was conducted in Jalingo metropolis, the state headquarters of Taraba State, Nigeria. Jalingo is located between latitude 8°50' and 8°90'N and between longitude 11°17' and 11°26'E (Fig. 1). Details about the study area can be found in Kanu et al. (2017). The study area is part of the Nigerian Basement Complex rocks and is mainly composed of migmatites, gneisses and the older granites (Obaje 2009).

To investigate the possible links between mineral magnetic and particle size within Jalingo metropolis, a strategy that combines land use pattern and traffic conditions was used. To this end, a number of sampling sites were selected for this study. A total of 154 top soil (0–10 cm) samples were randomly collected from residential areas ($n=25$), roadsides ($n=50$), dumpsites ($n=15$), mechanic workshops and motor park ($n=25$), farm land ($n=20$) and other locations: market/commercial area, sport complex, school environment ($n=19$). The samples were collected between November, 2011 and March, 2012.

Magnetic measurements

The magnetic measurements were performed following the procedure adopted by our previous work (Kanu et al. 2017). Magnetic susceptibility measurements were done using the Agico MFK1-FA Kappabridge instrument. Measurements were carried out at low frequency of 976 Hz and high frequency of 15,616 Hz at an alternating field of 200 A m⁻¹. From these, the frequency-dependent susceptibility ($\chi_{fd}\%$) parameter that is used to detect the presence of fine-grained (<0.03 μm) superparamagnetic (SP) ferrimagnetic minerals in samples (Basavaiah et al. 2012) was calculated. Here,

Fig. 1 Map of Jalingo showing sample points



$\chi_{fd}\%$ was calculated; thus, $\chi_{fd}(\%) = \frac{\chi_{lf} - \chi_{hf}}{\chi_{lf}} \times 100\%$, where χ_{lf} and χ_{hf} are low- and high-frequency magnetic susceptibility, respectively.

Anhyseretic remanent magnetization (ARM) was induced with peak alternating magnetic field of 100 mT and a constant DC bias field of 0.05 mT using a D-2000 AF demagnetizer (ASC scientific). The remanence acquired is measured using a Molspin spinner magnetometer, and the

values were converted to give the mass specific susceptibility of ARM (χ_{ARM}).

In order to determine the isothermal remanent magnetization (IRM), the samples were exposed to a forward field of 0.02 T and 1 T in a Molspin pulse magnetizer and the IRM was measured using the spinner magnetometer. The highest field that can be measured by the Molspin pulse magnetizer is 1 T and is used as saturation isothermal remanent magnetization (SIRM). Thereafter, back field demagnetization in steps of 20, 30, 100 and 300 mT was applied. The

soft isothermal remanent magnetization (Soft IRM) and S -ratio was then calculated from the following relations, respectively,

$$\text{Soft IRM} = \left(\frac{\text{SIRM} - \text{IRM}_{-30\text{ mT}}}{2} \right), \quad S_{\text{ratio}} = \frac{|\text{IRM}_{-300\text{ mT}}|}{\text{SIRM}}$$

Particle size/textural analyses

Particle size analyses were performed using a CILAS 1064 particle size analyzer with a detection range from 0.03 to 500 μm . Prior to measurement, the soil and dust samples were sieved using a sieve merge and the $< 125 \mu\text{m}$ size fractions were used for the particle size analysis. Approximately 200 mg of samples was dispersed in sodium hexametaphosphate $[\text{Na}(\text{PO}_3)_6]$ solution and is poured to the sample chamber and laser analyzed for the grain size spectrum for a 10 min program. For data reproducibility, three runs per sample were conducted. The results obtained from the laser granulometry were separated into different grades of fractions following Wentworth (1922) classification scheme; thus, clay $\leq 4 \mu\text{m}$, silt = 4–63 μm and sand $\geq 63 \mu\text{m}$. Other finer particles classes are $\text{PM}_{2.5} \leq 2.5 \mu\text{m}$ and $\text{PM}_{10} \leq 10 \mu\text{m}$.

Results and discussion

Mineral magnetic properties of the different sampling sites in Jalingo are shown in Table 1. The magnetic susceptibility values varied considerably from a low value of $0.21 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ to a high value of $912.56 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$. The highest mean χ_{lf} values are obtained in roadside soil ($219.13 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$) followed by residential area ($166.32 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$) and dumpsites ($131.43 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$). SIRM varied between $40.81 \times 10^{-5} \text{ A m}^2 \text{ kg}^{-1}$ and $4010.76 \times 10^{-5} \text{ A m}^2 \text{ kg}^{-1}$, with highest mean values obtained in dumpsites followed by roadside soils and residential areas. The values of χ_{lf} and SIRM in all the locations are high indicating abundance of ferrimagnetic and remanence carrying magnetic minerals. In all the sampling sites, high mean values of χ_{ARM} are obtained, implying high concentration of stable single-domain magnetite in the soils. Generally, the mean values of χ_{ARM} varied in the order farmland ($0.30 \times 10^{-5} \text{ m}^3 \text{ kg}^{-1}$) > dumpsites ($0.27 \times 10^{-5} \text{ m}^3 \text{ kg}^{-1}$) > residential area ($0.20 \times 10^{-5} \text{ m}^3 \text{ kg}^{-1}$) > roadside ($0.17 \times 10^{-5} \text{ m}^3 \text{ kg}^{-1}$) > other locations ($0.15 \times 10^{-5} \text{ m}^3 \text{ kg}^{-1}$) > mechanic workshop/motor park ($0.12 \times 10^{-5} \text{ m}^3 \text{ kg}^{-1}$). Usually, χ_{lf} and SIRM are regarded as approximate estimates of

Table 1 Summary statistics of magnetic parameters ($n = 154$)

Sampling locations	$\chi_{\text{lf}} (\times 10^{-8} \text{ m}^3 \text{ kg}^{-1})$	$\chi_{\text{fd}} (\%)$	$\chi_{\text{ARM}} (\times 10^{-5} \text{ m}^3 \text{ kg}^{-1})$	SIRM ($\times 10^{-5} \text{ A m}^2 \text{ kg}^{-1}$)	ARM/SIRM (dimensionless)	S -ratio (dimensionless)	Soft IRM ($\times 10^{-5} \text{ A m}^2 \text{ kg}^{-1}$)
Road side soil ($n = 50$)							
Range	21.17–614.89	1.50–7.68	0.01–0.32	63.45–2009.72	0.003–0.02	0.88–1.00	28.23–924.84
Mean \pm SD	219.13 \pm 138.80	2.98 \pm 1.42	0.17 \pm 0.07	1026.83 \pm 528.50	0.008 \pm 0.004	0.96 \pm 0.03	466.93 \pm 238.63
Residential area ($n = 25$)							
Range	16.08–912.56	0.30–14.72	0.05–0.45	113.53–4010.76	0.004–0.03	0.80–0.97	40.78–1598.98
Mean \pm SD	166.32 \pm 251.63	6.70 \pm 3.67	0.20 \pm 0.12	979.28 \pm 1187.61	0.01 \pm 0.01	0.91 \pm 0.05	390.22 \pm 447.29
Dump sites ($n = 15$)							
Range	53.82–352.34	4.01–9.76	0.12–0.57	358.81–3192.96	0.01–0.02	0.93–0.99	127.87–1529.71
Mean \pm SD	131.43 \pm 93.89	6.74 \pm 1.84	0.27 \pm 0.13	1138.52 \pm 859.10	0.01 \pm 0.01	0.97 \pm 0.02	521.71 \pm 411.74
Mechanic workshop and motor park ($n = 25$)							
Range	0.21–375.90	2.24–12.91	0.001–0.53	40.81–3967.96	0.00–0.03	0.84–0.99	10.54–1833.40
Mean \pm SD	79.67 \pm 97.28	5.60 \pm 3.02	0.12 \pm 0.12	663.98 \pm 866.71	0.01 \pm 0.01	0.95 \pm 0.05	297.43 \pm 399.91
Farm land ($n = 20$)							
Range	35.04–160.75	4.88–19.38	0.14–0.95	251.55–872.94	0.01–0.71	0.68–0.99	94.53–360.23
Mean \pm SD	75.26 \pm 37.41	12.52 \pm 3.18	0.30 \pm 0.23	412.80 \pm 138.95	0.03 \pm 0.02	0.90 \pm 0.07	166.02 \pm 67.90
Other locations ($n = 19$)							
Range	18.80–351.03	1.61–15.14	0.08–0.32	108.38–1110.82	0.004–0.03	0.69–1.00	42.29–485.56
Mean \pm SD	88.93 \pm 82.03	9.56 \pm 5.19	0.15 \pm 0.06	435.31 \pm 266.71	0.02 \pm 0.01	0.91 \pm 0.07	191.60 \pm 124.50
Total ($n = 154$)							
Range	0.21–912.56	0.30–19.38	0.001–0.945	40.81–4010.76	0.00–0.07	0.68–1.00	10.54–1833.40
Mean \pm SD	143.99 \pm 150.63	6.45 \pm 4.37	0.19 \pm 0.13	815.93 \pm 758.25	0.01 \pm 0.01	0.94 \pm 0.06	354.34 \pm 328.32

ferrimagnetic minerals contents in the samples (Shen et al. 2008). Specifically, SIRM is generally used as approximation of the total magnetic mineral concentration with grain size larger than SP/single-domain (SD) threshold. Unlike χ_{lf} , SIRM is not influenced by para- and diamagnetic minerals (Thompson and Oldfield 1986; Zhang et al. 2007). χ_{ARM} can give information about the amount of ferrimagnetic minerals in stable single-domain (SSD) region (Kanu et al. 2017). χ_{lf} , SIRM and χ_{ARM} are regarded as concentration-dependent magnetic parameters, while $\chi_{fd}\%$ and ARM/SIRM are parameters that depend on grain size. *S*-ratio and Soft IRM are indicative of magnetic mineralogy. From the $\chi_{fd}\%$ results, the mean values indicate that the roadside soil is overshadowed by coarse multidomain (MD) magnetic particles, while samples from farmland are dominated by SP grains. Other locations are composed of admixture of SSD and SP grains. High *S*-ratio (average, 0.94) and Soft IRM suggest the dominance of low coercivity soft magnetic minerals like magnetite and/or maghemite in the samples.

The results of particle size analysis by laser granulometry are presented in Table 2. The concentration of individual particle size classes is compared using box plots shown in Fig. 2. Results show that the concentration of sand fractions dominated the assemblage (51.76%) followed by the silt fraction (41.14%) and clay fraction (7.10%). However, based on individual land use, the silt fraction dominated the soils from dumpsites and farm land, while sand fractions

dominated the roadside soil, residential areas and mechanic workshop and motor park. In other locations, the ratio of silt to sand is very close (46.08%: 47.45%). High sand fractions in road side soils have also been observed in Southport, Merseyside, UK (Booth et al. 2007), Scunthorpe, North Lincolnshire, UK (Crosby et al. 2009) and Wolverhampton, UK (Crosby et al. 2013).

In terms of the respiratory health size classes, PM10 grains represent ~9%, while PM 2.5–4% of the sediments. The PM 2.5 and PM 10 are seen to be highly variable and have the highest values in the farmland. This may be attributed to accumulation of combustion products from bush burning and long distance travel vehicular emissions and even fertilizer application. The least values of the respiratory health-related sizes PM 2.5 and PM 10 are found in roadside soils (Table 2, Fig. 2), while the highest values are obtained in the farmland.

Correlation between magnetic properties and particle size classes

The values of the Pearson correlation coefficient (*r*) between magnetic concentration, mineralogical and grain and particle size parameters for Jalingo soils (*n* = 152) were computed using SPSS 20 statistical software and are presented in Table 3. Each magnetic parameter correlates significantly (at either *P* < 0.01 or *P* < 0.05) either positively or negatively with one particle size class or the other. It

Table 2 Summary statistics of textural parameters (*n* = 154)

Sampling locations	PM 2.5 (%)	PM10 (%)	Clay (%)	Silt (%)	Sand (%)
Road side soil (<i>n</i> = 50)					
Range	1.38–4.1	2.22–7.53	2.21–7.53	14.18–61.45	31.02–83.61
Mean ± SD	2.36 ± 0.68	4.73 ± 1.90	4.10 ± 1.31	27.95 ± 10.64	67.95 ± 11.81
Residential area (<i>n</i> = 25)					
Range	2.01–6.35	4.17–16.65	3.54–12.26	20.71–72.08	17.73–74.90
Mean ± SD	3.34 ± 1.31	7.72 ± 3.89	6.06 ± 2.67	39.99 ± 12.69	53.95 ± 14.50
Dump sites (<i>n</i> = 15)					
Range	2.72–5.75	6.38–16.92	4.83–11.08	48.22–77.37	11.55–46.95
Mean ± SD	3.96 ± 0.90	10.32 ± 3.05	7.24 ± 1.86	61.89 ± 7.91	30.87 ± 9.57
Mechanic workshop and motor park (<i>n</i> = 25)					
Range	1.45–10.39	1.63–21.55	2.11–18.85	14.96–79.18	9.61–82.32
Mean ± SD	4.02 ± 3.22	8.21 ± 7.18	7.09 ± 6.04	39.96 ± 19.81	52.95 ± 25.01
Jalingo farm land (<i>n</i> = 20)					
Range	2.19–23.00	3.20–42.04	3.41–41.04	35.23–71.90	0.00–57.63
Mean ± SD	8.80 ± 6.93	18.60 ± 13.23	16.19 ± 12.78	56.09 ± 11.66	27.71 ± 21.99
Other locations (<i>n</i> = 19)					
Range	1.47–5.90	2.11–17.31	2.19–11.68	19.03–69.03	20.65–78.78
Mean ± SD	3.56 ± 1.46	8.54 ± 4.01	6.48 ± 2.84	46.08 ± 15.69	47.45 ± 18.37
Total (<i>n</i> = 154)					
Range	1.38–23.00	1.63–42.02	2.11–41.04	14.18–79.18	0.00–83.61
Mean ± SD	3.94 ± 3.53	8.62 ± 7.43	7.10 ± 6.58	41.14 ± 17.54	51.76 ± 22.22

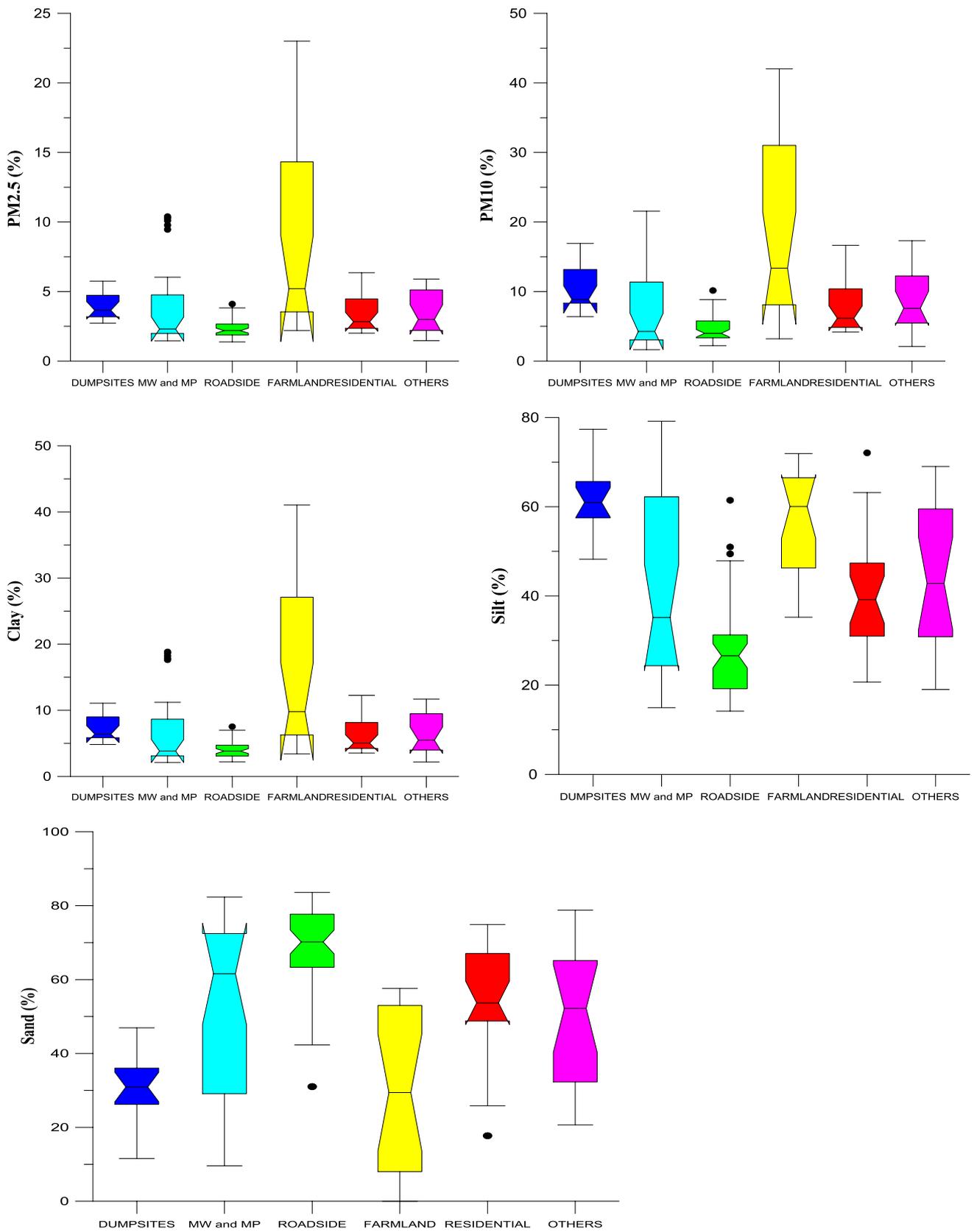


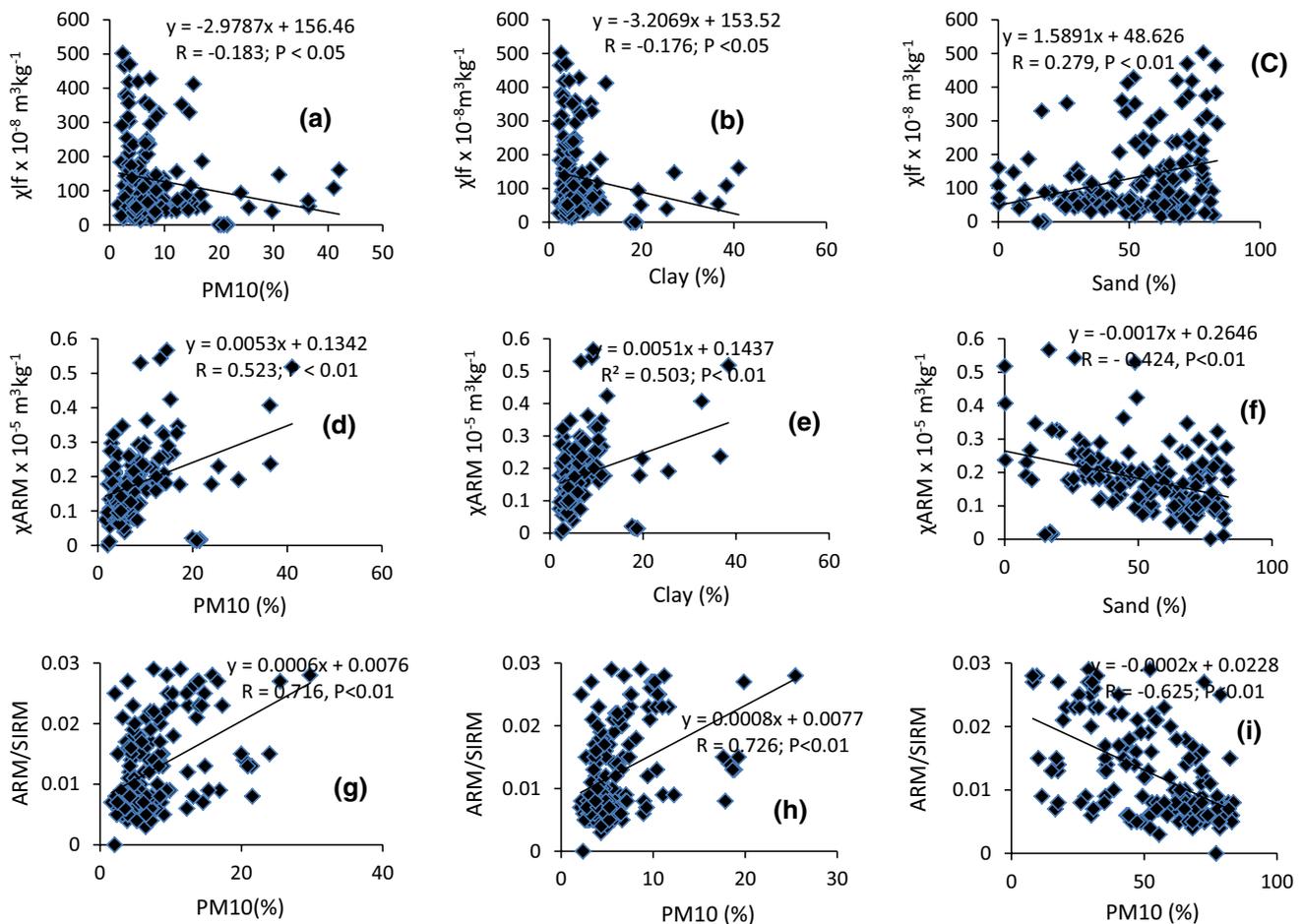
Fig. 2 Notched boxplots showing the variation of the particle size classes in the sampling locations

Table 3 Mineral magnetic and textural properties correlation for Jalingo metropolis ($n = 152$)

	PM 2.5 (%)	PM (10%)	Clay (%)	Silt (%)	Sand (%)
χ_{lf}	-0.179	-0.183*	-0.176*	-0.288**	0.279**
$\chi_{fd\%}$	0.549**	0.572**	0.551**	0.570**	-0.613**
χ_{ARM}	0.498**	0.523**	0.503**	0.349**	-0.424**
SIRM	-0.190*	-0.167*	-0.185*	-0.177*	0.195*
ARM/SIRM	0.727**	0.716**	0.726**	0.519**	-0.625**
S-ratio	-0.345**	-0.360**	-0.349**	-0.325**	0.360**
Soft IRM	-0.189*	-0.167*	-0.184*	-0.179*	0.196*

Significant correlation values are bolded

**Implies significant correlation at the 0.01 level (2-tailed), while * implies significant correlation at the 0.05 level (2-tailed)

**Fig. 3** Scatter plots of some magnetic parameters (χ_{lf} , χ_{ARM} and ARM/SIRM) and particle size classes (PM 10, clay and sand)

is also observed that each particle size class shows different strengths of correlation with the magnetic parameters. This implies that all particle size fractions contain a proportion of magnetic minerals, but the proportion varies with particle size (Oldfield et al. 2009). For the magnetic concentration parameters (χ_{lf} , SIRM and χ_{ARM}), weak negative correlations are observed between χ_{lf} and SIRM

and PM2.5, PM10, clay and silt fractions. The sand proportion shows positive and stronger association with χ_{lf} (0.279 , $P < 0.01$) and SIRM (0.195 , $P < 0.05$), suggesting that differences in the ferrimagnetic components could be used as proxy for the concentration of sand. In contrast to χ_{lf} and SIRM, χ_{ARM} exhibit stronger significant positive correlation with PM2.5 ($r = 0.498$, $P < 0.01$), PM10

Table 4 Mineral magnetic and textural properties correlation for dumpsites ($n=15$)

	PM 2.5 (%)	PM (10%)	Clay (%)	Silt (%)	Sand (%)
χ_{lf}	0.591*	0.705**	0.609*	0.699**	-0.696**
$\chi_{fd\%}$	0.045	0.046	0.056	-0.016	0.003
χ_{ARM}	0.600*	0.726**	0.618*	0.773**	-0.759**
SIRM	0.528	0.659*	0.548*	0.662**	-0.653*
ARM/SIRM	-0.154	-0.253	-0.175	-0.201	0.200
S-ratio	0.235	0.017	0.156	0.041	-0.063
Soft IRM	0.537*	0.670**	0.562*	0.646*	-0.642*

Significant correlation values are bolded

**Implies significant correlation at the 0.01 level (2-tailed), while * implies significant correlation at the 0.05 level (2-tailed)

Table 5 Mineral magnetic and textural properties correlation for road side soils ($n=50$)

	PM 2.5 (%)	PM (10%)	Clay (%)	Silt (%)	Sand (%)
χ_{lf}	-0.067	-0.067	-0.071	-0.113	0.110
$\chi_{fd\%}$	0.288	0.313*	0.294*	0.418**	-0.410**
χ_{ARM}	-0.014	0.028	0.000	-0.069	0.062
SIRM	-0.178	-0.152	-0.173	-0.224	0.221
ARM/SIRM	0.293*	0.334*	0.309*	0.423**	-0.416**
S-ratio	-0.358*	-0.345*	-0.357*	-0.398**	0.400**
Soft IRM	-0.201	-0.179	-0.197	-0.243	0.241

Significant correlation values are bolded

**Implies significant correlation at the 0.01 level (2-tailed), while * implies significant correlation at the 0.05 level (2-tailed)

($r=0.523$, $P<0.01$), clay ($r=0.503$, $P<0.01$) and silt ($r=0.349$, $P<0.01$) fractions and negative association with sand ($r=-0.424$, $P<0.01$). This result indicates that the maximum contribution of SD ferrimagnetic grains comes

Table 6 Mineral magnetic and textural parameters correlation for Jalingo mechanic village (JMV) and Jalingo central motor park (JMP) ($n=25$)

	PM 2.5 (%)	PM10 (%)	Clay (%)	Silt (%)	Sand (%)
χ_{lf}	-0.584**	-0.569**	-0.586**	-0.358	0.425*
$\chi_{fd\%}$	0.112	0.173	0.120	0.311	-0.275
χ_{ARM}	-0.376	-0.275	-0.367	0.134	-0.017
SIRM	-0.564**	-0.539**	-0.568**	-0.241	0.329
ARM/SIRM	0.307	0.373	0.318	0.451*	-0.434*
S-ratio	-0.909**	-0.890**	-0.905**	-0.743**	0.807**
Soft IRM	-0.607**	-0.584**	-0.611**	-0.293	0.380

Significant correlation values are bolded

**Implies significant correlation at the 0.01 level (2-tailed), while * implies significant correlation at the 0.05 level (2-tailed)

Table 7 Mineral magnetic and textural properties correlation for farm land ($n=20$)

	PM 2.5	PM 10	Clay	Silt	Sand
χ_{lf}	0.457*	0.444*	0.451*	0.131	-0.332
$\chi_{fd\%}$	0.731**	0.702**	0.725**	0.459*	-0.665**
χ_{ARM}	0.686**	0.670**	0.680**	0.367	-0.590**
SIRM	0.127	0.125	0.125	-0.039	-0.051
ARM/SIRM	0.800**	0.769**	0.794**	0.444	-0.696**
S-ratio	-0.047	-0.023	-0.042	0.122	-0.040
Soft IRM	0.346	0.331	0.343	0.054	-0.228

Significant correlation values are bolded

**Implies significant correlation at the 0.01 level (2-tailed), while * implies significant correlation at the 0.05 level (2-tailed)

Table 8 Mineral magnetic and textural properties correlation for residential areas ($n=24$)

	PM 2.5 (%)	PM (10%)	Clay (%)	Silt (%)	Sand (%)
χ_{lf}	0.301	0.263	0.319	-0.214	0.127
$\chi_{fd\%}$	0.065	0.116	0.054	0.465*	-0.417
χ_{ARM}	0.643**	0.652**	0.661**	0.245	-0.339
SIRM	0.137	0.117	0.154	-0.267	0.204
ARM/SIRM	0.235	0.316	0.228	0.689**	-0.646**
S-ratio	-0.253	-0.245	-0.273	0.195	-0.119
Soft IRM	0.204	0.172	0.219	-0.252	0.180

Significant correlation values are bolded

**Implies significant correlation at the 0.01 level (2-tailed), while * implies significant correlation at the 0.05 level (2-tailed)

from the finer size classes. In Wolverhampton (UK) soils, Crosby et al. (2013) also obtained negative correlation between χ_{lf} , SIRM and χ_{ARM} for lower size fractions and positive correlation with the medium and coarse sand fractions. The magnetic grain size indicators ($\chi_{fd\%}$ and ARM/

Table 9 Mineral magnetic and textural properties correlation for other locations ($n = 19$)

	PM 2.5	PM 10	Clay	Silt	Sand
χ_{lf}	-0.272	-0.159	-0.237	-0.111	0.132
$\chi_{fd\%}$	0.444	0.347	0.413	0.337	-0.352
χ_{ARM}	0.666**	0.654**	0.661**	0.696**	-0.697**
SIRM	-0.245	-0.146	-0.214	-0.101	0.119
ARM/SIRM	0.470*	0.412	0.450	0.375	-0.390
S-ratio	-0.241	-0.176	-0.221	-0.313	0.302
Soft IRM	-0.233	-0.134	-0.203	-0.067	0.089

Significant correlation values are bolded

**Implies significant correlation at the 0.01 level (2-tailed), while * implies significant correlation at the 0.05 level (2-tailed)

SIRM) show stronger relationship with the particle sizes than the concentration-dependent parameters. $\chi_{fd\%}$ and ARM/SIRM reveal strong positive association with PM2.5, PM10, clay and silt fractions and negative correlation with sand. The S-ratio and Soft IRM which indicate abundance

of soft coercive magnetic minerals (e.g., magnetite, maghemite) are negatively correlated with PM2.5, PM10, clay and silt fractions and positively correlated with sand fraction. This may indicate that the concentration of higher coercivity minerals increases with finer size fractions, but magnetite/maghemite increases in the coarse sand fraction.

In general, it is worth noting that the magnetic particle size relationship is divided into two broad categories: those that correlated positively with the fine size classes (PM2.5, PM10, clay and silt) but negatively with the proportion of sand (i.e., $\chi_{fd\%}$, χ_{ARM} and ARM/SIRM) and others that correlated negatively with the fine fractions but positively correlated with sand (χ_{lf} , SIRM, S-ratio and Soft IRM). This further indicates that $\chi_{fd\%}$, χ_{ARM} and ARM/SIRM which respond to fine-grained minerals are strongly positively correlated with fine size classes and negatively correlated with the coarse sand fraction. Also, for this study area, ARM/SIRM proved to be the most suitable proxy for estimating particle size fractions in Jalingo top soils, since it showed the highest significant correlation values with all the size classes. This is closely followed by χ_{ARM} and $\chi_{fd\%}$.

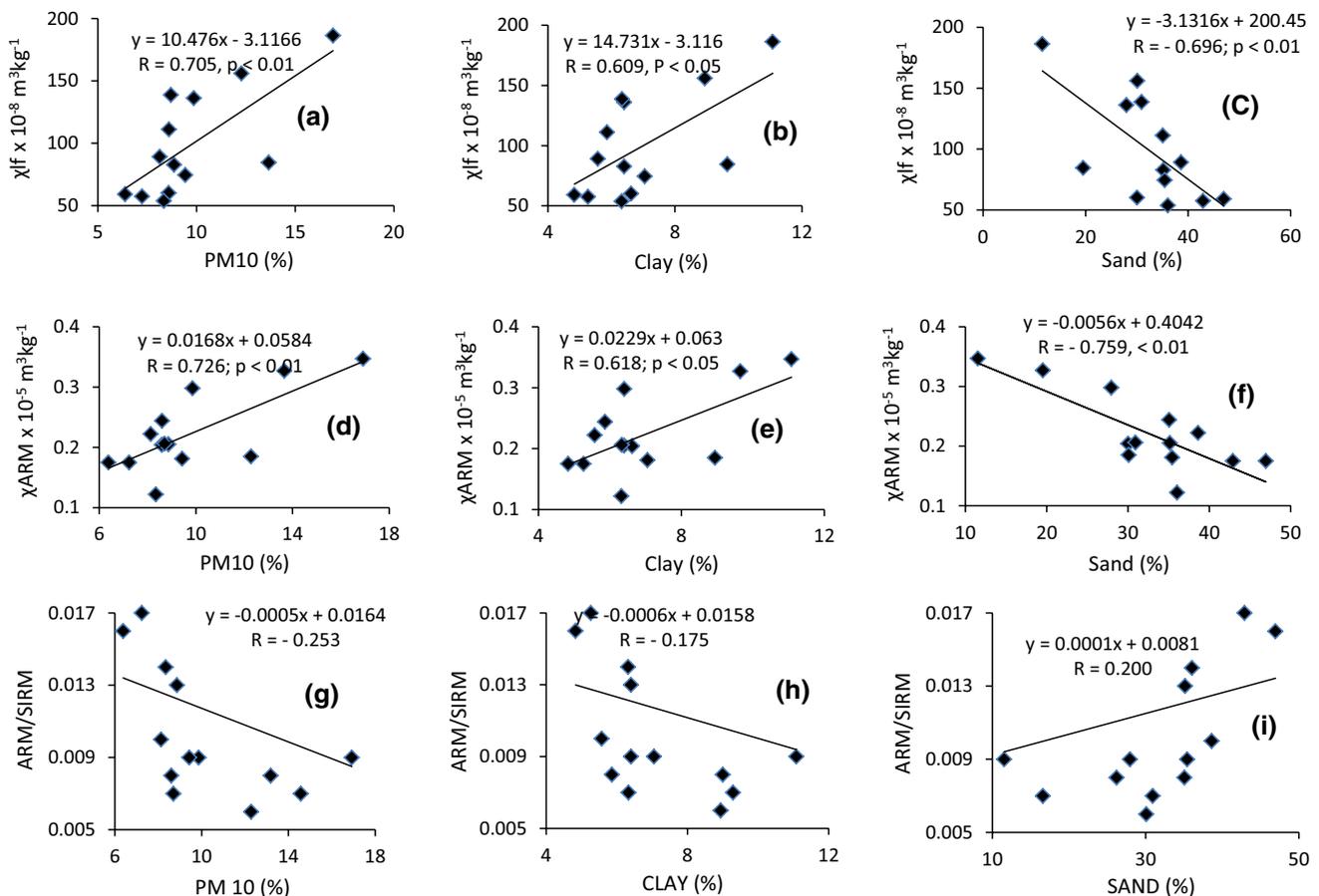


Fig. 4 Scatter plots of some magnetic parameters (χ_{lf} , χ_{ARM} and ARM/SIRM) and particle size classes (PM 10, clay and sand) for dumpsites sample

In comparison with previous studies, significant relationship between χ_{lf} , SIRM and χ_{ARM} and particle size classes has been found. Measurement of ARM is used to determine the amount of magnetite in fine textured (less than $0.1 \mu\text{m}$) clay size fraction, while χ_{lf} could be a useful indication of the quantity of coarse MD magnetite in large textured ($> 1 \mu\text{m}$) fractions (Oldfield et al. 1993). Clifton et al. (1999) found that χ_{lf} had strong correlation with sands and medium silts fractions, χ_{ARM} had strong association with clay and fine silts, while SIRM had strong association with very fine to medium silts. According to Zhang et al. (2001), $\chi_{fd}\%$ and χ_{ARM} could be used as a proxy for clay content. In terms of the potential of mineral magnetic measurements as a proxy for urban PM, Booth et al. (2005, 2007) found that χ_{lf} , χ_{ARM} and SIRM parameters correlated with all class sizes and suggested that χ_{lf} , χ_{ARM} and SIRM have potential as particle size proxies for several sedimentary environments. They, however, noted the importance of fully determining the nature of the relationship between sediment particle size and magnetic properties before applying mineral magnetic data as a size proxy. More recently, this was further explored by Booth et al. (2008) who discussed potential problems of

employing this technique. Crosby et al. (2009) found that in urban road-deposited sediment, χ_{ARM} and SIRM were strongly related to all the different PM classes (PM1.0, PM2.5 and PM10), but χ_{lf} did not. Saragnese et al. (2011) reported variable degree of correlation between the contents of PM 10 and various magnetic properties and attributed this variation to the diverse sources of origin of PM 10. The trends observed in the present study are not totally consistent with these earlier investigations, though significant similarities are evident. For example, while sand correlated negatively with χ_{lf} , χ_{ARM} and SIRM (Booth et al. 2005, 2007), it is positively correlated with χ_{lf} and SIRM and negatively associated with χ_{ARM} in this study. Also, while Crosby et al. (2013) obtained significant moderate negative correlation between χ_{lf} , χ_{ARM} and SIRM and clay, silt and sand contents, this study found significant ($P < 0.05$) but weak negative correlation between χ_{lf} and SIRM and clay and silt contents, and positive correlation with sand content. χ_{ARM} rather shows stronger significant but contrasting relationship.

Figure 3a–i shows bivariate scatter plots of some magnetic parameters (χ_{lf} , χ_{ARM} and ARM/SIRM) versus selected textural parameters (clay, sand and PM10). The

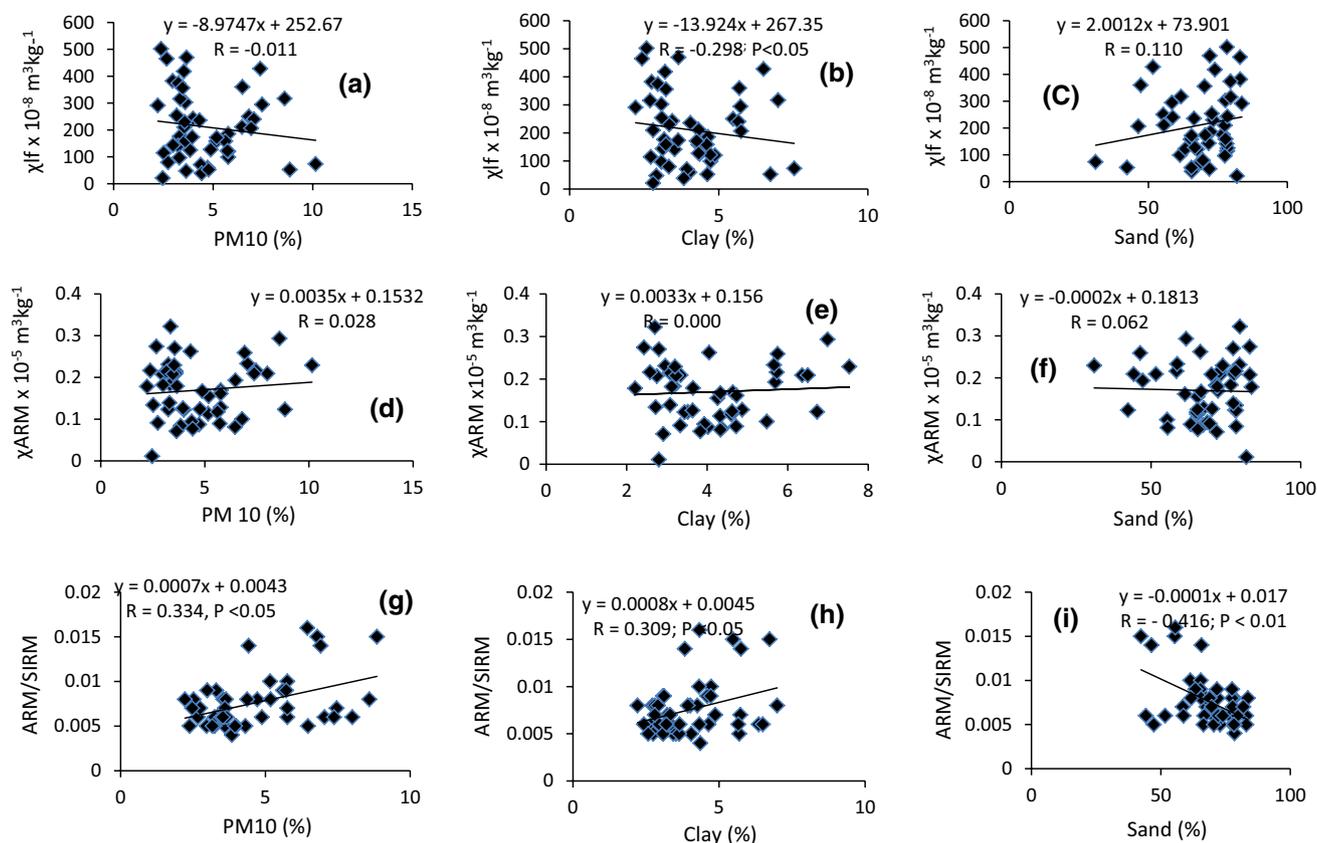


Fig. 5 Scatter plots of some magnetic parameters (χ_{lf} , χ_{ARM} and ARM/SIRM) and particle size classes (PM 10, clay and sand) for road side soils samples

plots of χ_{lf} and clay, sand and PM10 (Fig. 3a–c) show that the sample points are distributed all over the plot area, suggesting that the relationship does not give enough confidence to conclude that magnetic susceptibility could be a particle size proxy in the study area. Figure 3d–i shows relatively good degrees of scatter around the trend line, suggesting that χ_{ARM} and ARM/SIRM might be a good proxy for the contents of the textural parameters.

The study further assessed the dependence of the magnetic parameters and particle size relationship on land use (road side soils, dumpsites, mechanic workshops/motor parks, farmland, etc.). Results are presented in Tables 4, 5, 6, 7, 8 and 9. The results indicate that the strength and significance of the association between magnetic parameters and textural parameters can be different for specific environments within the same geological settings. Similar observation was found by Booth et al. (2008) within a sedimentary setting. The potential application of magnetic methods as proxy for particle size on dumpsites yielded interesting results shown in Table 4. Unlike the general case, χ_{lf} , χ_{ARM} , SIRM and Soft IRM are significantly positively

associated with the proportions of PM2.5, PM10, clay and silt and negatively correlated with sand fractions consistent with Booth et al. (2005, 2007). This suggests that in the dumpsites, χ_{lf} , χ_{ARM} , SIRM and Soft IRM could be effective particle size proxy. The grain size-dependent magnetic parameters ($\chi_{fd}\%$ and ARM/SIRM) show contrasting trend, though with very weak correlation coefficient that is not significant at $P < 0.01$ and $P < 0.05$. Figure 4a–i shows scatter plots of some magnetic parameters (χ_{lf} , χ_{ARM} and ARM/SIRM) versus selected textural parameters (clay, sand and PM10) for the dumpsites samples. The plots show good relationship between the particle sizes and χ_{lf} and χ_{ARM} (Fig. 4a–f), but not with ARM/SIRM (Fig. 4g–i). ARM/SIRM plots show greater scatter with the particle size and might not be a suitable particle size proxy for the dumpsites.

In Table 5, the Pearson correlation coefficients for Jalingo roadside soils are presented. The result indicate that χ_{lf} , SIRM and Soft IRM are poorly and negatively correlated with all the particle sizes except sand fraction. $\chi_{fd}\%$ and ARM/SIRM show significant positive relationship with PM10, clay and silt, and negative correlation with sand

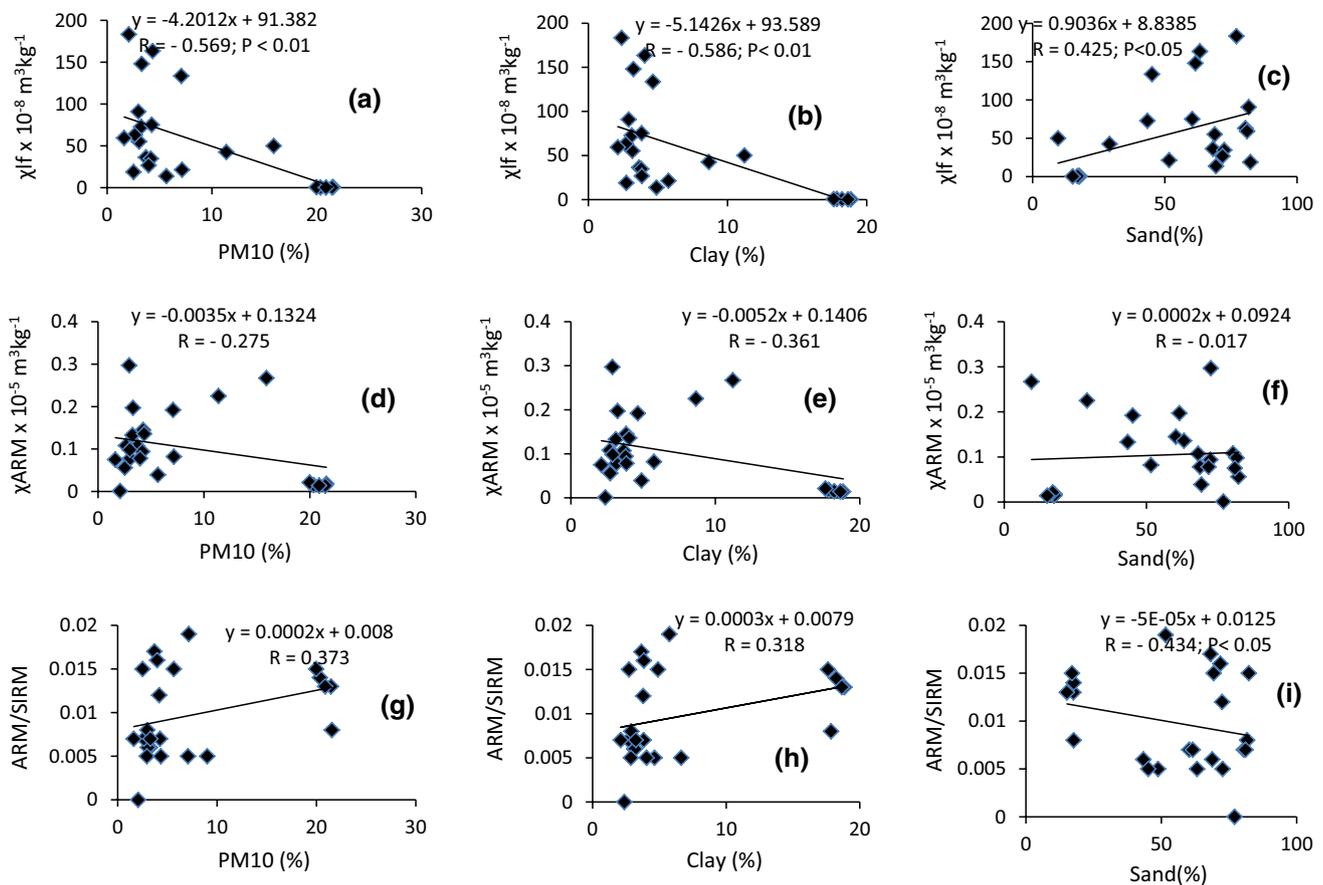


Fig. 6 Scatter plots of some magnetic parameters (χ_{lf} , χ_{ARM} and ARM/SIRM) and particle size classes (clay, sand and PM 10) for JMP and JMV soils samples

proportion. *S*-ratio is significantly negatively correlated with PM_{2.5}, PM₁₀, clay and silt fractions and positively correlated with sand. For this data set (road side soils), χ_{lf} , SIRM, χ_{ARM} and Soft IRM may not be reliable particle size proxies as they display minimal/poor relationship with particle sizes. In comparison with road-deposited sediment from Scunthorpe, North Lincolnshire, UK, Crosby et al. (2009) found that moderate significant relationships exist between clay content, PM₁₀, PM_{2.5} and PM_{1.0} and χ_{ARM} and SIRM with χ_{ARM} displaying the strongest correlation. In this study, the strongest correlation is found between $\chi_{fd}\%$, ARM/SIRM and *S*-ratio and PM_{2.5}, PM₁₀, silt and sand, suggesting that these magnetic parameters ($\chi_{fd}\%$, ARM/SIRM and *S*-ratio) could be considered as potential proxy for PM_{2.5}, PM₁₀ and silt in the road side soil. Figure 5 clearly summarizes these observations. Data points are highly scattered in the χ_{lf} , χ_{ARM} versus particle size plots as opposed to the ARM/SIRM versus particle size plots which indicates existence of better relationships (Fig. 5g–i).

For the Jalingo mechanic village (JMV) and Jalingo central motor park (JMP), *S*-ratio displays the strongest relationship with particle size classes. It is significantly negatively correlated with PM_{2.5} ($r = -0.909$, $P < 0.01$), PM₁₀

($r = -0.890$, $P < 0.01$), clay ($r = -0.905$, $P < 0.01$), silt ($r = -0.743$, $P < 0.01$) and positively correlated with fraction of sand ($r = 0.807$, $P < 0.01$) (Table 6). This indicates the abundance of soft magnetic mineral (e.g., magnetite) in the coarse fraction and the hard component (e.g., hematite) in the finer fractions. PM_{2.5}, PM₁₀ and clay are observed to relate significantly negatively with χ_{lf} , SIRM and Soft IRM, while ARM/SIRM shows significant positive relationship with silt ($r = 0.451$, $P < 0.05$) and negative correlation with sand ($r = -0.434$, $P < 0.05$) only. The poor correlation observed between χ_{ARM} and ARM/SIRM textural parameters in JMV and JMP (Table 6) may suggest the dominance of MD grains in the magnetic fraction over the finer SD and/or SP grains. This is because MD grains in most cases originate from polluted sources (Kanu et al. 2017). The data spread in Fig. 6 indicates that high level of confidence will not be given to χ_{lf} , χ_{ARM} and ARM/SIRM as particle size normalization parameters. Although χ_{lf} relationship with particle sizes is promising, the data are clearly not distributed normally.

The relationship between some mineral magnetic and textural parameters for a farmland in Jalingo is presented in Table 7 and Fig. 7. Here, significant negative correlations

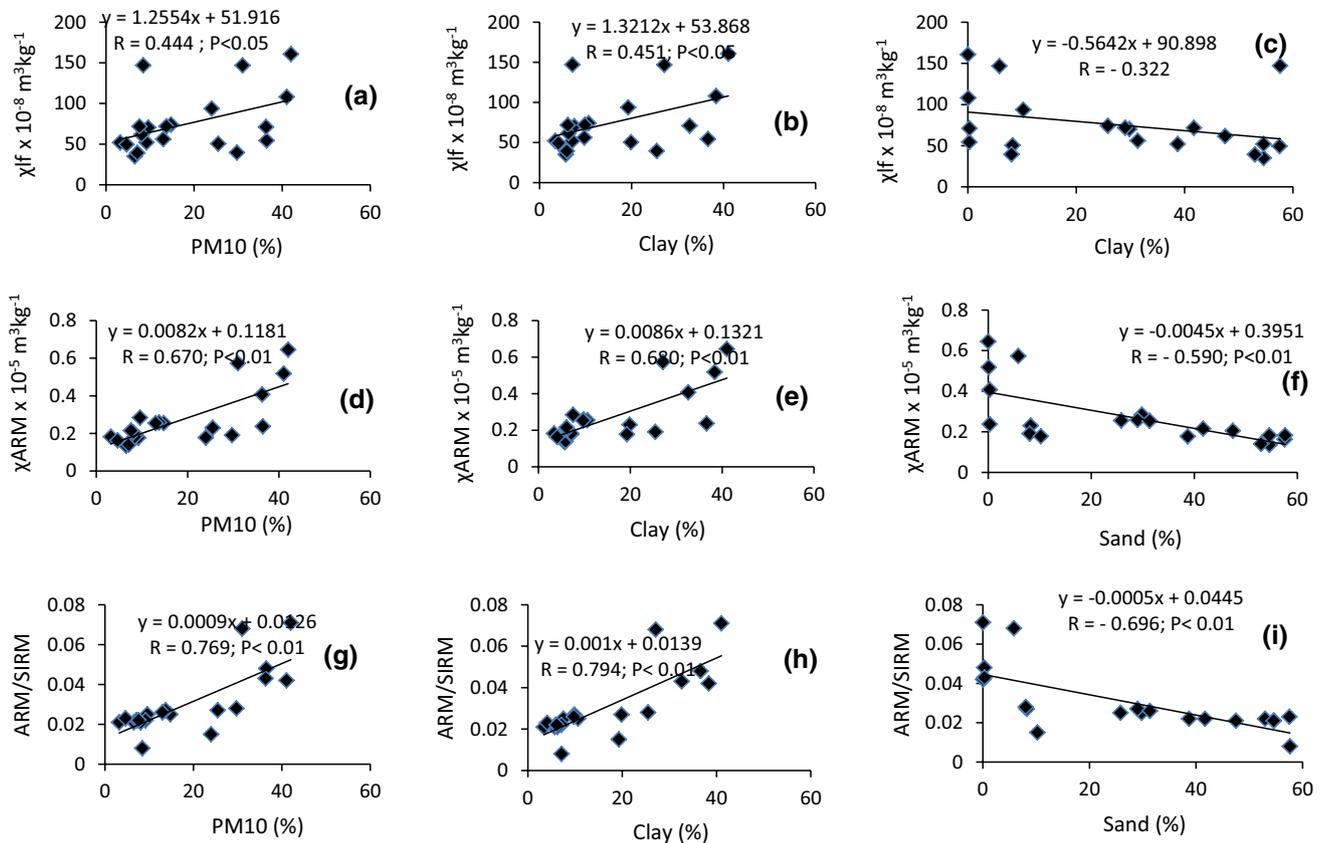


Fig. 7 Scatter plots of some magnetic parameters (χ_{lf} , χ_{ARM} and ARM/SIRM) and particle size classes (clay, sand and PM 10) for Jalingo farmland soils samples

are found between χ_{lf} , $\chi_{fd\%}$, χ_{ARM} and ARM/SIRM and fine fractions (PM2.5, PM10 and clay). These magnetic parameters are, however, found to exhibit significant negative correlation with sand fraction. Silt content is found to be related to $\chi_{fd\%}$ only. The data distribution is concentrated around the regression line (Fig. 7) compared to the situation in the JMP and JMV samples (Fig. 6). In this site, the $\chi_{fd\%}$, χ_{ARM} and ARM/SIRM displayed the strongest relationship and could be used as potential particle size proxy.

The situation in residential area is different from the other sites. In the Jalingo residential areas, χ_{ARM} significantly correlated positively with the fine size fractions (PM2.5, PM10 and clay), while ARM/SIRM displayed significant and positive correlation with silt and negative correlation with sand (Table 8 and Fig. 8). This implied that χ_{ARM} and ARM/SIRM could be considered appropriate particle size proxy in the residential area. Weiguo and Lizhong (2003) also obtained χ_{ARM} to be highly correlated with fine fraction ($< 4 \mu\text{m}$) for tidal flat sediments from Yangtze River, China.

Results of magnetic-particle size relationships for combination of other locations in Jalingo such as school environment, market place, sports complex, etc., are shown

in Table 9 and Fig. 9. In this case, χ_{ARM} appears to be the most efficient particle size proxy. It shows significant positive correlation at 1% confidence level with PM2.5 ($r = 0.666$), PM 10 ($r = 0.654$), clay ($r = 0.661$) and Silt ($r = 0.696$) and negative correlation with sand ($r = -0.697$). Also, ARM/SIRM exhibits moderate positive correlation which is significant at $P < 0.05$ with PM2.5 ($r = 0.470$).

The results indicated that when magnetic concentration parameters are employed as estimate for pollutant content, it is important to look into the changes caused by composition of particle size and possibly reduce such effects. Significant correlations between the magnetic parameters and the respiratory health-related size classes have shown the potential of using the technique as a possible alternative tool for monitoring PM, which could be linked to both health and pollution studies. It can be said that magnetic measurements have high possibility to be a reliable proxy for particle size. However, when data presented here are compared between different sampling sites and previous investigations, it is clear that there is no general trend that could be adduced. For example, one cannot conclude that χ_{ARM} is a clay size proxy or χ_{lf} is used as a sand proxy. Therefore, more case studies are needed to arrive at a conclusion; otherwise,

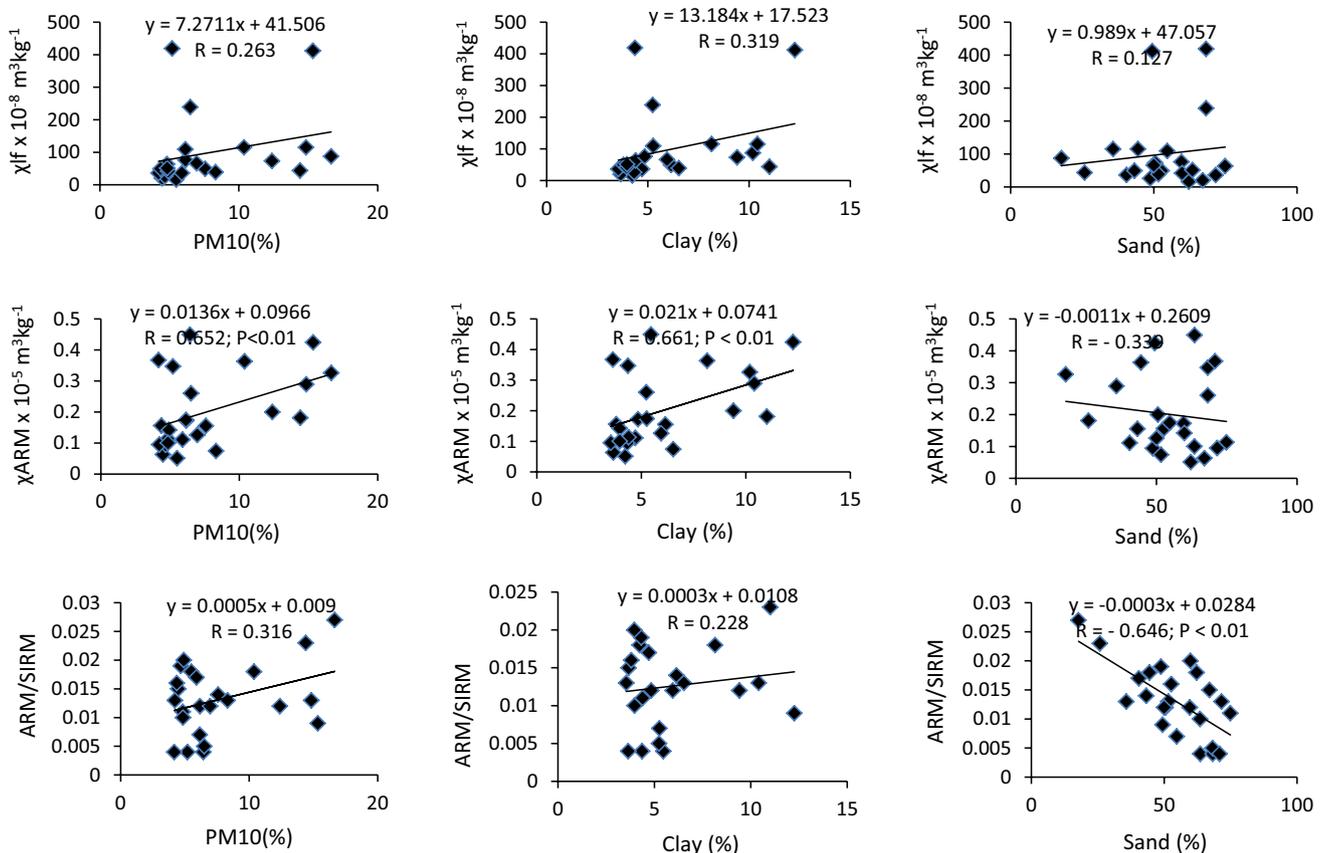


Fig. 8 Scatter plots of some magnetic parameters (χ_{lf} , χ_{ARM} and ARM/SIRM) and particle size classes (clay, sand and PM 10) for soils from Jalingo residential areas samples

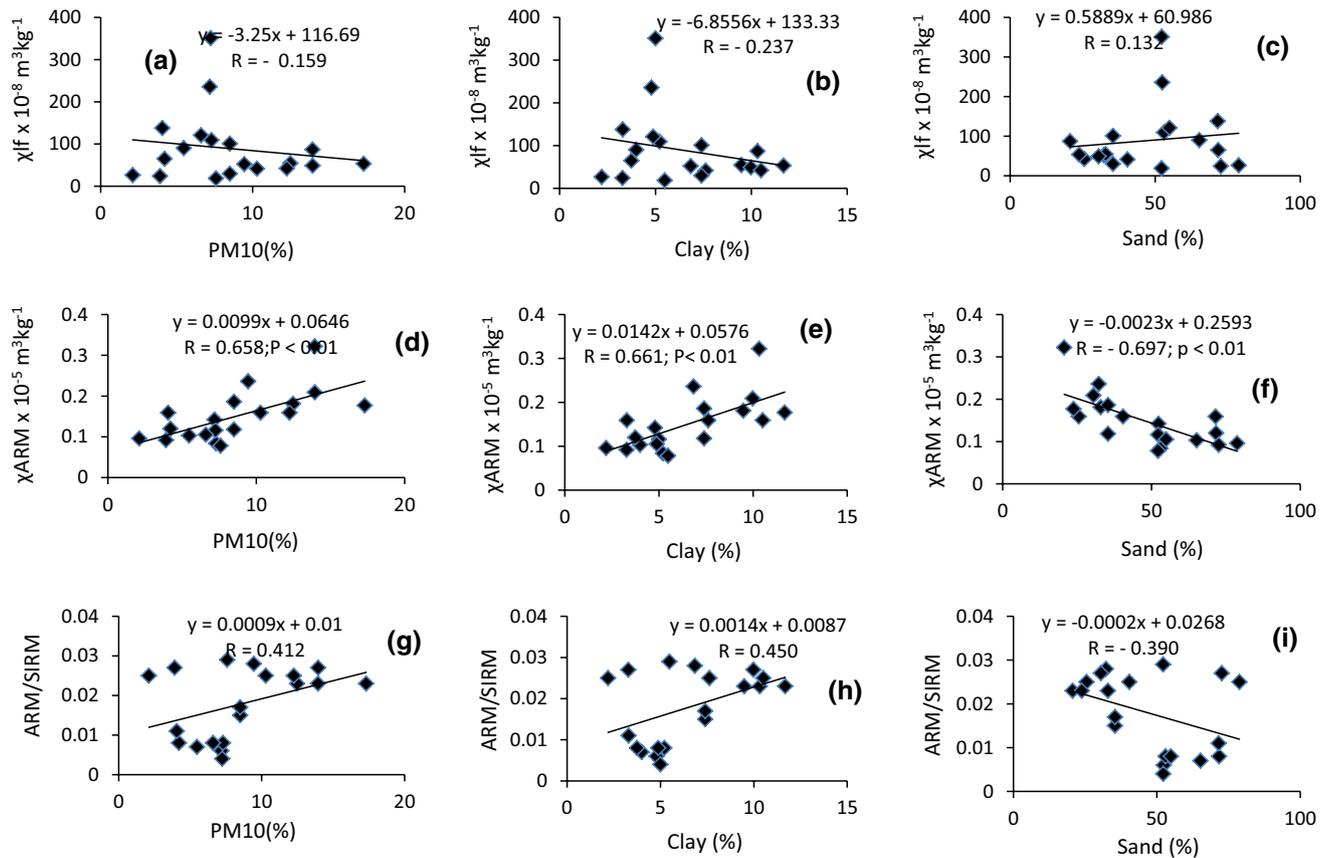


Fig. 9 Scatter plots of some magnetic parameters (χ_{lf} , χ_{ARM} and ARM/SIRM) and particle size classes (clay, sand and PM 10) for soils from other locations

caution should be exercised when using mineral magnetic methods as particle size proxy.

Conclusions

- Results obtained indicate that each particle size class shows different strengths of correlation with magnetic parameters, implying that each particle size class contains a proportion of magnetic minerals.
- Some magnetic properties ($\chi_{fd}\%$, χ_{ARM} and ARM/SIRM) are positively correlated with finer fractions (PM2.5, PM10, clay and silt) but negatively correlated with sand proportion, while χ_{lf} , SIRM and Soft IRM reveal opposite trend of correlation.
- Land use affects the manner in which magnetic parameters relate with particle size. It varies from one land use pattern to another.
- Assessment of all selected magnetic parameters used for the study shows that the magnetic parameters that indicate fine-grained ferrimagnetic minerals (ARM/SIRM and χ_{ARM}) have the strongest and most signifi-

cant correlation values and most consistently significant ($P < 0.01$) relationship with most particle size classes and could be considered most suitable particle size proxy for the study area. Thus, magnetic techniques have the potential to be considered as proxy for particle size.

- Although significant relationship between magnetic and particle size parameters has been obtained, the trend does not follow predictable pattern of other studies. This implies that the application of magnetic data as proxy for particle size may not be universal; hence, for a particular soil or environmental setting, the nature of the association should be examined before attempting to use magnetic properties as particle size proxy.

Acknowledgements This research was partly funded by the Taraba State University TETFund grant, and the authors are deeply grateful. The Department of Science and Technology (DST), India, and the Indian Institute of Geomagnetism (IIG), Mumbai, India, are also appreciated for access and use of the facilities at the Environmental Magnetism Laboratory. The authors wish to thank Drs. K. Deenadayalan and P. K. Das for assistance during laboratory measurements.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Adachi K, Tainosho Y (2005) Single particle characterization of size-fractionated road sediments. *Appl Geochem* 20:849–859
- Basavaiah N, Blaha U, Das PK, Deenadayalan K, Sadashiv MB, Schulz H (2012) Evaluation of environmental magnetic pollution screening in soils of basaltic origin: results from Nashik Thermal Power Station, Maharashtra, India. *Environ Sci Pollut Res* 19:3028–3038
- Beckwith PR, Ellis JB, Revitt DM, Oldfield F (1986) Heavy metal and magnetic relationships for urban source sediments. *Phys Earth Planet Inter* 42(1–2):67–75
- Booth CA, Walden J, Neal A, Smith JP (2005) Use of mineral magnetic concentration data as a particle size proxy: a case study using marine, estuarine and fluvial sediments in the Carmarthen Bay area, South Wales, U.K. *Sci Total Environ* 347:241–253
- Booth CA, Winspear CM, Fullen MA, Worsley AT, Power AL, Holden VJC (2007) A pilot investigation into the potential of mineral magnetic measurements as a proxy for urban roadside particulate pollution. In: Borrego CA, Brebbia CA (eds) *Air pollution XV*. WIT Press, Ashurst, pp 391–400
- Booth CA, Fullen MA, Winspear CM, Walden J, Worsley AT, Marcinkonis S, Coker AO (2008) Problems and potential of mineral magnetic measurements as a soil particle size proxy. *J Environ Eng Landsc Manag* 16(3):151–158
- Brimblecombe P (2011) Air pollution episodes. In: Nriagu JO (ed) *Encyclopedia of environmental health*. Elsevier, Burlington, pp 39–45
- Burnett RT, Cakmak S, Brook JR, Krewski D (1997) The role of particulate size and chemistry in the association between summertime ambient air pollution and hospitalization for cardio respiratory diseases. *Environ Health Perspect* 105:614–620
- Chen Y, Xie S (2012) Temporal and spatial visibility trends in the Sichuan Basin, China, 1973 to 2010. *Atmos Res* 112:25–34
- Chen FH, Wu RJ, Pompei D, Oldfield F (1995) Magnetic property and particle size variations in the late Pleistocene and Holocene parts of the Dadongling loess section near Xining, China. *Quat Proc* 4:27–40
- Chen T, Gokhale J, Shofer S, Kuschner W (2007) Outdoor air pollution: particulate matter health effects. *Am J Med Sci* 333(4):235–243
- Clifton J, McDonald P, Plater A, Oldfield F (1999) Derivation of a grain-size proxy to aid the modelling and prediction of radionuclide activity in saltmarshes and mud flats of the Eastern Irish Sea. *Estuar Coast Shelf Sci* 48:511–518
- Crosby CJ (2012) Application of mineral magnetic measurements as a pollution proxy for urban road deposited sediment. Ph.D. thesis, University of Wolverhampton, 345 pp
- Crosby CJ, Booth CA, Worsley AT, Fullen MA, Searle DE, Khatib JM, Winspear CM (2009) Application of mineral magnetic concentration measurements as a particle size proxy for urban road deposited sediments. *Air Pollut XVII*:153–162
- Crosby CJ, Booth CA, Fullen MA (2013) Mineral magnetic measurements as particle size proxy for urban roadside soil pollution (part 1). *Environ Sci Process Impacts*. <https://doi.org/10.1039/c3em00344b>
- Crosby CJ, Fullen MA, Booth CA, Searle DE (2014) A dynamic approach to urban road deposited sediment pollution monitoring (Marylebone Road, London, UK). *J Appl Geophys* 105:10–20
- Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG, Speizer FE (1993) An association between air pollution and mortality in six U.S. cities. *N Engl J Med* 329(24):1743–1759
- Hatfield RG (2014) Particle size-specific magnetic measurements as a tool for enhancing our understanding of the bulk magnetic properties of sediments. *Minerals* 4:758–787. <https://doi.org/10.3390/min4040758>
- Hill M, Lanos P, Chauvin A, Vitali D, Laubenheimer F (2007) An archaeomagnetic investigation of a Roman amphorae workshop in Albinia (Italy). *Geophys J Int* 169(2):471–482
- Hoffmann V, Knab M, Appel E (1999) Magnetic susceptibility mapping of roadside pollution. *J Geochem Explor* 66(1–2):313–326
- Jordanova D, Jordanova N, Lanos P, Petrov P, Tsacheva T (2012) Magnetism of outdoor and indoor settled dust and its utilization as a tool for revealing the effect of elevated particulate air pollution on cardiovascular mortality. *Geochem Geophys Geosyst* 13(08):Q08Z49. <https://doi.org/10.1029/2012gc004160>
- Kanu MO, Basavaiah N, Meludu OC, Oniku AS (2017) Investigating the potential of using environmental magnetism techniques as pollution proxy in urban road deposited sediment. *Int J Environ Sci Technol*. <https://doi.org/10.1007/s13762-017-1356-5>
- Lippmann M (2011) Particulate matter (PM) air pollution and health: regulatory and policy implications. *Air Qual Atmos Health* 5:1–5
- Lippmann M, Ito K, Nadas A, Burnett RT (2003) Association of particulate matter components with daily mortality and morbidity in urban populations. *Res Rep* 95:5–82
- Luo X, Yu S, Li X (2011) Distribution, availability, and sources of trace metals in different particle size fractions of urban soils in Hong Kong: implications for assessing the risk to human health. *Environ Pollut* 159(5):1317–1326
- Maher BA, Moore C, Matzka J (2008) Spatial variation in vehicle-derived metal pollution identified by magnetic and elemental analysis of roadside tree leaves. *Atmos Environ* 42(2):364–373
- Matzka J, Maher BA (1999) Magnetic biomonitoring of roadside tree leaves: identification of spatial and temporal variations in vehicle-derived particulates. *Atmos Environ* 33:4565–4569
- Maxbauer DP, Feinberg JM, Fox DL (2016) Magnetic mineral assemblages in soils and paleosols as the basis for paleoprecipitation proxies: a review of magnetic methods and challenges. *Earth Sci Rev* 155:28–48
- Meza-Figueroa D, O-Villanueva MD, Parra MLD (2007) Heavy metal distribution in dust from elementary schools in Hermosillo, Sonora, Mexico. *Atmos Environ* 41:276–288
- Obaje NG (2009) *Geology and mineral resources of Nigeria, vol 120*. Lecture notes in earth sciences. Springer, Berlin. <https://doi.org/10.1007/978-3-540-92685-61>
- Oldfield F, Richardson N, Appleby PG, Yu L (1993) ²⁴¹Am and ¹³⁷Cs activity in fine-grained saltmarsh sediments from parts of the N.E. Irish Sea shoreline. *J Environ Radioact* 19:1–24
- Oldfield F, Hao Q, Bloemendal J, Gibbs-Eggar Z, Patil S, Guo Q (2009) Links between bulk sediment particle size and magnetic grain-size: general observations and implications for Chinese loess studies. *Sedimentology* 56:2091–2106
- Petrovský E, Kapička A, Jordanova N, Knab M, Hoffmann V (2000) Low-field magnetic susceptibility: a proxy method of estimating increased pollution of different environmental systems. *Environ Geol* 39(3–4):312–318
- Saragnese F, Lanci L, Lanza R (2011) Nanometric-sized atmospheric particulate studied by magnetic analyses. *Atmos Environ* 45:450–459
- Shen M, Hu S, Blaha U, Yan H, Roster W, Appel E, Hoffmann V (2008) Magnetic properties of urban soil profiles and their significance for traffic pollution-case study of the capital airport expressway in Beijing. *Front Earth Sci China* 2(4):400–407

- Thompson R, Morton DJ (1979) Magnetic susceptibility and particle-size distribution in recent sediments of the Loch Lomond drainage basin, Scotland. *J Sediment Petrol* 49(3):0801–0812
- Thompson R, Oldfield F (1986) *Environmental magnetism*. Allen and Unwin, London
- Weiguo Z, Lizhong Y (2003) Magnetic properties of tidal flat sediments of the Yangtze Estuary and its relationship with particle size. *Sci China (Ser D)* 46(9):954–955
- Wentworth CK (1922) A scale of grade and class terms for clastic sediments. *J Geol* 30(5):372–392
- World Health Organization (2006) *Health risks of particulate matter from long-range transboundary air pollution*. European Centre for Environment and Health, Bonn
- Zan J, Fang X, Yang S, Yan M (2015) Bulk particle size distribution and magnetic properties of particlesized fractions from loess and paleosol samples in Central Asia. *Geochem Geophys Geosyst* 16:101–111. <https://doi.org/10.1002/2014GC005616>
- Zhang W, Yu L, Hutchinson SM (2001) Diagenesis of magnetic minerals in the intertidal sediments of the Yangtze Estuary, China, and its environmental significance. *Sci Total Environ* 266:160–175
- Zhang W, Yu L, Lu M, Hutchinson SM, Feng H (2007) Magnetic approach to normalizing heavy metal concentrations for particle size effects in intertidal sediments in the Yangtze Estuary, China. *Environ Pollut* 147(1):238–244
- Zheng HB, Oldfield F, Yu LH, Shaw J, An ZS (1991) The magnetic properties of particle-sized samples from the Luo-Chuan Loess section—evidence for pedogenesis. *Phys Earth Planet Inter* 68:250–258



Magnetic interpretation utilizing a new inverse algorithm for assessing the parameters of buried inclined dike-like geological structure

Khalid S. Essa¹ · Mahmoud Elhussein¹

Received: 5 September 2018 / Accepted: 28 January 2019 / Published online: 5 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

A new algorithm has been established to interpret magnetic anomaly data due to inclined dike-like structure. This algorithm uses first horizontal derivative anomalies attained from magnetic anomaly data utilizing filters of sequential window lengths. The final estimated parameters are the half-width, the depth, angle of magnetization and amplitude factor of an inclined dike-like geological structure. A minimum variance criterion is used for selecting the most suitable variables. This algorithm has been realized to theoretical data without and with random noise. The effects of interference due to near structures have additionally been studied. The method was then applied to two field examples from Turkey and Peru, which demonstrate its effectiveness and accurateness. Thus, it is a respectable correspondence among the model parameters retrieved from this approach, drilling information, and the outcomes published in the literature. For example, in Turkey, we applied the technique to gauge the source variables and also the results were precise to $w=64.74$ m and $h=87.65$ m (2% and 3% errors, respectively) based on information from Aydin and Gelişli (Jeofizik 10:41–49, 1996).

Keywords Magnetic anomaly · Inclined dike · Depth · Half-width

Introduction

In magnetic elucidation problems, the anomalies owing to an inclined dike model have a wide use, e.g., oil, mineral, groundwater exploration and others (Smith et al. 2005; Abdelrahman et al. 2007a; Abdelrahman et al. 2012; Abo-Ezz and Essa 2016; Baiyegunhi and Gwavava 2017; Essa et al. 2018; Gadirov et al. 2018). However, in practice, an inclined dike configuration is often approximated by a vertical dike. This simple model may not be geologically accurate, but it is often times utilized in the geophysical interpretation to get the depth, width, and the thickness of a class of dike structures (Mehanee 2015; Mehanee and Essa 2015). Normal and rapid elucidation of such easy geometric models is administrated by hand through graphical approaches involving characteristic curves (Bean 1966; Koulomzine et al. 1970; Rao et al. 1972). However, the precision of the

estimated outcomes by these methods depends on the accuracy of the residual field separated from an observed field.

In addition, several numerical-approaches are adjusted to understand the magnetic anomaly caused by an inclined dike-like structure, such as curve matching techniques, which depend on trial and error of fitting the measured and calculated profile (Dondurur and Pamukçu 2003). Werner (1953) designed the deconvolution method (Werner deconvolution) to investigate inclined magnetized dikes by separating the anomaly owing to a selected dike from the interference of neighboring dikes. In spectral analysis approaches (Cassano and Rocca 1974; Sengupta and Das 1975; Bhimasankaram et al. 1978), the source depth is evaluated from the power band of their magnetic anomalies; the drawbacks of this method are that the precision of the outcomes is affected by the precision of the drawn interpolating line and it is difficult to apply this method when there are more than one body positioned at various depths (Cassano and Rocca 1974).

Rao et al. (1981) utilized a fancy gradient technique that depends on a few distinctive points (designated on the amplitude/phase plots) from which the variables of the dike are determined. Keating and Pilkington (1990) devised an

✉ Khalid S. Essa
khalid_sa_essa@yahoo.com

¹ Geophysics Department, Faculty of Science, Cairo University, Giza P.O. 12613, Egypt

automatic numerical routine to infer the vertical magnetic gradient profiles of an inclined dike. Abdelrahman et al. (2007b) suggested a semiautomatic least squares approach that convolves the inclined dike model with a moving average filter. However, the challenge facing this approach in estimating the model parameters is choosing the correct origin location of the dike from the measured magnetic data. Cooper (2012) discussed a new semiautomatic method to interpret the magnetic anomalies of dike by determining the fraction of the derivatives of the whole magnetic field. The disadvantage of this method is that it is applicable to thin dikes with different dips. Another method relies upon the amplitude of the analytic signal, which was devised by Cooper (2015) for the magnetic anomaly from a thin sheet to calculate the depth to the upper surface, amount of dip, and product of susceptibility and thickness of the dike. Al-Garni (2015) used a neural network method to estimate several parameters of dipping dikes. Essa and Elhussien (2017) discussed a semiautomatic method that depends on the second horizontal derivative of the measured magnetic anomaly for delineating dipping dikes sources. The shortcoming of the usage of higher order of derivatives is that sensitivity to noise is emphasized.

In this paper, we establish a new algorithm for calculating the response of an inclined dike model using the first derivative. The suggested method relies on evaluating the horizontal derivative anomalies from measured magnetic data utilizing several filters of successive window lengths (graticule spacings). The advantages of using this methodology are adjusted to estimate the depth and half-width of the covered inclined dike structures and also it is less sensitive to noise. The technique relies on calculating the variance (Var) of depths for every half-width. The tiniest variance is employed as a principle condition for deciding the most effective depth and half-width of the buried structure and does not need any graphical utilities. The accuracy of this methodology is demonstrated at synthetic examples, using simulated data generated from a model with random errors and a statistical distribution. The suggested method is then applied to real field examples from Turkey and Peru. There is an acceptable agreement between the results gained by our approach and those attained from different methods. Thus, the depths estimated from this approach are consistent with those attained from drilling information and published ones.

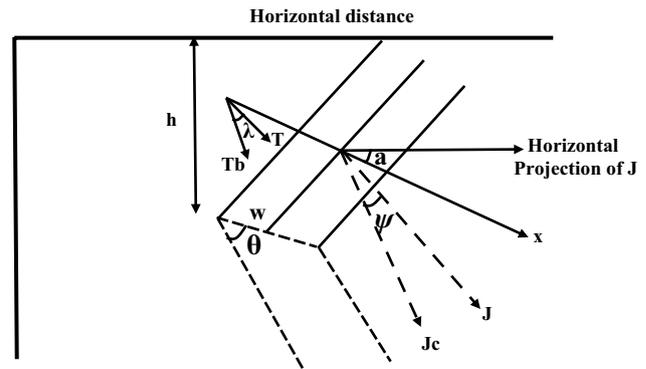


Fig. 1 A sketch diagram of inclined dike source

The method

Hood (1964), McGrath and Hood (1970) and Essa and Elhussien (2017) present an equation for the magnetic anomaly response of an inclined dike (Fig. 1) as follows:

$$H(x_j, h, w, \theta) = A_c \left[\sin \left(\theta \times \frac{\pi}{180} \right) \left(\tan^{-1} \left(\frac{x_j + w}{h} \right) - \tan^{-1} \left(\frac{x_j - w}{h} \right) \right) - \frac{\cos \left(\theta \times \frac{\pi}{180} \right)}{2} \ln \left(\frac{(x_j + w)^2 + h^2}{(x_j - w)^2 + h^2} \right) \right], \quad j = 1, 2, 3, 4, \dots, N \quad (1)$$

where h (m) is the depth to the upper surface of an inclined dike, w (m) is the half-width of the inclined dike, θ ($^\circ$) is the angle of magnetization, x_j (m) are the horizontal coordinates of the observation points, and A_c (nT) is the amplitude factor related to magnetization.

Using the following three points ($x_j - s$, x_j , $x_j + s$) on the anomaly profile, the first horizontal derivative (H_x) of the magnetic anomaly is given by subsequent expression:

$$H_x(x_j, h, w, \theta, s) = \frac{H(x_j + s) - H(x_j - s)}{2s}, \quad (2)$$

where $s = 1, 2, \dots, M$ spacing units which are called window lengths.

Substituting Eq. (1) in Eq. (2), the FHD (first horizontal derivative) of the magnetic anomaly is:

$$H_x(x_j, h, w, \theta, s) = \frac{A_c}{2s} \left[\sin \left(\theta \times \frac{\pi}{180} \right) \left(\tan^{-1} \left(\frac{x_j + s + w}{h} \right) + \tan^{-1} \left(\frac{-x_j - s + w}{h} \right) + \tan^{-1} \left(\frac{-x_j + s - w}{h} \right) + \tan^{-1} \left(\frac{x_j - s - w}{h} \right) \right) + \frac{\cos \left(\theta \times \frac{\pi}{180} \right)}{2} \ln \left(\frac{\left(\frac{(x_j - s + w)^2 + h^2}{(x_j - s - w)^2 + h^2} \right)}{\left(\frac{(x_j + s + w)^2 + h^2}{(x_j + s - w)^2 + h^2} \right)} \right) \right] \quad (3)$$

From the above equation and by rearrangement, we can gauge the depth (h_f), the half-width (w), the index angle (θ) and the amplitude factor (A_c) (“Appendix”) as follows:

$$h_f = \frac{3s + w}{\tan\left(\frac{P}{q}\right)} \tag{4}$$

where the factors P and q are defined in “Appendix”.

$$\theta_c = \tan^{-1} \left[\frac{\frac{H_x(+s) \ln\left(\frac{(s-w)^2+h^2}{(s+w)^2+h^2}\right)}{H_x(0)} - \frac{1}{2} \ln\left(\frac{(2s-w)^2+h^2}{(2s+w)^2+h^2}\right)}{\tan^{-1}\left(\frac{2s+w}{h}\right) + \tan^{-1}\left(\frac{w-2s}{h}\right) + 2 \tan^{-1}\left(\frac{-w}{h}\right)} \right] \tag{5}$$

$$A_c = \frac{2sH_x(0)}{\cos\theta \ln\left(\frac{(s-w)^2+h^2}{(s+w)^2+h^2}\right)} \tag{6}$$

The RMS (root-mean-square error) is estimated at various w values from the following equation:

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^N [H(x_j) - H_c(x_j)]^2}{N}} \tag{7}$$

where $H(x_j)$ is the measured field and $H_c(x_j)$ is the computed field. This is considered the misfit between the observed and calculated anomalies. The procedures and flowchart of applying our technique are summarized in Fig. 2.

The limitation of our method is based on: (a) the number of data points (larger than 21 points), (b) the convergence and stability of the minimizer, (c) a priori information accessible from different techniques that can be obtained as underlying model data.

Uncertainty analysis

We examined the uncertainties in estimating the model parameters (h, w, θ, A) of an inclined dike source by our method. For this purpose, each noise-free and noise-corrupted magnetic anomaly has been scrutinized and the impact of interference from neighboring bodies is considered.

Analysis of synthetic example

The total magnetic anomaly (ΔH) for an inclined dike with the parameters: $A_c = 1500$ nT, $\theta = 40^\circ$, $h = 9$ m, $w = 4$ m, and profile length = 100 m, is

$$\begin{aligned} \Delta H(x_j) &= A_c \left[\left(\sin\left(40^\circ \times \frac{\pi}{180}\right) \right) \left(\tan^{-1}\left(\frac{x_j+4}{9}\right) - \tan^{-1}\left(\frac{x_j-4}{9}\right) \right) \right. \\ &\quad \left. - \frac{\cos\left(40^\circ \times \frac{\pi}{180}\right)}{2} \ln\left(\frac{(x_j+4)^2+9^2}{(x_j-4)^2+9^2}\right) \right] \end{aligned} \tag{8}$$

Equation (2) is utilized to evaluate the horizontal derivative anomalies from the magnetic anomaly (ΔH) utilizing diverse s values ($s = 5, 6, 7, 8$ and 9 m). We estimated h

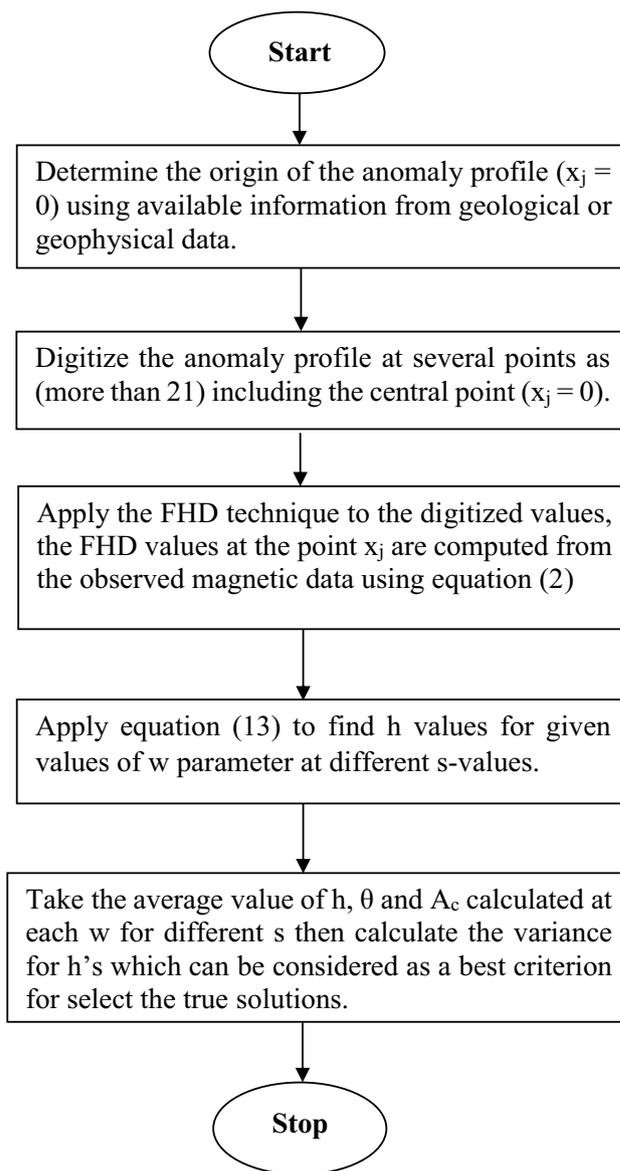


Fig. 2 Generalized scheme for parameters estimation by using our method

Table 1 Numerical results for a synthetic example with and without 10% random noise ($h=9$ m, $\theta=40^\circ$, $A=1500$ nT, $w=4$ m, and profile length = 100 m)

w (m)	Parameters	s (m)					Average	Variance	RMS error (nT)
		5	6	7	8	9			
Without random noise									
2	h (m)	10.08	9.95	9.86	9.79	9.73	9.88	0.019	16.780
	θ ($^\circ$)	40.03	39.82	39.63	39.50	39.41	39.68	0.062	
	A_c (nT)	3308.05	3243.50	3199.82	3167.38	3141.66	3212.08	4326.168	
2.5	h (m)	9.89	9.78	9.71	9.65	9.60	9.73	0.013	13.769
	θ ($^\circ$)	40.03	39.85	39.69	39.59	39.52	39.73	0.043	
	A_c (nT)	2602.81	2559.89	2531.01	2509.59	2492.63	2539.19	1896.490	
3	h (m)	9.65	9.57	9.52	9.47	9.43	9.53	0.007	10.066
	θ ($^\circ$)	40.02	39.89	39.77	39.69	39.64	39.80	0.023	
	A_c (nT)	2123.22	2096.83	2079.18	2066.14	2055.82	2084.24	709.624	
3.5	h (m)	9.36	9.31	9.28	9.26	9.23	9.29	0.002	5.499
	θ ($^\circ$)	40.01	39.94	39.87	39.83	39.80	39.89	0.007	
	A_c (nT)	1771.97	1759.43	1751.10	1744.98	1740.15	1753.52	158.209	
4	h (m)	9.00	9.00	9.00	9.00	9.00	9.00	0.000	0.000
	θ ($^\circ$)	40.00	40.00	40.00	40.00	40.00	40.00	0.000	
	A_c (nT)	1500.00	1500.00	1500.00	1500.00	1500.00	1500.00	0.000	
4.5	h (m)	8.57	8.63	8.67	8.70	8.73	8.66	0.004	6.622
	θ ($^\circ$)	39.99	40.08	40.16	40.21	40.24	40.14	0.011	
	A_c (nT)	1279.78	1291.91	1299.78	1305.50	1310.01	1297.40	142.710	
5	h (m)	8.05	8.19	8.28	8.35	8.41	8.26	0.020	14.579
	θ ($^\circ$)	39.97	40.19	40.36	40.47	40.53	40.30	0.053	
	A_c (nT)	1094.24	1118.92	1134.64	1146.00	1154.88	1129.74	574.383	
With 10% random noise									
2	h (m)	10.25	10.38	10.30	9.53	9.55	10.00	0.180	22.382
	θ ($^\circ$)	40.17	39.89	37.51	41.59	39.47	39.73	2.169	
	A_c (nT)	3374.51	3408.67	3369.71	3015.63	3050.18	3243.74	37,417.392	
2.5	h (m)	10.07	10.21	10.15	9.38	9.42	9.85	0.168	20.016
	θ ($^\circ$)	40.16	39.91	37.57	41.48	39.58	39.74	2.001	
	A_c (nT)	2655.82	2691.61	2666.48	2388.97	2419.82	2564.54	21,659.687	
3	h (m)	9.83	10.00	9.96	9.20	9.25	9.65	0.152	17.345
	θ ($^\circ$)	40.15	39.93	37.43	41.40	39.71	39.73	2.071	
	$A_c A_c$ (nT)	2167.26	2206.18	2191.64	1966.41	1995.52	2105.40	13,202.675	
3.5	h (m)	9.54	9.75	9.73	8.99	9.05	9.41	0.135	14.510
	θ ($^\circ$)	40.14	39.97	37.72	41.75	39.88	39.89	2.057	
	A_c (nT)	1809.56	1852.75	1847.06	1660.31	1688.84	1771.70	8238.740	
4	h (m)	9.19	9.34	9.45	8.73	8.81	9.11	0.103	12.587
	θ ($^\circ$)	40.12	40.01	37.83	41.73	40.09	39.95	1.931	
	A_c (nT)	1532.81	1581.31	1583.55	1426.73	1455.50	1515.98	5184.879	
4.5	h (m)	8.77	9.08	9.13	8.42	8.24	8.73	0.155	13.154
	θ ($^\circ$)	40.10	40.07	37.96	41.95	40.34	40.08	2.018	
	A_c (nT)	1328.88	1393.89	1373.68	1218.22	1230.85	1309.10	6533.090	
5	h (m)	8.35	8.66	8.75	8.02	8.11	8.38	0.105	14.301
	θ ($^\circ$)	40.07	40.14	38.13	42.24	40.65	40.24	2.161	
	A_c (nT)	1150.44	1183.52	1284.84	1089.00	1120.05	1165.57	5677.898	

Bold represents the best-fit parameters estimated

values at various w for each s value and afterward calculated the average depth and variance (Table 1) by using our inversion approach. Table 1 demonstrates the outcomes for noise-free data. The estimated parameters from the planned method are in fair agreement with the model of the 2D inclined dike ($h=9$ m and $w=4$ m). In the end, we can observe that the minimum variance ($\text{Var}=0$) occurs at the true depth ($h=9$ m) and true half-width ($w=4$ m).

Subsequently, the real data are polluted with random noise; we imposed 10% random noise on the composite anomaly to see the impact of this noise on our algorithm. The FHD anomalies were computed using the same s values mentioned above.

Table 1 also shows the computational results for the noisy magnetic data. The average depth is 9.11 m, which is an error of 1.22% from the true h value and the estimated half-width is $w=4$ m. The obtained solution for the different parameters corresponds to a minimum variance with an RMS which equal to 12.587 nT. This demonstrates that our method is valuable when applied to noisy magnetic data.

Analysis of the interference effect

The magnetic anomalies may be distorted due to nearby structures; a synthetic model composed of a 2D inclined dike (with $h=9$ m, $\theta=40^\circ$, $A_c=1500$ nT and $w=4$ m) and a vertical thin dike (with $h=4$ m, $\alpha=45^\circ$, $K=2500$ nT and $x_o=40$ m) was computed using a profile length of 120 m. The analysis by our method was applied to this synthetic data to evaluate and study the response of this neighboring structure on assessing the body parameters (h , w , θ and A_c) by using our new approach.

Figure 3 shows the composite magnetic anomaly due to the two structures with and without 10% random noise. The FHD technique was utilized to the magnetic anomaly without random noise using four graticule spacings ($s=5$, 6, 7 and 8 m).

The values of h at several w for each s value were appraised by our method, and then the average and Var values of the different factors were calculated at various w for each s value (Table 2). From Table 2, the average depth of 8.33 m, w of 4.5 m, θ of 40.04° and A_c of 1208.03 nT correspond to the minimum RMS ($\text{RMS}=136.337$ nT). The percentage error in depth is 7.44%, in half-width is 12.5%, in magnetization angle is 0.1% and in amplitude factor is 19.46%.

The FHD technique was repeated with 10% random noise for the same previous spacings ($s=5$, 6, 7 and 8 m). From Table 2, the average depth h_c is 8.19 m, w is 4.5 m, θ_c is

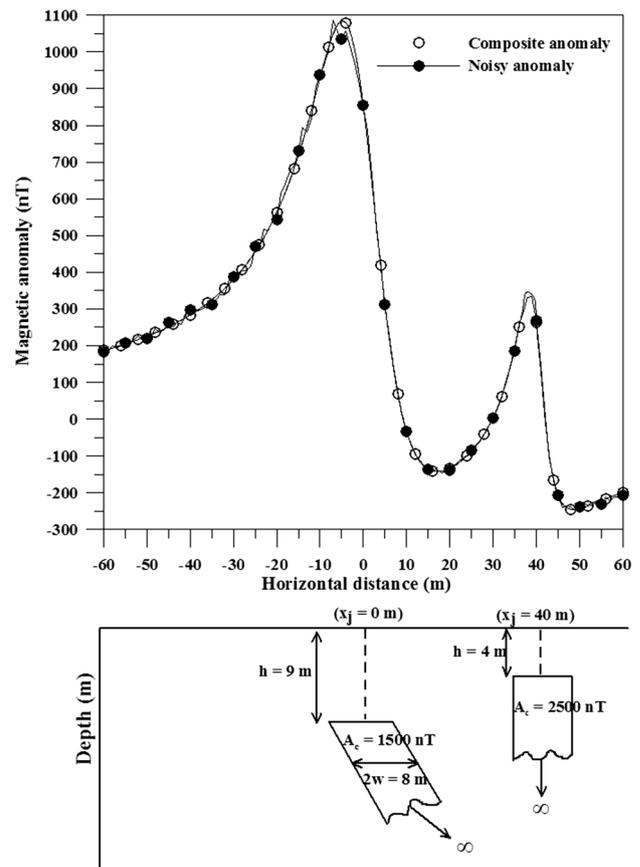


Fig. 3 A composite magnetic field anomaly of a buried inclined dike with $h=9$ m, $\theta=40^\circ$, $A_c=1500$ nT and $w=4$ m, and a vertical thin dike with $h=4$ m, $\alpha=45^\circ$, $K=2500$ nT and $x_o=40$ m, and profile length=120 m with and without 10% random noise and a sketch diagram of the 2 sources

40.99° and A_c is 1176.46 nT; the corresponding minimum ($\text{RMS}=134.398$ nT). The error of computed depth is 9% from the true depth (9 m), half-width is 12.5% from the true value (4 m), magnetization angle is 2.475% from the true value (40°); finally, the amplitude factor is 21.57% from the actual value (1500 nT).

Analysis of the effect of choosing origin

The selection of an incorrect origin of the magnetic profile from real data leads to errors in determining the inclined dike structure parameters. To determine the implication of choosing an incorrect origin, we used a synthetic model of $A_c=1500$ nT, $\theta=50^\circ$, $h=9$ m, $w=4$ m, and profile length=100 m) with an error (1.5 m) in the source location (see Eq. 9) as follows:

Table 2 Numerical results for a synthetic example composed of a 2D inclined dike (with $h=9$ m, $\theta=40^\circ$, $A_c=1500$ nT and $w=4$ m), and a vertical dike (with $h=4$ m, $\alpha=45^\circ$, $K=2500$ nT and $x_o=40$ m) with and without 10% random noise

w (m)	Parameters	s (m)				Average	Variance	RMS error (nT)
		5	6	7	8			
Without random noise								
2	h (m)	9.93	9.72	9.50	9.22	9.59	0.092	137.853
	θ ($^\circ$)	40.00	39.73	39.46	39.15	39.59	0.131	
	A_c (nT)	3176.17	3073.87	2977.91	2872.04	3025.00	16,951.172	
2.5	h (m)	9.74	9.55	9.34	9.08	9.43	0.081	138.152
	θ ($^\circ$)	39.99	39.77	39.54	39.26	39.64	0.099	
	A_c (nT)	2498.44	2425.31	2354.59	2274.55	2388.22	9192.195	
3	h (m)	9.50	9.33	9.15	8.89	9.22	0.068	138.609
	θ ($^\circ$)	39.99	39.82	39.63	39.39	39.71	0.065	
	A_c (nT)	2037.46	1985.84	1933.35	1871.55	1957.05	5055.692	
3.5	h (m)	9.20	9.07	8.91	8.67	8.96	0.052	139.323
	θ ($^\circ$)	39.98	39.88	39.75	39.56	39.79	0.033	
	A_c (nT)	1699.71	1665.47	1627.29	1579.48	1642.99	2667.185	
4	h (m)	8.84	8.76	8.62	8.41	8.66	0.036	140.333
	θ ($^\circ$)	39.98	39.95	39.90	39.78	39.90	0.008	
	A_c (nT)	1438.07	1418.98	1392.86	1356.50	1401.60	1247.512	
4.5	h (m)	8.41	8.37	8.29	8.23	8.33	0.006	136.337
	θ ($^\circ$)	39.97	40.06	40.10	40.04	40.04	0.003	
	A_c (nT)	1226.03	1221.09	1205.73	1179.263	1208.03	442.457	
5	h (m)	7.89	7.93	7.89	7.74	7.86	0.007	141.172
	θ ($^\circ$)	39.96	40.19	40.34	40.37	40.22	0.036	
	A_c (nT)	1147.21	1156.39	1151.17	1087.69	1135.62	1034.965	
With 10% random noise								
2	h (m)	9.05	9.08	10.62	9.43	9.54	0.540	135.247
	θ ($^\circ$)	44.41	42.05	36.29	39.46	40.55	12.164	
	A_c (nT)	2773.66	2800.02	3361.49	2936.05	2967.81	73,947.141	
2.5	h (m)	8.85	8.90	10.26	9.29	9.33	0.429	135.758
	θ ($^\circ$)	44.41	42.10	36.34	39.56	40.60	12.015	
	A_c (nT)	2178.22	2207.29	2660.25	2325.65	2342.85	48,839.218	
3	h (m)	8.60	8.68	10.07	9.11	9.12	0.459	136.399
	θ ($^\circ$)	44.42	42.17	36.41	39.68	40.67	11.830	
	A_c (nT)	1772.42	1805.25	2186.81	1914.02	1919.63	35,390.776	
3.5	h (m)	8.29	8.41	9.85	8.89	8.86	0.502	137.313
	θ ($^\circ$)	44.43	42.26	36.49	39.84	40.76	11.609	
	A_c (nT)	1474.29	1511.73	1843.31	1615.79	1611.28	27,512.103	
4	h (m)	7.91	8.08	9.57	8.63	8.55	0.562	138.631
	θ ($^\circ$)	44.45	42.37	36.59	40.04	40.86	11.353	
	A_c (nT)	1242.48	1285.46	1580.68	1388.16	1374.20	22,684.575	
4.5	h (m)	7.65	7.89	9.11	8.12	8.19	0.414	134.398
	θ ($^\circ$)	44.38	42.52	36.76	40.28	40.99	10.727	
	A_c (nT)	1053.65	1103.32	1341.55	1207.32	1176.46	16,212.923	
5	h (m)	6.88	7.22	8.88	7.97	7.74	0.786	143.002
	θ ($^\circ$)	44.52	42.72	36.89	40.58	41.18	10.759	
	A_c (nT)	893.17	951.11	1199.41	1058.88	1025.64	18,134.564	

Bold represents the best-fit parameters estimated

Table 3 Numerical results for a synthetic example composed of a 2D inclined dike (with $h=9$ m, $\theta=50^\circ$, $A_c=1500$ nT and $w=4$ m), with incorrect origin

w (m)	Parameters	s (m)				Average	Variance	RMS error (nT)
		5	6	7	8			
2	h (m)	9.25	9.06	8.91	8.79	9.00	0.038	88.894
	θ ($^\circ$)	57.07	55.83	54.38	52.86	55.04	3.317	
	A_c (nT)	3122.66	3040.55	2982.16	2938.89	3021.06	6322.521	
2.5	h (m)	9.05	8.88	8.75	8.64	8.83	0.031	87.734
	θ ($^\circ$)	57.06	55.86	54.44	52.95	55.08	3.152	
	A_c (nT)	2452.09	2396.34	2356.67	2327.48	2383.15	2909.129	
3	h (m)	8.80	8.66	8.55	8.46	8.62	0.022	86.464
	θ ($^\circ$)	57.04	55.89	54.52	53.07	55.13	2.946	
	A_c (nT)	1995.06	1959.30	1933.78	1915.10	1950.81	1198.323	
3.5	h (m)	8.50	8.39	8.30	8.23	8.35	0.013	85.357
	θ ($^\circ$)	57.01	55.93	54.62	53.21	55.19	2.696	
	A_c (nT)	1659.33	1640.15	1626.27	1616.27	1635.50	348.218	
4	h (m)	8.12	8.06	8.01	7.97	8.04	0.004	83.692
	θ ($^\circ$)	56.28	55.99	54.75	53.60	55.15	1.513	
	A_c (nT)	1398.38	1394.07	1390.47	1388.12	1392.76	20.004	
4.5	h (m)	7.64	7.66	7.79	7.62	7.68	0.006	84.439
	θ ($^\circ$)	56.94	56.07	54.92	53.63	55.39	2.061	
	A_c (nT)	1185.92	1195.94	1202.01	1206.78	1197.66	80.937	
5	h (m)	7.11	7.19	7.24	7.28	7.21	0.005	85.765
	θ ($^\circ$)	56.90	56.18	55.14	53.92	55.53	1.674	
	A_c (nT)	1005.60	1030.32	1046.05	1057.85	1034.96	510.024	

Bold represents the best-fit parameters estimated

$$\Delta H(x_i) = A_c \left[\left(\sin \left(50^\circ \times \frac{\pi}{180} \right) \right) \left(\tan^{-1} \left(\frac{(x_j - 1.5) + 4}{9} \right) - \tan^{-1} \left(\frac{(x_j - 1.5) - 4}{9} \right) \right) - \frac{\cos \left(50^\circ \times \frac{\pi}{180} \right)}{2} \times \ln \left(\frac{((x_j - 1.5) + 4)^2 + 9^2}{((x_j - 1.5) - 4)^2 + 9^2} \right) \right]. \tag{9}$$

Our investigation begins by utilizing Eq. (2) for the FHD separation anomalies from the magnetic anomaly (ΔH) using available s values ($s=5, 6, 7$ and 8 m).

Table 3 displays the results after applying the error in the horizontal coordinate. The estimated parameters (h , w , θ and A_c) are 8.04 m, 4 m, 55.15° and 1392.76 nT, respectively, and correspond to the minimum RMS of 83.692 nT. The error in computed depth is 10.66%, in half-width is 0%, in index parameter is 10.3% while the error of estimated amplitude factor is 7.15%.

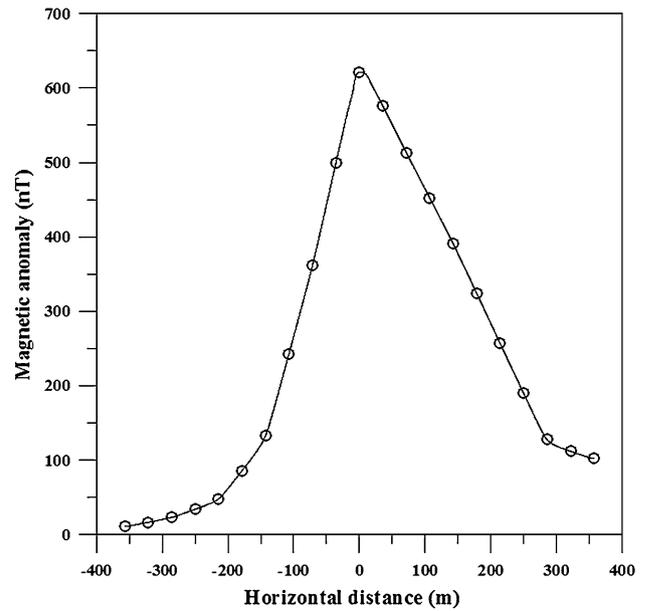


Fig. 4 A vertical component magnetic anomaly Bayburt–Sarihan (northeast of Turkey) skarn dike by Dondurur and Pamukçu (2003)

Field examples

To test the validity and the rationality of the anticipated approach, we have explored two real field cases with increasing complexity of the geological sources collected from the available published literature.

Magnetic anomaly of the Bayburt–Sarihan skarn zone, Turkey

A vertical component magnetic field profile was taken over the Bayburt–Sarihan (northeast of Turkey) skarn dike by Dondurur and Pamukçu (2003) (Fig. 4). The length of this profile is 714.5 m, and it was digitized at 8.93 m sample interval. We have applied our new technique to determine the body parameters (z , d) using five successive windows

($s = 62.51, 71.44, 80.37, 89.3,$ and 98.23 m) for the horizontal derivative anomalies. Table 4 shows that the parameters with the best-fit parameters are: $w = 64.74$ m, $h = 87.65$ m, $\theta = -71.37^\circ$ and $A_c = -409.25$ nT (RMS = 69.353). Table 5 shows the comparison with other interpretation methods from the published literature.

Magnetic anomaly in the Marcona district, Peru

The magnetic field profile taken by Gay (1963) near the magnetic equator in the Marcona district, Peru, is shown in Fig. 5. The profile length is 1125 m and was digitized at 14.06 m. Following the previous outlined process, the horizontal derivative anomalies were calculated from Eq. (2) by using seven s value ($s = 126.54, 140.6, 154.66, 168.72, 182.78, 196.84$ and 210.9 m). For our algorithm,

Table 4 Numerical results for Bayburt–Sarihan (northeast of Turkey) skarn zone field example

w (m)	Parameters	s (m)					Average	Variance	RMS error (nT)
		62.51	71.44	80.37	89.3	98.23			
62.51	h (m)	-71.44	-71.19	-71.31	-71.49	-71.46	90.62	221.299	70.726
	θ (°)	-71.44	-71.19	-71.31	-71.49	-71.46	-71.38	0.016	
	A_c (nT)	-297.61	-362.74	-438.02	-518.46	-574.60	-438.29	12,626.513	
64.74	h (m)	67.44	78.70	89.65	99.85	102.63	87.65	216.427	69.353
	θ (°)	-71.38	-71.21	-71.30	-71.49	-71.46	-71.37	0.013	
	A_c (nT)	-277.08	-341.67	-415.35	-493.68	-518.49	-409.25	10,283.907	
66.98	h (m)	63.74	75.84	87.29	97.83	103.81	85.70	264.218	76.328
	θ (°)	-71.32	-71.13	-71.28	-71.48	-71.46	-71.33	0.020	
	A_c (nT)	-256.98	-321.38	-393.73	-470.17	-523.82	-393.22	11,675.083	
69.21	h (m)	59.60	72.74	84.77	95.68	101.90	82.94	293.448	79.520
	θ (°)	-71.24	-71.10	-71.27	-71.48	-71.46	-71.31	0.025	
	A_c (nT)	-237.06	-301.74	-373.02	-447.81	-500.41	-372.01	11,345.857	
71.44	h (m)	54.88	69.34	82.07	93.41	99.88	79.92	330.915	83.035
	θ (°)	-71.15	-71.06	-71.25	-71.47	-71.46	-71.28	0.034	
	A_c (nT)	-216.96	-282.57	-353.10	-426.45	-478.14	-351.44	11,127.507	
73.67	h (m)	49.32	65.60	79.16	90.99	97.75	76.57	381.454	86.962
	θ (°)	-71.04	-71.02	-71.24	-71.47	-71.47	-71.25	0.048	
	A_c (nT)	-196.10	-263.68	-333.82	-405.96	-456.88	-331.29	11,050.705	

Bold represents the best-fit parameters estimated

Table 5 Comparison between numerical results of different methods for Bayburt–Sarihan (northeast of Turkey) skarn zone field example

Parameters	Method					
	Aydin and Gelişli method (1996)	Dondurur and Pamukçu method (2003)			Abdelrahman et al. method (2007b)	Present method
		Inverse solution	Hilbert transform	Power spectrum		
h_{avg} (m)	94	97	100	98	100	87.65
w (m)	66	76	75	70	79	64.74
θ (°)	–	–	–	–	–	-71.37
A_c (nT)	–	–	–	–	–	-409.25

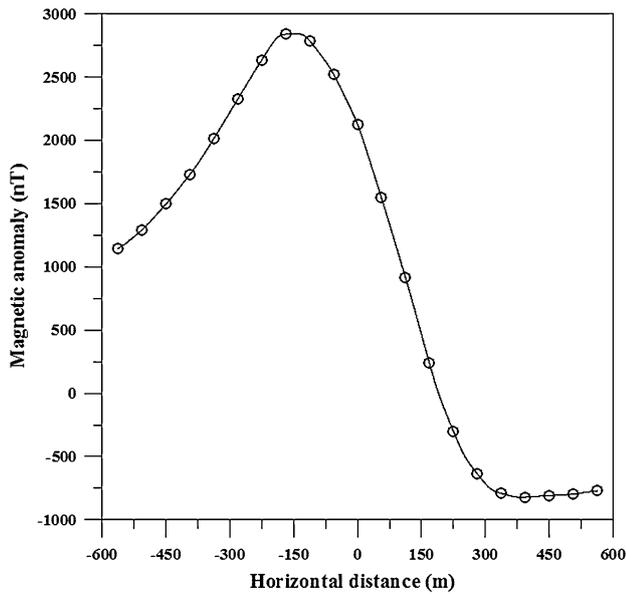


Fig. 5 Observed magnetic anomaly near the magnetic equator at Marcona district, Peru (Gay 1963)

the retrieved parameters are as follows: $h = 138.28$ m, $w = 196.84$ m, $\theta = 40.49^\circ$ and $A_c = 1862.75$ nT which are the best-fit model parameters (RMS = 48.5567 nT) (Table 6). The estimated results from our process are in sensible agreement with those published in the literature (Table 7).

Conclusions

The estimation of the parameters of an inclined dike are very important in geophysical exploration. We have devised an algorithm, which is based on the first horizontal derivative technique to gauge the body parameters. Our new approach is easy, semiautomatic and it does not necessitate any graphical supports. To verify the accurateness and pertinence of our method, the approach has been applied to synthetic data without and with random noise, and applied it also to a more complicated synthetic model including interference effect, again with and without random noise.

From our outcomes, we show that the best-fit parameters for the inclined dike can be determined with reasonable accuracy from our new algorithm even if the observed data are tainted with noise. Furthermore, the method has been

Table 6 Numerical results for Marcona district field example, Peru

w (m)	Parameters	s (m)								Average	Variance	RMS error (nT)
		126.54	140.6	154.66	168.72	182.78	196.84	210.9				
182.78	h (m)	162.75	161.89	161.78	169.32	171.39	218.26	264.72	187.16	1571.434	167.75	
	θ ($^\circ$)	43.25	42.90	42.78	41.40	40.27	37.45	35.45	40.50	9.029		
	A_c (nT)	2015.96	2001.90	2012.43	2084.94	2104.35	2581.12	3075.90	2268.09	168,446.901		
186.295	h (m)	151.30	152.95	154.12	162.74	165.46	213.67	260.92	180.17	1735.987	147.413	
	θ ($^\circ$)	43.01	42.83	42.83	41.52	40.44	37.54	35.50	40.53	8.664		
	A_c (nT)	1872.86	1889.78	1915.10	1996.94	2023.01	2498.11	2987.91	2169.10	176,661.701		
189.81	h (m)	138.10	143.06	145.75	155.64	159.12	208.87	256.97	172.50	1947.912	124.877	
	θ ($^\circ$)	42.67	42.74	42.90	41.66	40.64	37.64	35.55	40.54	8.253		
	A_c (nT)	1722.36	1775.31	1816.74	1908.94	1942.11	2416.72	2902.14	2069.19	187,800.103		
193.325	h (m)	122.55	131.94	136.48	147.94	152.29	203.82	252.87	163.98	2230.391	100.424	
	θ ($^\circ$)	42.15	42.62	42.98	41.83	40.88	37.74	35.61	40.54	7.791		
	A_c (nT)	1561.62	1657.20	1716.34	1820.29	1861.19	2336.77	2818.42	1967.40	202,915.960		
196.84	h (m)	103.96	119.25	116.08	129.50	134.93	159.58	204.68	138.28	1163.967	48.5567	
	θ ($^\circ$)	41.26	42.43	43.07	42.03	41.15	37.86	35.67	40.49	7.327		
	A_c (nT)	1588.76	1533.68	1612.39	1730.15	1779.66	2258.01	2536.59	1862.75	146,803.465		
200.355	h (m)	82.56	104.36	113.95	129.98	136.71	192.94	244.17	143.52	3154.515	49.782	
	θ ($^\circ$)	39.60	42.12	43.19	42.27	41.47	37.49	35.73	40.27	7.700		
	A_c (nT)	1211.73	1402.29	1502.24	1637.19	1696.66	2180.25	2656.48	1755.26	249,038.578		
203.87	h (m)	62.28	86.38	99.06	119.06	127.55	187.03	239.55	131.56	3801.377	51.463	
	θ ($^\circ$)	36.75	41.55	43.36	42.60	41.88	38.13	35.80	40.01	9.285		
	A_c (nT)	1063.55	1259.95	1379.91	1539.05	1610.86	2103.22	2577.95	1647.78	274,887.994		

Bold represents the best-fit parameters estimated

Table 7 Comparison between numerical results of different methods for Marcona district field example, Peru

Parameters	Method					
	Gay method (1963)	Koulomzine et al. method (1970)		Pal method (1985)	Al-Garni method (2015)	Present method
		With (φ)	With (μ)			
h_{averg} (m)	124	126.7	135.5	132.6	130	138.28
w (m)	186	205.95	202.75	193.75	191.7	196.84
θ ($^\circ$)	–	–	–	–	–	40.49
A_c (nT)	–	–	–	–	–	1862.75

relevantly applied to two field data from Turkey and Peru, obtained over mineral rich deposits. The limitation of our approach concerns the convergence of s -values, i.e., not all s -values can give results. The good agreement of the results acquired from our method with those published in the literature shows that the inclined dike parameters can be reasonably and efficiently determined.

Acknowledgements We would like to thank Prof. Eleftheria E. Papadimitriou, Editor-in-Chief, Prof. Ralf Schaa, Associate Editor, and the two capable expert reviewers for their keen interest, valuable comments on the manuscript, and improvements to this work.

Appendix

Hood (1964), McGrath and Hood (1970) and Essa and Elhussein (2017) represent an equation for the total magnetic anomaly of an inclined dike (Fig. 1) as follows:

$$\begin{aligned}
 H(x_j, h, w, \theta) &= A_c \left[\sin \left(\theta \times \frac{\pi}{180} \right) \left(\tan^{-1} \left(\frac{x_j + w}{h} \right) - \tan^{-1} \left(\frac{x_j - w}{h} \right) \right) \right. \\
 &\quad \left. - \frac{\cos \left(\theta \times \frac{\pi}{180} \right)}{2} \ln \left(\frac{(x_j + w)^2 + h^2}{(x_j - w)^2 + h^2} \right) \right], \\
 j &= 1, 2, 3, 4, \dots N
 \end{aligned}
 \tag{10}$$

where h (m) is the depth to the top of the inclined dike, w (m) is the half-width of the inclined dike, θ ($^\circ$) is the angle of magnetization (index parameter), x_j (m) are the horizontal coordinates, and A (nT) is the amplitude factor.

Using three observation points ($x_j - s, x_j, x_j + s$) along the anomaly profile; the first horizontal derivative (H_x) of the total magnetic anomaly is given by the following expression:

$$H_x(x_j, h, w, \theta, s) = \frac{H(x_j + s) - H(x_j - s)}{2s}, \tag{11}$$

where $s = 1, 2, \dots M$ spacing units which is called the graticule spacing or window length.

Substituting Eq. (10) in Eq. (11), the first horizontal derivative (FHD) of the total magnetic anomaly is given by:

$$\begin{aligned}
 H_x(x_j, h, w, \theta, s) &= \frac{A_c}{2s} \left[\sin \left(\theta \times \frac{\pi}{180} \right) \left(\tan^{-1} \left(\frac{x_j + s + w}{h} \right) \right. \right. \\
 &\quad \left. \left. + \tan^{-1} \left(\frac{-x_j - s + w}{h} \right) \right) \right. \\
 &\quad \left. + \tan^{-1} \left(\frac{-x_j + s - w}{h} \right) + \tan^{-1} \left(\frac{x_j - s - w}{h} \right) \right) \\
 &\quad \left. + \frac{\cos \left(\theta \times \frac{\pi}{180} \right)}{2} \ln \left[\frac{\left(\frac{(x_j - s + w)^2 + h^2}{(x_j - s - w)^2 + h^2} \right)}{\left(\frac{(x_j + s + w)^2 + h^2}{(x_j + s - w)^2 + h^2} \right)} \right] \right].
 \end{aligned}
 \tag{12}$$

By substituting $x_j = 0$ in Eq. (12), the amplitude factor (A_c) can be estimated as follows when $s > w$:

$$A_c = \frac{2sH_x(0)}{\cos \left(\theta \times \frac{\pi}{180} \right) \ln \left(\frac{(s-w)^2 + h^2}{(s+w)^2 + h^2} \right)}. \tag{13}$$

Using Eq. (13), Eq. (12) can be written as the follows:

$$\begin{aligned}
 H_x(x_j, h, w, \theta, s) &= \frac{H_x(0)}{\cos \left(\theta \times \frac{\pi}{180} \right) \ln \left(\frac{(s-w)^2 + h^2}{(s+w)^2 + h^2} \right)} \\
 &\quad \times \left[\sin \left(\theta \times \frac{\pi}{180} \right) \left(\tan^{-1} \left(\frac{x_j + s + w}{h} \right) \right. \right. \\
 &\quad \left. \left. + \tan^{-1} \left(\frac{-x_j - s + w}{h} \right) \right) \right. \\
 &\quad \left. + \tan^{-1} \left(\frac{-x_j + s - w}{h} \right) + \tan^{-1} \left(\frac{x_j - s - w}{h} \right) \right) \\
 &\quad \left. + \frac{\cos \theta}{2} \ln \left[\frac{\left(\frac{(x_j - s + w)^2 + h^2}{(x_j - s - w)^2 + h^2} \right)}{\left(\frac{(x_j + s + w)^2 + h^2}{(x_j + s - w)^2 + h^2} \right)} \right] \right].
 \end{aligned}
 \tag{14}$$

Putting $x_j = +s, x_j = -s, x_j = +2s, x_j = -2s$, we get the following four equations, respectively:

$$\begin{aligned}
 H_x(+s) &= \frac{H_x(0)}{\cos\left(\theta \times \frac{\pi}{180}\right) \ln\left(\frac{(s-w)^2+h^2}{(s+w)^2+h^2}\right)} \\
 &\times \left[\sin\left(\left(\theta \times \frac{\pi}{180}\right) \times \frac{\pi}{180}\right) \left(\tan^{-1}\left(\frac{2s+w}{h}\right) + \tan^{-1}\left(\frac{w-2s}{h}\right) + 2 \tan^{-1}\left(-\frac{w}{h}\right)\right) \right. \\
 &\left. + \frac{\cos\left(\theta \times \frac{\pi}{180}\right)}{2} \ln\left(\frac{(2s-w)^2+h^2}{(2s+w)^2+h^2}\right) \right] \tag{15}
 \end{aligned}$$

$$\begin{aligned}
 H_x(-s) &= \frac{H_x(0)}{\cos\left(\theta \times \frac{\pi}{180}\right) \ln\left(\frac{(s-w)^2+h^2}{(s+w)^2+h^2}\right)} \\
 &\times \left[\sin\left(\theta \times \frac{\pi}{180}\right) \left(2 \tan^{-1}\left(\frac{w}{h}\right) + \tan^{-1}\left(\frac{2s-w}{h}\right) + \tan^{-1}\left(\frac{-2s-w}{h}\right)\right) \right. \\
 &\left. + \frac{\cos\left(\theta \times \frac{\pi}{180}\right)}{2} \ln\left(\frac{(2s-w)^2+h^2}{(2s+w)^2+h^2}\right) \right] \tag{16}
 \end{aligned}$$

$$\begin{aligned}
 H_x(+2s) &= \frac{H_x(0)}{\cos\left(\theta \times \frac{\pi}{180}\right) \ln\left(\frac{(s-w)^2+h^2}{(s+w)^2+h^2}\right)} \\
 &\times \left[\sin\left(\theta \times \frac{\pi}{180}\right) \left(\tan^{-1}\left(\frac{3s+w}{h}\right) + \tan^{-1}\left(\frac{w-3s}{h}\right) + \tan^{-1}\left(\frac{-s-w}{h}\right)\right) \right. \\
 &\left. + \tan^{-1}\left(\frac{s-w}{h}\right) + \frac{\cos\left(\theta \times \frac{\pi}{180}\right)}{2} \ln\left(\frac{\frac{(s+w)^2+h^2}{(s-w)^2+h^2}}{\frac{(3s+w)^2+h^2}{(3s-w)^2+h^2}}\right) \right] \tag{17}
 \end{aligned}$$

$$\begin{aligned}
 H_x(-2s) &= \frac{H_x(0)}{\cos\left(\theta \times \frac{\pi}{180}\right) \ln\left(\frac{(s-w)^2+h^2}{(s+w)^2+h^2}\right)} \\
 &\times \left[\sin\left(\theta \times \frac{\pi}{180}\right) \left(\tan^{-1}\left(\frac{w-s}{h}\right) + \tan^{-1}\left(\frac{s+w}{h}\right) + \tan^{-1}\left(\frac{3s-w}{h}\right) + \tan^{-1}\left(\frac{-3s-w}{h}\right)\right) \right. \\
 &\left. + \frac{\cos\left(\theta \times \frac{\pi}{180}\right)}{2} \ln\left(\frac{\frac{(3s-w)^2+h^2}{(3s+w)^2+h^2}}{\frac{(s-w)^2+h^2}{(s+w)^2+h^2}}\right) \right] \tag{18}
 \end{aligned}$$

By subtracting Eq. (16) from Eq. (15) and Eq. (18) from Eq. (17), we get the following two equations, respectively:

$$\begin{aligned}
 H_x(+s) - H_x(-s) &= \frac{H_x(0) \sin\left(\theta \times \frac{\pi}{180}\right)}{\cos\left(\theta \times \frac{\pi}{180}\right) \ln\left(\frac{(s-w)^2+h^2}{(s+w)^2+h^2}\right)} \\
 &\times \left[2 \tan^{-1}\left(\frac{2s+w}{h}\right) + 2 \tan^{-1}\left(\frac{w-2s}{h}\right) + 4 \tan^{-1}\left(-\frac{w}{h}\right) \right] \tag{19}
 \end{aligned}$$

$$\begin{aligned}
 H_x(+2s) - H_x(-2s) &= \frac{H_x(0) \sin\left(\theta \times \frac{\pi}{180}\right)}{\cos\left(\theta \times \frac{\pi}{180}\right) \ln\left(\frac{(s-w)^2+h^2}{(s+w)^2+h^2}\right)} \\
 &\times \left[2 \tan^{-1}\left(\frac{3s+w}{h}\right) + 2 \tan^{-1}\left(\frac{w-3s}{h}\right) + 2 \tan^{-1}\left(\frac{-s-w}{h}\right) + 2 \tan^{-1}\left(\frac{s-w}{h}\right) \right] \tag{20}
 \end{aligned}$$

Then by dividing Eq. (19) by Eq. (20), we get the following:

$$q = \frac{\tan^{-1}\left(\frac{2s+w}{h}\right) + \tan^{-1}\left(\frac{w-2s}{h}\right) + 2 \tan^{-1}\left(-\frac{w}{h}\right)}{\tan^{-1}\left(\frac{3s+w}{h}\right) + \tan^{-1}\left(\frac{w-3s}{h}\right) + \tan^{-1}\left(\frac{-s-w}{h}\right) + \tan^{-1}\left(\frac{s-w}{h}\right)}, \tag{21}$$

where

$$q = \frac{H_x(+s) - H_x(-s)}{H_x(+2s) - H_x(-2s)}.$$

From Eq. (21) and by rearrangement, we can calculate h_f from the following equation:

$$h_f = \frac{3s + w}{\tan\left(\frac{p}{q}\right)}, \quad (22)$$

where

$$P = q \tan^{-1}\left(\frac{3s - w}{h_i}\right) + q \tan^{-1}\left(\frac{s + w}{h_i}\right) + q \tan^{-1}\left(\frac{w - s}{h_i}\right) + \tan^{-1}\left(\frac{2s + w}{h_i}\right) + \tan^{-1}\left(\frac{w - 2s}{h_i}\right) + 2 \tan^{-1}\left(-\frac{w}{h_i}\right).$$

Equation (22) can be deciphered for h using the standard methods for solving nonlinear equations (Press et al. 1986), and its iteration form can be expressed as:

$$h_f = f(h_i), \quad (23)$$

where h_i is the initial depth estimate and h_f is the revised depth, for the next iteration h_f will be used as h_i . The iteration stops when $|h_f - h_i| \leq e$, where e is a small predetermined real number close to zero.

From Eq. (15), we can calculate the magnetization angle as follows:

$$\theta c = \tan^{-1} \left[\frac{H_x(+s) \cdot \ln\left(\frac{(s-w)^2 + h^2}{(s+w)^2 + h^2}\right) - \frac{1}{2} \ln\left(\frac{(2s-w)^2 + h^2}{(2s+w)^2 + h^2}\right)}{M_x(0) \left[\tan^{-1}\left(\frac{2s+w}{h}\right) + \tan^{-1}\left(\frac{w-2s}{h}\right) + 2 \tan^{-1}\left(\frac{-w}{h}\right) \right]} \right]. \quad (24)$$

References

- Abdelrahman EM, Abo-Ezz ER, Essa KS, El-Araby TM, Soliman KS (2007a) A new least-squares minimization approach to depth and shape determination from magnetic data. *Geophys Prospect* 55:433–446
- Abdelrahman EM, Abo-Ezz ER, Soliman KS, EL-Araby TM, Essa KS (2007b) A least-squares window curves method for interpretation of magnetic anomalies caused by dipping dikes. *Pure Appl Geophys* 164:1027–1044
- Abdelrahman EM, Abo-Ezz ER, Essa KS (2012) Parametric inversion of residual magnetic anomalies due to simple geometric bodies. *Explor Geophys* 43:178–189
- Abo-Ezz ER, Essa KS (2016) A least-squares minimization approach for model parameters estimate by using a new magnetic anomaly formula. *Pure Appl Geophys* 173:1265–1278
- Al-Garni MA (2015) Interpretation of magnetic anomalies due to dipping dikes using neural network inversion. *Arab J Geosci* 8:8721–8729
- Aydin A, Gelişli K (1996) Magnetic studies for the skarn zone of Sarihan-Bayburt. *Jeofizik* 10:41–49 (in Turkish with English Abstract)
- Baiyegunhi C, Gwavava O (2017) Magnetic investigation and 2 D gravity profile modelling across the Beattie magnetic anomaly in the southeastern Karoo Basin, South Africa. *Acta Geophys* 65:119–138
- Bean RJ (1966) A rapid graphical solution for the aeromagnetic anomaly of the two-dimensional tabular body. *Geophysics* 31:963–970
- Bhimasankaram VLS, Mohan NI, Seshagiri Rao SV (1978) Interpretation of magnetic anomalies of dikes using Fourier transforms. *Geophys Explor* 16:259–266
- Cassano E, Rocca F (1974) Interpretation of magnetic anomalies using spectral estimation techniques. *Geophys Prospect* 23:663–681
- Cooper GRJ (2012) The semi-automatic interpretation of magnetic dyke anomalies. *Comput Geosci* 44:95–99
- Cooper GRJ (2015) Using the analytic signal amplitude to determine the location and depth of thin dikes from magnetic data. *Geophysics* 80:1–6
- Dondurur D, Pamukçu OA (2003) Interpretation of magnetic anomalies from dipping dike model using inverse solution, power spectrum and Hilbert transform methods. *J Balk Geophys Soc* 6:127–139
- Essa KS, Elhussein M (2017) A new approach for the interpretation of magnetic data by a 2-D dipping dike. *J Appl Geophys* 136:431–443
- Essa KS, Nady AG, Mostafa MS, Elhussein M (2018) Implementation of potential field data to depict the structural lineaments of the Sinai Peninsula, Egypt. *J Afr Earth Sci* 147:43–53
- Gadirov VG, Eppelbaum LV, Kuderavets RS, Menshove OI, Gadirov KV (2018) Indicative features of local magnetic anomalies from hydrocarbon deposits: examples from Azerbaijan and Ukraine. *Acta Geophys*. <https://doi.org/10.1007/s11600-018-0224-0>
- Gay SP (1963) Standard curves for interpretation of magnetic anomalies over long tabular bodies. *Geophysics* 28:161–200
- Hood P (1964) The Königsberger ratio and the dipping-dyke equation. *Geophys Prospect* 12:440–456
- Keating PB, Pilkington M (1990) An automated method for the interpretation of magnetic vertical-gradient anomalies. *Geophysics* 55:336–343
- Koulomzine T, Lamontagne Y, Nadeau A (1970) New methods for the direct interpretation of magnetic anomalies caused by inclined dikes of infinite length. *Geophysics* 35:812–830
- McGrath PH, Hood PJ (1970) The dipping dike case, a computer curve matching method of magnetic interpretation. *Geophysics* 35:831–848
- Mehanee SA (2015) Tracing of paleo-shear zones by self-potential data inversion: case studies from the KTB, Rittsteig, and Grossensees graphite-bearing fault planes. *Earth Planets Space* 67(1):14
- Mehanee SA, Essa KS (2015) 2.5D regularized inversion for the interpretation of residual gravity data by a dipping thin sheet: numerical examples and case studies with an insight on sensitivity and non-uniqueness. *Earth Planets Space* 67(1):130
- Pal PC (1985) A gradient analysis based simplified inversion strategy for the magnetic anomaly of an inclined and infinite thick dike. *Geophysics* 50:1179–1182
- Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) *Numerical recipes, the art of scientific computing*. Cambridge University Press, Cambridge
- Rao BSR, Prakasa Rao TKS, Gopala Rao D, Kesavamani M (1972) Derivatives and dike anomaly interpretation. *Pure Appl Geophys* 99:120–129
- Rao DA, Babu HVR, Narayan PVS (1981) Interpretation of magnetic anomalies due to dikes: the complex gradient method. *Geophysics* 46:1572–1578
- Sengupta S, Das SK (1975) Interpretation of magnetic anomalies of dikes by Fourier transforms. *Pure Appl Geophys* 113:625–633
- Smith RS, Salem A, Lemieux J (2005) An enhanced method for source parameter imaging of magnetic data collected for mineral exploration. *Geophys Prospect* 53:655–665
- Werner S (1953) Interpretation of magnetic anomalies of sheet-like bodies. *Sveriges Geologiska Undersökning. Ser C Årsbok* 43(6):49



Edge-based finite-element modeling of 3D frequency-domain electromagnetic data in general dispersive medium

Mingxin Yue^{1,2} · Xiaoping Wu¹

Received: 16 April 2018 / Accepted: 30 January 2019 / Published online: 8 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

The geophysical electromagnetic (EM) theories are commonly based on the assumption that the conductivity of underground media is frequency independent. However, due to the existence of induced polarization (IP) effect, many earth materials are dispersive, and their electrical conductivity varies significantly with frequency. Therefore, the conventional numerical techniques are not proper for EM forward modeling in general dispersive medium. We present a new algorithm for modeling three-dimensional (3D) EM data containing IP phenomena in frequency domain by using an edge-based finite element algorithm. In this research, we describe the dispersion behavior of earth media by using a Cole–Cole complex conductivity model. Our algorithm not only models land and airborne EM surveys but also provides more flexibility in describing the surface topography with irregular hexahedral grids. We have validated the developed algorithm using an analytic solution over a half-space model with and without IP effect. The capabilities of our code were demonstrated by modeling coupled EM induction and IP responses in controlled-source audio magnetotelluric (CSAMT) and airborne electromagnetic (AEM) examples. This algorithm will have important guiding significance for survey planning in the dispersive areas, and it could be taken as a forward solver for practical 3D inversion incorporated IP parameters.

Keywords Frequency-domain electromagnetic method · Edge-based finite element · Dispersion · Induced polarization effect

Introduction

The frequency domain electromagnetic (EM) method has the advantages of large exploration depth, high resolution and strong anti-interference ability. It has been widely used for various geophysical fields such as geologic mapping (Sengpiel 1983), the detection of hydrocarbon (Constable 2010), mineral resource exploration (Vallée et al. 2011) as well as environmental and engineering investigations (Delefortrie et al. 2014; Scheuermann 2016). In the practical geophysical

electromagnetic applications, we usually consider the physical properties of the earth materials are frequency independent. However, the conductivity of earth materials varies significantly with frequency, and it is manifested by induced polarization (IP) effect in geophysical EM survey (Ward et al. 1988).

IP effect is an electromagnetic phenomenon; it can be characterized by a frequency-dependent complex conductivity in frequency domain. Many parametric models have been developed to describe the nature of this phenomenon, where the Cole–Cole relaxation model is the simplest and commonly used one (Cai et al. 2017a, b). Based on the a large number of measurements of rock and ore samples, Pelton et al. (1978) pointed out that the dispersive resistivity can be expressed as follows:

$$\sigma(\omega) = \sigma_0 \left(1 - \eta \left(1 - \frac{1}{1 + (i\omega\tau)^c} \right) \right)^{-1} \quad (1)$$

In this equation, σ_0 is the electrical conductivity at zero-frequency (DC), τ is a characteristic constant, η is chargeability, and c is frequency dependent ranging from

✉ Mingxin Yue
staryue@mail.ustc.edu.cn

Xiaoping Wu
wpx@ustc.edu.cn

¹ Laboratory of Seismology and Physics of Earth's Interior, School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, China

² Key Laboratory of Exploration Technologies for Oil and Gas Resources (Yangtze University), Ministry of Education, Wuhan 430100, China

zero to one; when $c = 1$, it is the popular Debye model. However, this phenomenon is not usually included in traditional EM numerical algorithm. Hence, numerical analysis of dispersive materials with suitable methods is of considerable importance, especially in the 3D geological environment. Recently, Marchant et al. (2014) used an implicit time-stepping approach to analyze IP effects in time-domain electromagnetic responses and found negative values appear in time-domain EM data in the presence of chargeable materials. Then, Cai et al. (2017a, b) extended this approach using finite element method to make it more computationally efficient. Furthermore, Commer et al. (2017) modeled dispersive TEM fields using an explicit time-stepping finite difference time-domain (FDTD) algorithm. Most work focused on the effect of dispersive physical properties is in time-domain electromagnetic data. Although studies relating to inverting IP data to obtain complex resistivity information have been published (Commer et al. 2011; Xu and Zhdanov 2015), the IP effects in frequency domain EM data in a complex 3D dispersive environment have not been discussed in the literature. In this case, the interpretation of frequency-domain EM data in the field urgently needs a forward solver that can deal with EM induction and IP effect simultaneously.

There are several techniques for 3D frequency-domain EM numerical simulations, such as integral equation method (IE) (e.g., Raiche 1974; Xiong and Tripp 1997), finite difference method (FD) (e.g., Wang and Hohmann 1993; Newman and Alumbaugh 1995; Groot-Hedlin 2006), and finite element method (FE) (e.g., Hu et al. 2015; Chung et al. 2015; Ansari and Farquharson 2014). IE method works effectively for the models with compact bodies in a layered background. However, the computational demand is rigorous when handling those complex models with large conductivity contrast (Zhdanov et al. 2006). FD method is one of the most commonly used approaches in 3D EM numerical simulations. However, it encounters some problems when the topography and complex structures appear in the models (Weiss and Newman 2002; Sasaki and Nakazato 2003; Mittet 2010). Compared with the other two methods, the FEM which uses the unstructured mesh is more suitable for 3D EM modeling in a complex model geometry (Grayver and Kolev 2015; Soloveichik et al. 2018). Recently, FE method has been used extensively to model 3D EM problems. However, the use of traditional nodal basis functions to vector problem cannot satisfy the condition that the electromagnetic field is continuous on the physical boundary and usually results in the appearance of spurious solutions (Jin 2015). Recently, a new vector basis function solves these problems by assigning degrees of freedom to be determined on the edges of the discrete element and only requiring the electromagnetic field to be tangential continuous. At the

same time, the basis function of edge-based element satisfies the condition of zero divergence automatically (Cai et al. 2014, 2017a, b). However, up to now, few have taken the IP effects into account, 3D forward modeling for fully coupled EM-IP problem is in urgently need.

In this study, we use our newly developed edge-based finite element (EDFE) forward modeling code to analyses and discuss IP effects in land and airborne electromagnetic survey in detail. The outline is as follows. In “Methodology” section, we adopt edge-based FE method to solve EM problems in a 3D dispersive medium. In this algorithm, the electrical conductivity is modeled by Cole–Cole relaxation model. To make our code more practical, irregular hexahedral grids were used. In “Numerical examples” section, we first validate the accuracy and efficiency of our 3D EDFE method. Further, we use two synthetic examples of controlled-source audio magnetotelluric (CSAMT) and airborne electromagnetic (AEM) to demonstrate the effect of dispersive materials and different Cole–Cole parameters on EM responses. Finally, we present a model results for AEM system over 3D rugged topography.

Methodology

Governing equations

In low-frequency electromagnetic applications, we can ignored the displacement currents, assuming the temporal dependence is $e^{-i\omega t}$, and the diffusive forms of electromagnetic field are described as (Zhdanov 2009):

$$\nabla \times \vec{E} = i\omega\mu\vec{H} \quad (2a)$$

$$\nabla \times \vec{H} = \sigma\vec{E} + \vec{j} \quad (2b)$$

where \vec{E} and \vec{H} represent the electric and magnetic field, respectively. $\omega = 2\pi f$, is the angular frequency; \vec{j} is the extraneous electric or magnetic source; μ represents magnetic permeability; σ is the conductivity which is the inverse of the resistivity (ρ), and it is represented in the form of complex conductivity by Cole–Cole model in this study.

Since the conductivity in the air is close to zero, when solving the electromagnetic field using the above two formulas, the first term $\sigma\vec{E}$ at the right end of Eq. (2b) is relatively small and then Eq. (2a) and Eq. (2b) will form weak coupling, resulting in the equation solution not converging. Therefore, the rotations of the upper equations are converted into second-order non-homogeneous Helmholtz equations:

$$\nabla \times \nabla \times \vec{E} = i\omega\mu\sigma\vec{E} + i\omega\mu\vec{j} \quad (3)$$

To solve the problem of field singularities and avoid excessive meshing near the source, we divided the electromagnetic

Fig. 1 An illustration of an edge element. **a** The $\xi\eta\zeta$ -coordinate system. **b** The xyz -coordinate system

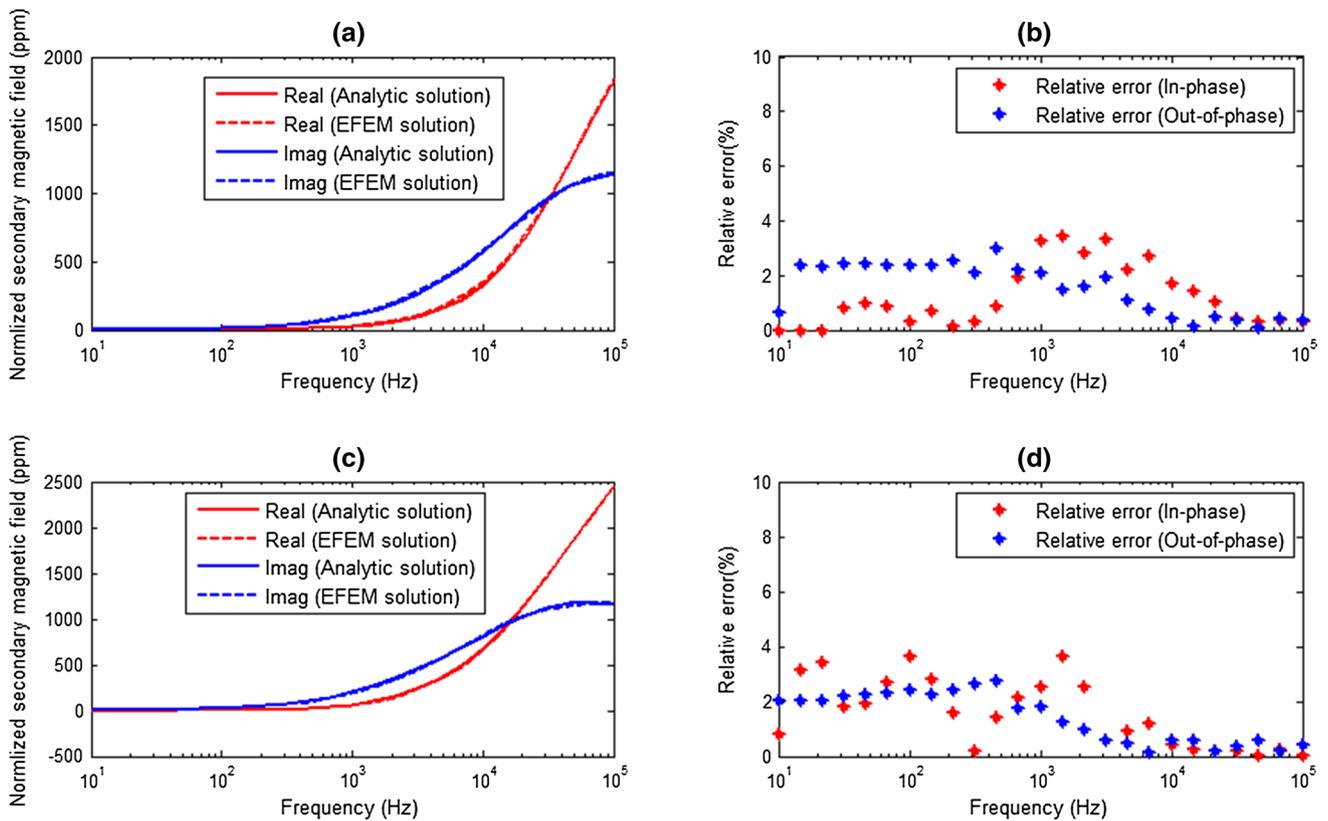
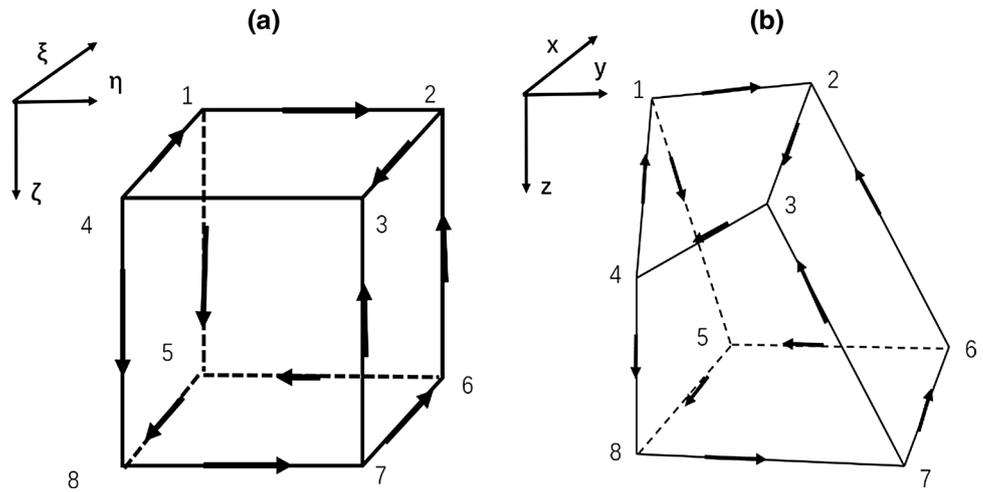


Fig. 2 Comparison of the normalized secondary vertical magnetic field from 3D numerical results and analytic solutions (Kim et al. 2011). **a, b** Responses of a non-dispersive model with a conductivity of 0.01 S/m and its relative error. **c, d** Responses of a dispersive model and its relative error

field is into the background field and the scattering field. The equation for the anomalous electric field is described as follows:

$$\nabla \times \nabla \times \vec{E}_s = i\omega\mu\sigma\vec{E}_s + i\omega\mu\hat{\sigma}\vec{E}_b \tag{4}$$

ity of 0.01 S/m and its relative error. **c, d** Responses of a dispersive model and its relative error

In Eq. (4), subscript p and s stand for the background field and the scattering field. $\hat{\sigma} = \sigma - \sigma_0$ is the anomalous conductivity. In the above formula, the background field \vec{E}_b needs to be pre-calculated over a 1D layered earth (Ward et al. 1988). Here, the above differential equation was solved using the first type of boundary conditions (Dirichlet boundary condition).

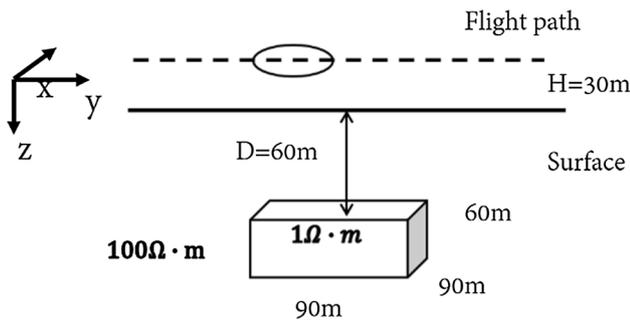


Fig. 3 A simple 3D model for verifying the accuracy of our proposed method

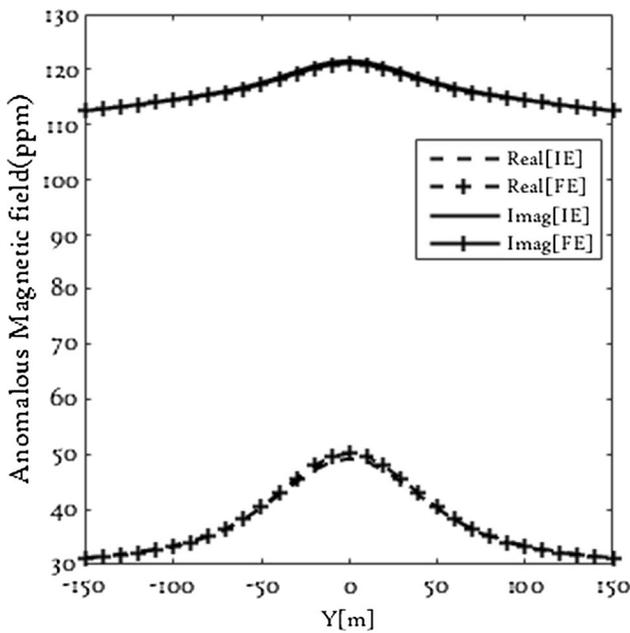


Fig. 4 Comparison of vertical anomalous magnetic field from 3D EDFE and IE solutions

Once the scattering electric field is calculated from Eq. (4), we can get the scattering magnetic field easily based on following formulation:

$$\vec{H}_s = -\frac{1}{i\omega\mu} \nabla \times \vec{E}_s \tag{5}$$

The total field becomes the sum of the background field and the scattering field.

Finite element approximation and analysis

We use EDFE method to solve the above partial differential equation Eq. (4). EDFE method assigns degrees that can satisfy the tangential continuity of the electromagnetic field and eliminate no-physical solutions (Noh et al. 2016). To make the

codes more capable of modeling arbitrary three-dimensional geometries, irregular hexahedral elements were used in this work (Grayver and Kolev 2015).

As shown in Fig. 1, we define the center coordinates of element as (x_0^e, y_0^e, z_0^e) , and h_x^e, h_y^e, h_z^e are the element lengths in three direction along the Cartesian coordinate system, respectively (Jin 2015). We adopted vector basis function $\vec{\varphi}_k^e$, the E-field components inside the element is written as follows:

$$\vec{E}^e = \sum_{k=1}^{12} \vec{\varphi}_k^e E_k^e \tag{6}$$

By Coordinate conversion, the edge-based element basis functions in $\xi\eta\zeta$ -coordinate system can be written as:

$$\vec{\varphi}_k^e = \frac{h_k^e}{8} (1 + \zeta\zeta_k) (1 + \eta\eta_k) \nabla\xi, \quad k = 1, 2, 3, 4 \tag{7a}$$

$$\vec{\varphi}_k^e = \frac{h_k^e}{8} (1 + \zeta\zeta_k) (1 + \xi\xi_k) \nabla\eta, \quad k = 5, 6, 7, 9 \tag{7b}$$

$$\vec{\varphi}_k^e = \frac{h_k^e}{8} (1 + \eta\eta_k) (1 + \xi\xi_k) \nabla\zeta, \quad k = 10, 11, 12, 13 \tag{7c}$$

Based on Eq. (6) and Eq. (4), we apply weighted residual approach, and the weak form of the secondary E-field Helmholtz equation is described as:

$$F(\vec{E}_s^e) = \int \vec{\varphi}_k^e \left[\nabla \times \nabla \times \vec{E}_s - i\omega\mu\sigma\vec{E}_s - i\omega\mu\Delta\sigma\vec{E}_p \right] d\mathbf{v} \tag{8}$$

The following discretized form of Eq. (8) is derived:

$$F(E_s)_k = \sum_{e=1}^{Ne} \left[A^e E_s^e - i\omega\mu\sigma_e B^e E_s^e - i\omega\mu\Delta\sigma_e B^e E_p^e \right] \tag{9}$$

where A^e and B^e are element stiffness matrices

$$A_{kj}^e = \int (\nabla \times \vec{\varphi}_k^e) \cdot (\nabla \times \vec{\varphi}_j^e) d\mathbf{v} \tag{10}$$

$$B_{kj}^e = \int \vec{\varphi}_k^e \cdot \vec{\varphi}_j^e d\mathbf{v} \tag{11}$$

After finite analysis, a large sparse equation system can be obtained:

$$\sum_{e=1}^N \left\{ \sum_{k=1}^{12} \left(\sum_{j=1}^{12} K_{kj}^e E_{sj}^e - P_k^e \right) \right\} = 0 \tag{12}$$

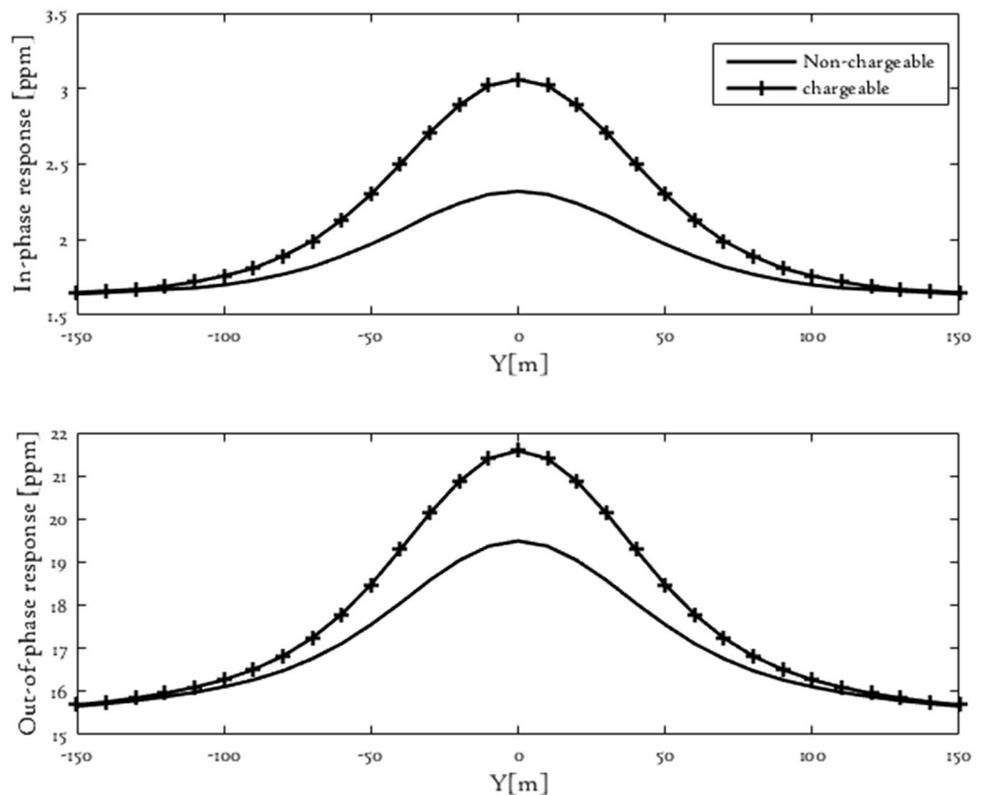
where

$$K_{kj}^e = A_{kj}^e - i\omega\mu\sigma B_{kj}^e \tag{13}$$

$$P_i^e = i\omega\mu\Delta\sigma \int E_p \vec{\varphi}_k^e d\mathbf{v} \tag{14}$$

There are two ways to solve this large, sparse system of equations, such as iterative solver and direct solver (da Silva et al. 2012). The geophysical EM simulations are usually

Fig. 5 Normalized components of anomalous magnetic field responses of an anomalous body in a resistive half-space for a 100-Hz source



multi-source and multi-frequency problems. Direct solvers have great advantages for such problems (da Silva et al. 2012; Bo et al. 2015). In this study, a parallel direct solver called PARDISO was used to solve this large equation system (Schenk and Gärtner 2004; Kumbhar et al. 2011).

Numerical examples

Verification of accuracy and efficiency

We first validate our 3D EDFE code using analytic solutions over a half-space model. The configuration is a horizontal coplanar (HCP) loop with an 8-m offset, and the vertical magnetic dipole set at 30 m above the earth surface. The calculation domain is set to $2 \text{ km} \times 2 \text{ km} \times 1 \text{ km}$ with a 5 km extension along x , y , z directions, and totally 80,000 hexahedron elements were discretized. We did two calculations: a non-dispersive model with the resistivity of $100 \Omega \text{ m}$ and a dispersive model with the IP parameters with $c = 0.5$, $\eta = 0.5$, $\tau = 0.1 \text{ s}$ and $\sigma_0 = 0.01 \text{ S/m}$. The vertical secondary magnetic field was calculated at 25 frequencies from 10^1 Hz to 10^5 Hz which distributed in logarithmic scale uniformly. The total run time of this model is 18s on a personal computer with an i7 CPU at 2.9 GHz and 16 G RAM.

As shown in Fig. 2a, c, the vertical anomalous magnetic field of our 3D numerical results and analytic solutions

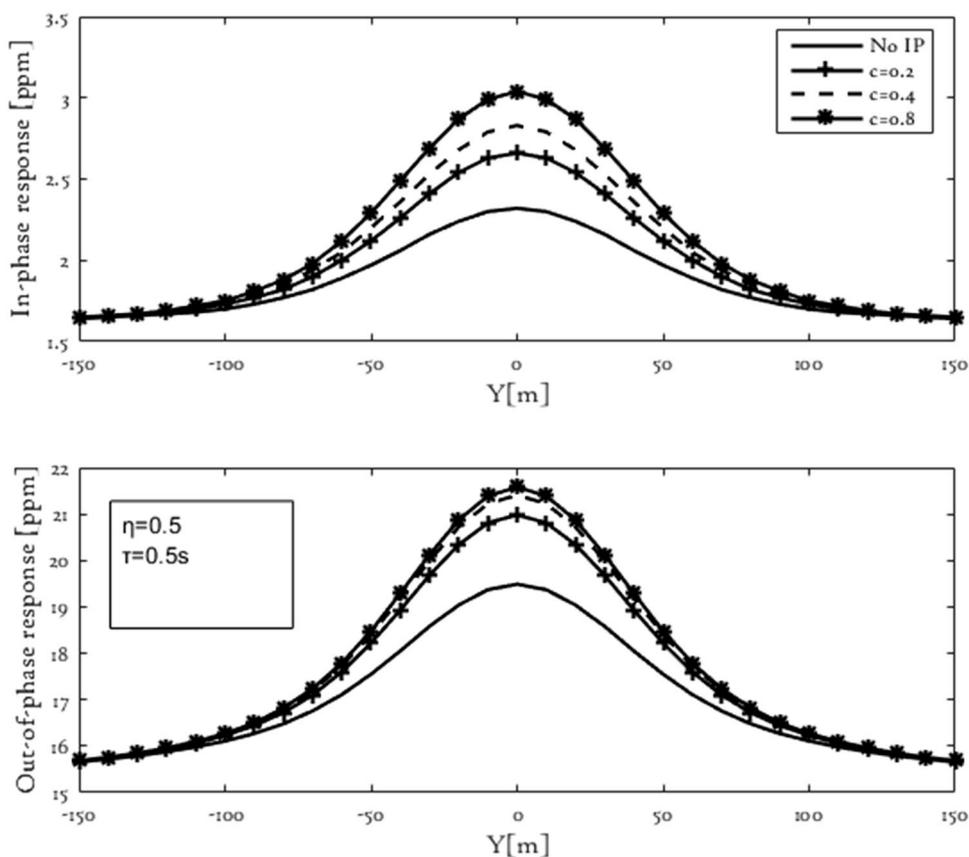
showed high agreement over both non-dispersive and dispersive model. The relative error value is less than 4% (Fig. 2b, d).

The accuracy of our algorithm was also verified by the integral equation solutions (Xiong and Tripp 1997) for a fully 3D problem. The EDFE and IE solutions of 3-D models are described in Fig. 3. The transmitter and receiver configuration is a horizontal loop–loop system with an 8-m offset, and the height of this system is 30 m above the surface. The calculate frequency is 925 Hz. The results of our proposed EDFE method and integral equation approach for the vertical secondary magnetic field responses are shown in Fig. 4. It is in high agreement with IE solutions.

IP effect in airborne EM data

To well understand the IP effect in Airborne EM data, a single chargeable block is buried in a uniform background which is the same as in the previous example (shown in Fig. 3). We first consider the anomalous body exhibits Debye dispersion with $\sigma_0 = 1 \text{ S/m}$, $\tau = 0.5 \text{ s}$, and $\eta = 0.5$. Two forward modeling were run: the responses of the model with a non-chargeable anomalous body, and a chargeable anomalous body, respectively. Figure 5 presents the secondary magnetic field responses from the two simulations. Due to the IP effect, the responses of the model are significantly increased. The results show the varied maximum value of

Fig. 6 Normalized components of anomalous secondary magnetic field responses for different c



50% in in-phase component and the varied maximum value of 30% in out-of-phase component.

Then, we study a general chargeable model with different c and η values. We calculated the responses along the y -direction direction profile for a 100-Hz source. Figure 6 reveals the AEM responses become larger with the frequency dependence c increases. It shows that their varied maximum value of 30% at $c=0.8$ in in-phase component and varied maximum value of 20 per cent at $c=0.8$ in out-of-phase component. Figure 7 reveals the AEM responses increase with the increase in chargeability. The results show that their varied maximum value of 30% at $\eta=0.5$ in in-phase component and varied maximum value of 10 per cent at $\eta=0.1$ in out-of-phase component. Even if the Cole–Cole model parameter values is quite small in this model study, but the IP effect is still significant in AEM responses.

Figure 8 presents the responses of previous 3D model in dispersive medium. We can obviously see that AEM responses are significantly disturbed by the IP effect if the medium is dispersive. It shows the responses will become larger and an anomaly in the dispersive medium becomes a wider distance interval than in a non-dispersive medium. So, if the medium is dispersive and not modeled properly, the information of anomaly will be masked.

IP effect in land EM data

To study the IP effect in land EM survey, a 3D CSAMT modeling is carried out. The geometry of this model is shown in Fig. 9. In this test, we use an y -oriented electric dipole source, and the length of the source is 1km. In this model, a conductive anomalous body and a resistive anomalous body are buried in a half space background. The measurement range is from -1000 m to 1000 m with a constant 100 m spacing in the y -direction at $x=0$, total 21 sounding points.

We ran four numerical tests, and the dispersive properties of the physical unit are shown in Table 1. Three frequencies are 10 Hz, 100 Hz and 1000 Hz. We plot the amplitude and phase data component of inline electric field in Fig. 10. The conductive body shows a significantly anomaly along the y -direction. However, the resistive body shows little change along this profile. It shows the E_y responses are more sensitive to the conductive anomaly than to the resistive anomaly, and the response of the phase data in the horizontal direction is higher than amplitude data. In CSAMT responses, anomalous bodies with IP effect have little impact on the results, because compared to the background, its size is too small. However, we can clearly see that the background with IP effect has a huge impact on CSAMT responses.

Fig. 7 Normalized components of anomalous magnetic field responses for different η

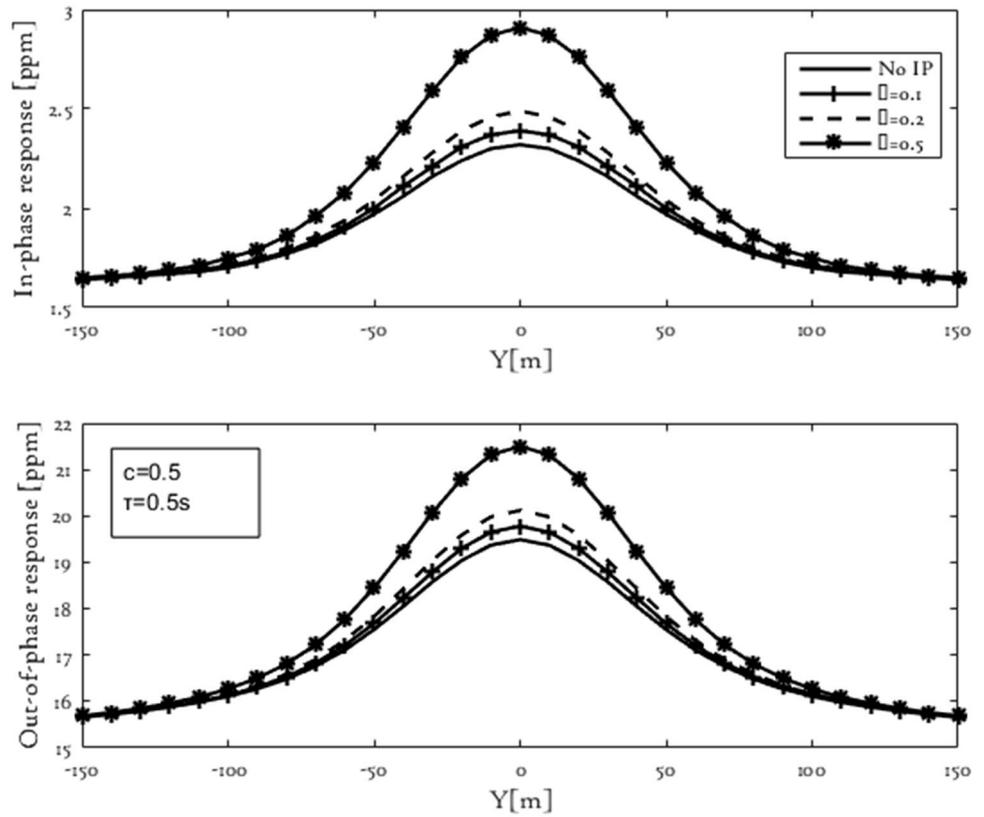
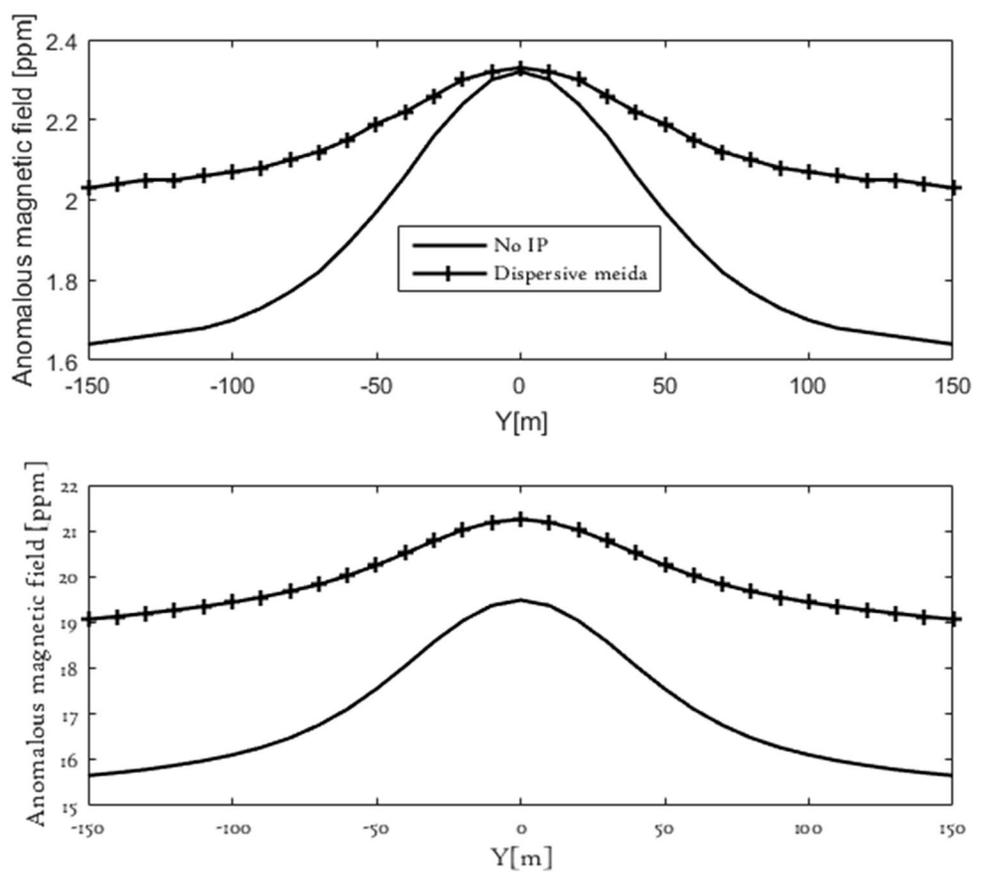


Fig. 8 The in-phase (top) and out-of-phase (bottom) responses of model shown in Fig. 3. The black solid line denotes responses of a non-dispersive medium, and the arrow line denotes the responses of a dispersive medium with Cole–Cole parameter of $\sigma_0=0.01$ S/m, $\eta=0.2$, $\tau=0.5$, and $c=0.5$



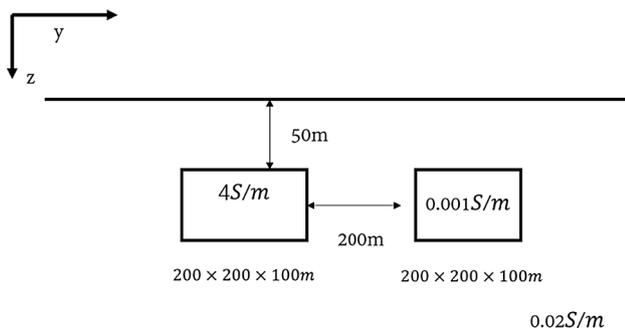


Fig. 9 The cross section of the CSAMT model

Table 1 Classification of attributes in CSAMT model. IP parameters: $\eta = 0.5$, $\tau = 0.2$, $c = 1$

Test	Resistive anomaly	Conductive anomaly	Background
T01	Non-chargeable	Non-chargeable	Non-chargeable
T02	Chargeable	Non-chargeable	Non-chargeable
T03	Non-chargeable	Chargeable	Non-chargeable
T04	Non-chargeable	Non-chargeable	Chargeable

Next, to qualitatively categorize effects of IP parameters on CSAMT responses, different combinations of τ and c in the test 4 (T04) were calculated. The transmitter frequency is 100 Hz. Figure 11 shows the amplitude responses of inline

electric field. Figure 11a displays the amplitude variations with frequency dependence c , and it shows the amplitude decreases with increasing frequency exponent c ; Fig. 11b displays the amplitude and phase variations with time constant τ and chargeability η . It shows the amplitude values will decrease with increasing chargeability and the results are not changed significantly when time constant τ changes.

Analysis of IP effect in a topographic model

It has been demonstrated by Zhang et al. (2016) that the topography gives an important response on AEM data. To illustrate capabilities of our code for simulating models with terrain, we design a model with two hills and a valley as shown in Fig. 12a, b. The hills have an altitude of 70 m and 80 m, while the valley is 40 m below the sea level. A conductive anomalous body with a DC conductivity of 0.2S/m is embedded in the background with a DC conductivity of 0.01S/m. The size of plate is 400 × 20 × 70 m. The height of AEM system is always 30 m above the ground over the whole flight. The HCP offset is set to 8 m.

We ran four simulations: the electromagnetic response of the uniform half space, the model in presence of a conductive anomaly, the model without conductive anomaly, and the response of the model with a non-chargeable conductive anomalous body in the chargeable half space with IP parameters are of $\eta = 0.2$, $\tau = 0.2$ s, and $c = 1$. To perform a more quantitative analysis of IP effect and topographic effect in

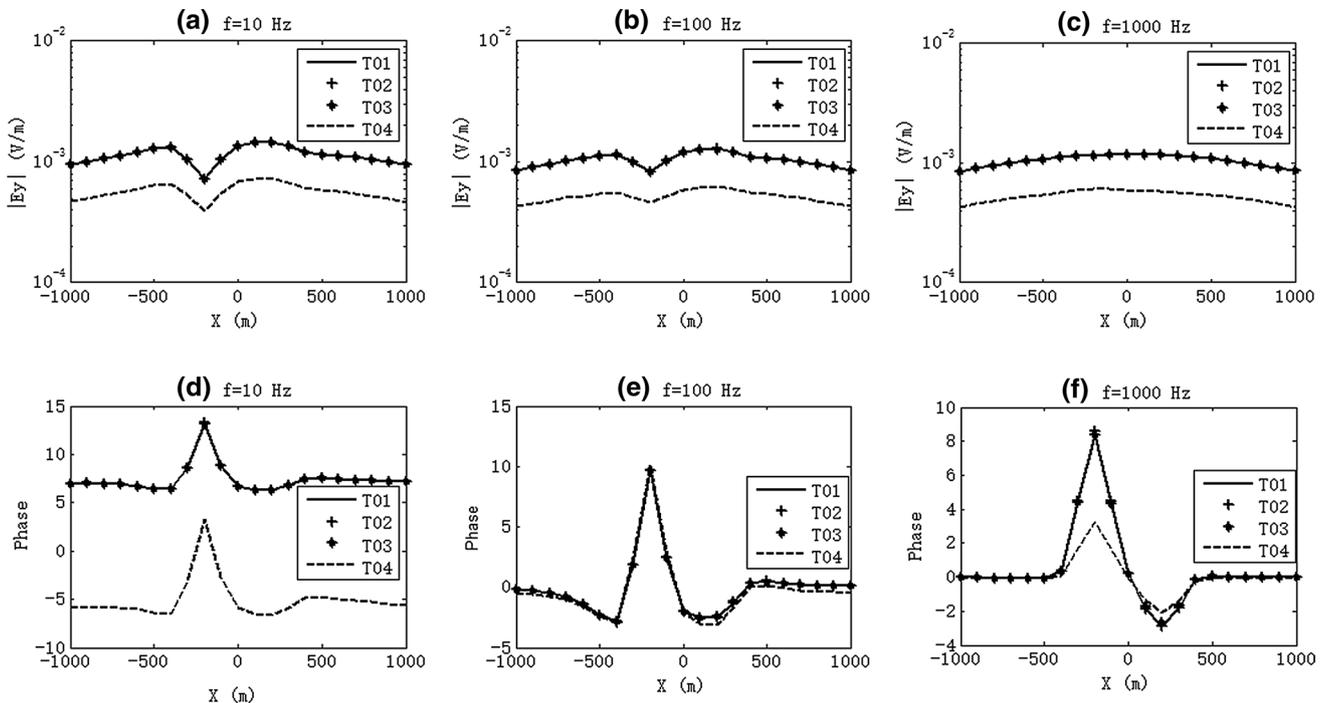


Fig. 10 Responses of the CSAMT model. a–c Amplitude; e, f phase

Fig. 11 Responses for the CSAMT model

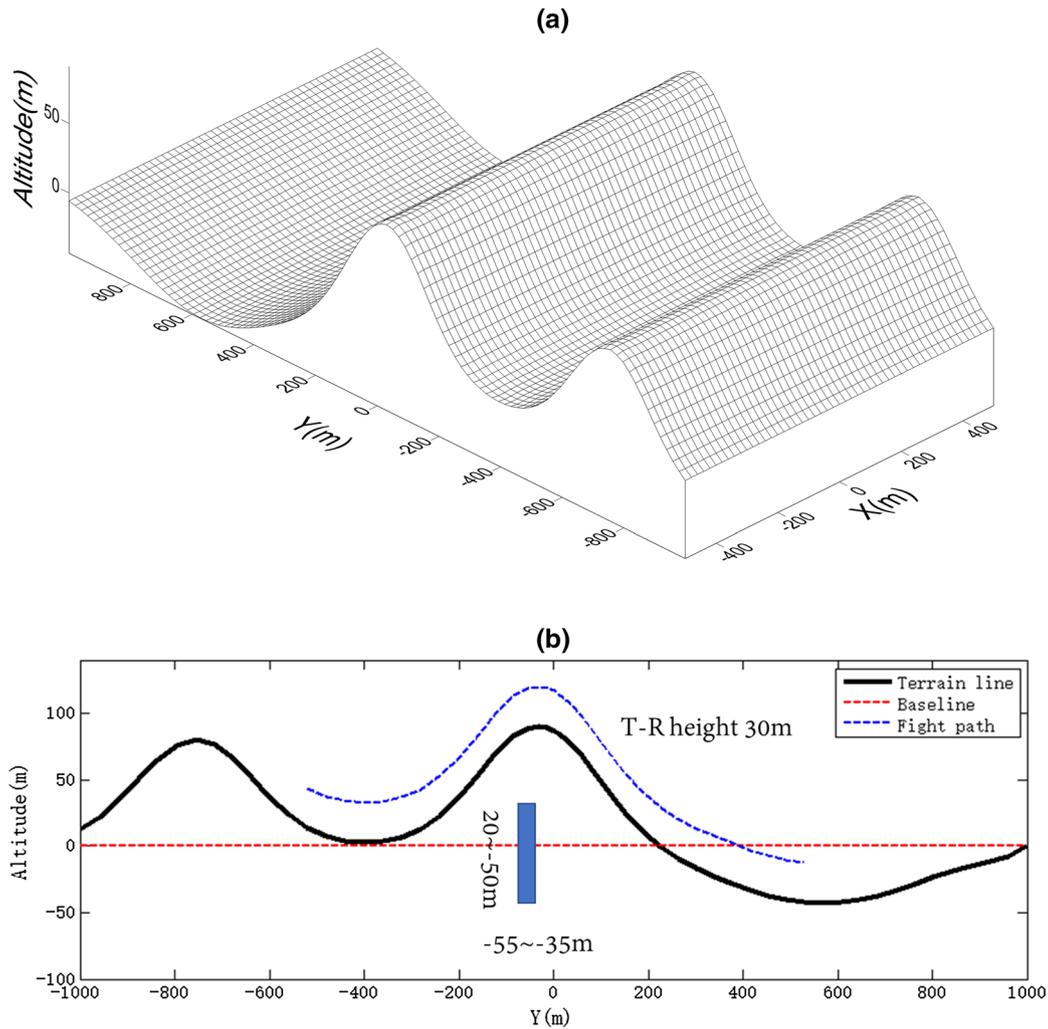
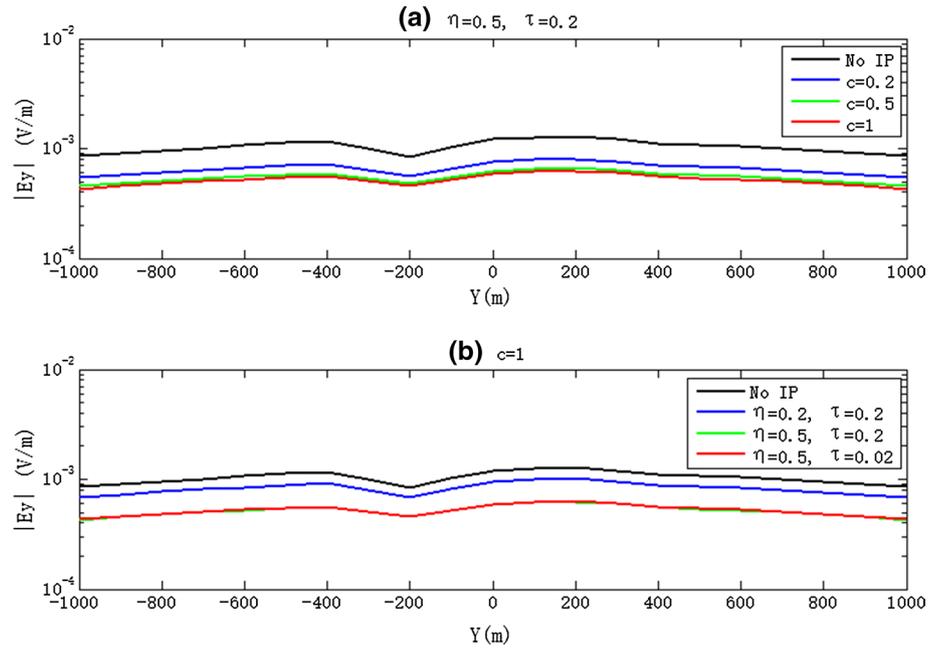


Fig. 12 AEM system over a topographic model. **a** 3D view of this model. **b** 2D view in XZ-plane. Blue rectangle box represents the range of the conductor

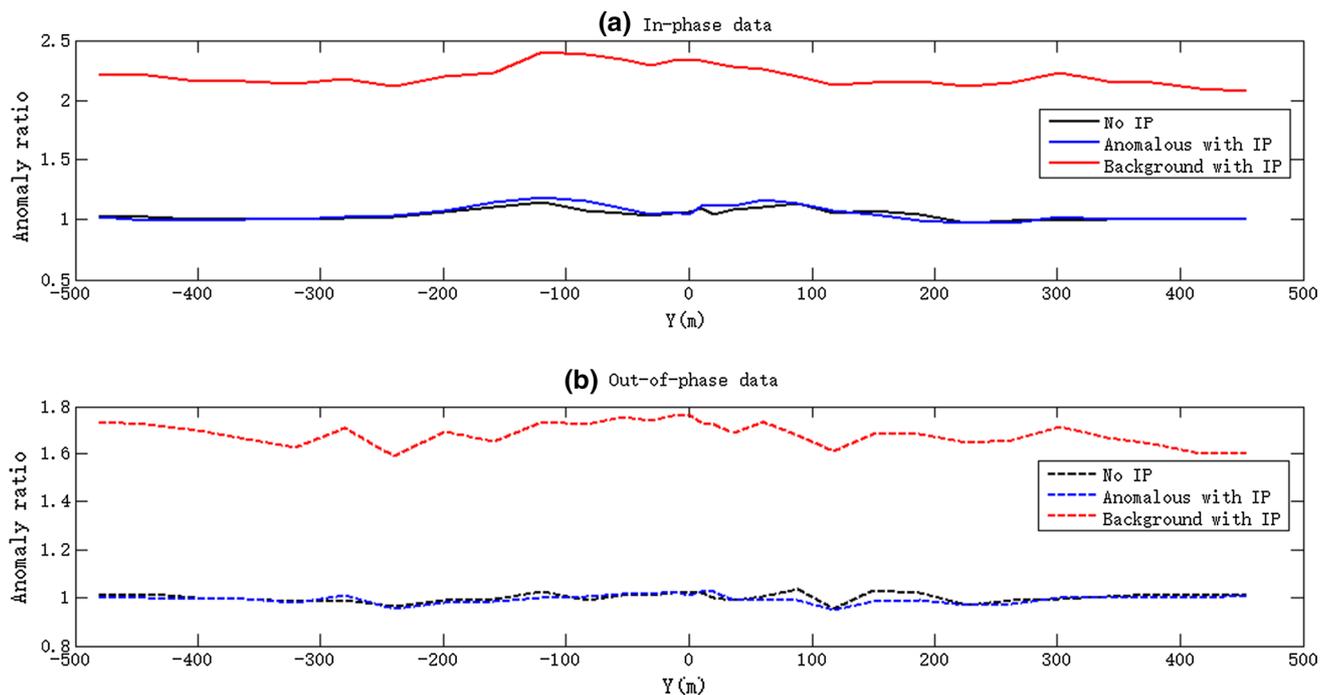


Fig. 13 Anomaly ratio of the topographic model as shown in Fig. 12

AEM data, the corresponding ratio with or without anomalous body is calculated. Figure 13 shows the anomaly ratio of vertical magnetic field response, we can see the topography affects EM data a lot, it may result in wrong results if not considering topography in interpretation of data. It is not very easy to recognize conductive anomaly in the presence of topography, whether or not it is dispersive. The AEM response was generally dominated by the IP effect if the background medium is dispersive, and it is very difficult to detect basement conductor buried in chargeable medium if IP effect is not modeled.

Conclusions

We have presented a frequency-domain electromagnetic forward modeling solver in 3D general dispersive medium based on an edge finite-element approach. The irregular hexahedral meshes used in this paper can describe surface topography properly. Using Cole–Cole conductivity model, we can fully simulate EM-IP responses. This method can be used to model not only land EM but also airborne EM data. From the results of AEM and CSAMT examples, we can clearly see that in presence of chargeable materials the standard EM modeling (EM induction-only) yields inaccurate results if the IP effect is not modeled. Our approach could be very useful in survey plan in the presence of

chargeable materials. It is worth mentioning that our method can become a useful forward solver for extracting IP parameters from EM data.

Acknowledgements The author acknowledges two anonymous reviewers for their useful comments and suggestions. This work was supported by The National Key R&D Program of China (2018YFC0603500), Natural Science Foundation of China (41874084, 41674076), and China Scholarship Council. Funding was provided by Open Fund of Key Laboratory of Exploration Technologies for Oil and Gas Resources (Yangtze University), Ministry of Education (Grant No. K2016-08).

References

- Ansari S, Farquharson CG (2014) 3D finite-element forward modeling of electromagnetic data using vector and scalar potentials and unstructured grids. *Geophysics* 79(4):E149–E165
- Bo H, Xiang-Yun HU, Huang YF, Peng RH, Jian-Hui LI, Cai JC (2015) 3-d frequency-domain csem modeling using a parallel direct solver. *Chin J Geophys (in Chinese)* 58(8):2812–2826
- Cai H, Xiong B, Han M, Zhdanov M (2014) 3D controlled-source electromagnetic modeling in anisotropic medium using edge-based finite element method. *Comput Geosci* 73:164–176
- Cai H, Hu X, Li J, Endo M, Xiong B (2017a) Parallelized 3D CSEM modeling using edge-based finite element with total field formulation and unstructured mesh. *Comput Geosci* 99:125–134
- Cai H, Hu X, Xiong B, Zhdanov MS (2017b) Finite-element time-domain modeling of electromagnetic data in general dispersive medium using adaptive Padé series. *Comput Geosci* 109:194–205

- Chung Y, Son J, Lee TJ, Kim HJ, Shin C (2015) Three-dimensional modelling of controlled-source electromagnetic surveys using an edge finite-element method with a direct solver. *Geophys Prospect* 62(6):1468–1483
- Commer M, Newman GA, Williams KH, Hubbard SS (2011) 3D induced-polarization data inversion for complex resistivity. *Geophysics* 76(3):F157–F171
- Commer M, Petrov PV, Newman GA (2017) FDTD modelling of induced polarization phenomena in transient electromagnetics. *Geophys J Int* 209(1):387–405
- Constable S (2010) Ten years of marine CSEM for hydrocarbon exploration. *Geophysics* 75(5):75A67–75A81
- da Silva NV, Morgan JV, MacGregor L, Warner M (2012) A finite element multifrontal method for 3D CSEM modeling in the frequency domain. *Geophysics* 77(2):E101–E115
- Delefortrie S, Saey T, Van De Vijver E, De Smedt P, Missiaen T, Demerre I, Van Meirvenne M (2014) Frequency domain electromagnetic induction survey in the intertidal zone: limitations of low-induction-number and depth of exploration. *J Appl Geophys* 100:14–22
- Grayver AV, Kolev TV (2015) Large-scale 3D geoelectromagnetic modeling using parallel adaptive high-order finite element method. *Geophysics* 80(6):E277–E291
- Groot-Hedlin CD (2006) Finite-difference modeling of magnetotelluric fields: error estimates for uniform and nonuniform grids. *Geophysics* 71(3):97
- Hu YC, Li TL, Fan CS, Wang DY, Li JP (2015) Three-dimensional tensor controlled-source electromagnetic modeling based on the vector finite-element method. *Appl Geophys* 12(1):35–46
- Jin JM (2015) *The finite element method in electromagnetics*. Wiley, Hoboken
- Kim HJ, Song Y, Lee KH (2011) High-frequency electromagnetic inversion for a dispersive layered earth. *J Geomagn Geoelectr* 49(11):1439–1450
- Kumbhar A, Chakravarthy K, Keshavamurthy R, Rao GV (2011) Utilization of parallel solver libraries to solve structural and fluid problems. *Intel Math Kernel Library—White Papers*
- Marchant D, Haber E, Oldenburg DW (2014) Three-dimensional modeling of IP effects in time-domain electromagnetic data. *Geophysics* 79(6):E303–E314
- Mittet R (2010) High-order finite-difference simulations of marine CSEM surveys using a correspondence principle for wave and diffusion fields FDTD simulation of marine CSEM surveys. *Geophysics* 75(1):F33–F50
- Newman GA, Alumbaugh DL (1995) Frequency-domain modelling of airborne electromagnetic responses using staggered finite differences. *Geophys Prospect* 43(8):1021–1042
- Noh K, Oh S, Seol SJ, Lee KH, Byun J (2016) Analysis of anomalous electrical conductivity and magnetic permeability effects using a frequency domain controlled-source electromagnetic method. *Geophys J Int* 204(3):1550–1564
- Pelton WH, Ward SH, Hallof PG, Sill WR, Nelson PH (1978) Mineral discrimination and removal of inductive coupling with multifrequency IP. *Geophysics* 43(3):588–609
- Raiche AP (1974) An integral equation approach to three-dimensional modelling. *Geophys J Int* 36(2):363–376
- Sasaki Y, Nakazato H (2003) Topographic effects in frequency-domain helicopter-borne electromagnetics. *Explor Geophys* 34(2):24–28
- Schenk O, Gärtner K (2004) Solving unsymmetric sparse systems of linear equations with PARDISO. *Future Gener Comput Syst* 20(3):475–487
- Scheuermann A (2016) Electric and electromagnetic measurement methods in civil and environmental engineering. *J Geophys Eng* 13(2):E1
- Sengpiel KP (1983) Resistivity/depth mapping with airborne electromagnetic survey data. *Geophysics* 48(2):181–196
- Soloveichik YG, Persova MG, Domnikov PA, Koshkina YI, Vagin DV (2018) Finite-element solution to multidimensional multisource electromagnetic problems in the frequency domain using non-conforming meshes. *Geophys J Int* 212(3):2159–2193
- Vallée MA, Smith RS, Keating P (2011) Metalliferous mining geophysics—state of the art after a decade in the new millennium. *Geophysics* 76(4):W31–W50
- Wang T, Hohmann GW (1993) A finite-difference time-stepping solution for three-dimensional electromagnetic modeling. *Geophysics* 58(1):1646
- Ward SH, Hohmann GW, Nabighian MN (1988) Electromagnetic theory for geophysical applications In: *Electromagnetic methods in applied geophysics* (vol 1, no. 3, pp 131–311)
- Weiss CJ, Newman GA (2002) Electromagnetic induction in a fully 3-D anisotropic earth. *Geophysics* 67(4):1104–1114
- Xiong Z, Tripp AC (1997) 3-D electromagnetic modeling for near-surface targets using integral equations. *Geophysics* 62(4):1097–1106
- Xu Z, Zhdanov MS (2015) Three-dimensional Cole–Cole model inversion of induced polarization data based on regularized conjugate gradient method. *IEEE Geosci Remote Sens Lett* 12(6):1180–1184
- Zhang B, Yin CC, Liu YH, Cai J (2016) 3D modeling on topographic effect for frequency-/time-domain airborne EM systems. *Chin J Geophys (in Chinese)* 59(4):1506–1520
- Zhdanov MS (2009) *Geophysical electromagnetic theory and methods*, vol 43. Elsevier, Amsterdam
- Zhdanov MS, Lee SK, Yoshioka K (2006) Integral equation method for 3D modeling of electromagnetic fields in complex structures with inhomogeneous background conductivity. *Geophysics* 71(6):G333–G345



Prediction of porosity and gas saturation for deep-buried sandstone reservoirs from seismic data using an improved rock-physics model

Yaneng Luo¹ · Handong Huang¹ · Morten Jakobsen² · Yadi Yang³ · Jinwei Zhang⁴ · Yanjie Cai⁵

Received: 5 December 2018 / Accepted: 23 February 2019 / Published online: 2 March 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

In recent years, many important discoveries have been made in global deep oil and gas exploration, which indicates that deep exploration has gradually become one of the most important areas in current and future hydrocarbon exploration. However, the prediction of deep reservoirs is very challenging due to their low porosity and complex pore structure characteristics caused by the burial depth and diagenesis. Rock physics provides a link between the geologic reservoir parameters and seismic elastic properties and has evolved to become a key tool of quantitative seismic interpretation. Based on the mineral component and pore structure analysis of studied rocks, we propose an improved rock-physics model by introducing a third feldspar-related pore for deep-buried sandstone reservoirs to the traditional Xu–White model. This modelling process consists of three steps: first, rock matrix modelling using time-average equations; second, dry rock modelling using a multi-pore analytical approximation; and third, fluid-saturated rock modelling using a patchy distribution. It has been used in total porosity estimation, S-wave velocity prediction and rock-physics template establishment. The applicability of the improved rock-physics model is verified by a theoretical quartz-water model test and a real data total porosity estimation compared with the traditional Xu–White model and the density method. Then, a rock-physics template is generated by the improved rock-physics model for porosity and gas saturation prediction using seismic data. This template is carefully calibrated and validated by well-log data at both the well-log scale and seismic scale. Finally, the feasibility of the established rock-physics template for porosity and gas saturation prediction is validated by a deep-buried sandstone reservoir application in the East China Sea.

Keywords Deep exploration · Gas-bearing sandstone · Pore structure · Rock-physics modelling · Seismic prediction

Introduction

With the development of deep exploration technologies in both the geological, geophysical theory and engineering, many giant oil and gas fields have been discovered in the world's deep petroliferous basins (Dyman et al. 2003; Dutton and Loucks 2010; Sun et al. 2013; Pang et al. 2015; Lai et al. 2017; Wang et al. 2018; Yuan et al. 2019a). However, the characterization of deep reservoirs is challenging due to the low porosity and complex pore structures caused by the burial depth and diagenesis (Zou et al. 2012; Rezaee et al. 2012; Lai et al. 2016). Rock physics is an effective tool to combine reservoir parameters (porosity, shale volume, fluid saturation, etc.) with seismic elastic parameters (P-wave velocity, S-wave velocity, density, etc.) (Avseth et al. 2005; Mavko et al. 2009; Best 2014). Depending on whether the combination is stepwise or unified, the application workflow may be grouped into two categories: one is

✉ Handong Huang
huanghandongcup@163.com

¹ State Key Laboratory of Petroleum Resource and Prospecting, Unconventional Natural Gas Research Institute, China University of Petroleum, Beijing 102249, China
² Department of Earth Science, University of Bergen, 5020 Bergen, Norway
³ Research Institute of Petroleum Exploration and Development, PetroChina Company Limited, Beijing 100083, China
⁴ College of Earth Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, Shandong, China
⁵ Gudao Oil Production Plant of Shengli Oilfield, China Petrochemical Corporation, Dongying 257231, Shandong, China

the sequential or cascaded approach (Buland et al. 2008), and the other is the joint or simultaneous approach (Bosch et al. 2009; Yuan et al. 2019b). Since we mainly focus on the establishment and verification of an improved rock-physics model, we adopt the first cascaded approach to obtain the rock-physics modelling results. Rock-physics models can be used to simulate and infer possible geological scenarios beyond available observation data. They play an important role in both the seismic forward modelling analysis and the quantitative interpretation of seismic inversion results (Avseth et al. 2005).

For high-porosity (larger than 20%) sandstones, combination of granular medium Hertz–Mindlin contact model and Hashin–Shtrikman bounds is often applied to calculate the dry rock moduli under different sorting and cementation trends (Dvorkin and Nur 1996; Ødegaard and Avseth 2004; Mavko et al. 2009; Ruiz and Dvorkin 2009; Avseth et al. 2014). For this type of modelling, the key parameters are the amount and localization of the cement within the grain structure and the coordination number of grains (the average number of contacts per sphere). Since the initial cementation effect will cause a large velocity increase with only a small decrease in porosity, this contact-cement model is often found to overpredict shear stiffnesses in cemented sandstones when the nonuniform contacts and heterogeneous stress are not taken into account (Bachrach and Avseth 2008; Avseth et al. 2010).

For low-porosity sandstones, the influence of pore structure on elastic parameters is more significant than the lithology variation or pore size (Berge et al. 1993; Johansen et al. 2002), which is the main reason for the complex relationship between the elastic parameters and reservoir parameters of deep sandstone reservoirs. The inclusion models (Mavko et al. 2009) can take all these microstructure factors into account and are usually adopted for this kind of modelling. Kuster and Toksöz (1974) derived expressions for the effective elastic moduli by using the long-wavelength first-order scattering theory on the assumption of the noninteraction between inclusions, which means the porosity must be much smaller than the pore aspect ratio. To eliminate this limitation, the differential effective medium (DEM) model calculates the effective elastic moduli by incrementally adding inclusions to the background rock matrix (Norris 1985; Zimmerman 1985; Berryman 1992). This allows the porosity to be higher than the aspect ratio, and the results are always consistent with the rigorous bounds. However, since the system of ordinary differential equations is coupled in the DEM model, the solution must be computed by a numerical iterative process, which will cause tremendous computation time in the real 3D seismic inversion applications. Moreover, for multiple pores, the DEM model is sensitive to the order in which the pores are added to the background rock matrix (Mavko et al. 2009). To overcome these deficiencies

of the DEM model, many approximate analytical expressions are made under different assumptions. By assuming the variations of Poisson's ratio caused by the changes of crack porosity are negligible to first order, Berryman, Pride and Wang (2002) deduced the analytical expressions for rocks with dry and saturated cracks. Keys and Xu (2002) decoupled the DEM system by assuming a constant dry rock Poisson's ratio and obtained the analytical dry rock approximation formula, which shows a clear relationship between the elastic parameters and pore parameters and can be easily used for the porosity and pore aspect ratio estimation (Vernik and Kachanov 2010; Bai et al. 2013; Zhao et al. 2013; Keys et al. 2017). Li and Zhang (2012) obtained the analytical expression of the dry rock modulus ratio for different specific pore shapes by assuming the linear relationship between the polarization factors in the DEM system and the modulus ratio. In addition, Jakobsen et al. (2003a, b) derived the common tensor form inclusion model, which can consider other complicated pore microscopic features using the T-matrix method.

The objective of our study is to establish a suitable rock-physics model for porosity and gas saturation prediction of deep-buried sandstone reservoirs. This paper is organized as follows: We first introduce the geological characteristics of studied rocks. Then, we illustrate the theory and methodology of the improved rock-physics model. Next, we apply this model in total porosity estimation, S-wave velocity prediction, and rock-physics template establishment. Next, we use the established template for porosity and gas saturation prediction from real seismic data to demonstrate its performance. Finally, we discuss the parameter calibration rules, the assumptions and uncertainties, and the scale problem by using the improved rock-physics model. Several key conclusions are also drawn at the end of this paper.

Geological setting

The data used in this study are from the East China Sea shelf basin, which is one of the key areas for offshore natural gas exploration in China (Sun et al. 2013). The main target layer is a Paleogene sandstone layer buried more than 4 km depth with an average porosity of 13%. Since the study area is close to the source area, the sandstone has a low compositional maturity and contains unstable feldspar minerals (Hu et al. 2013; Wu et al. 2017). Figure 1 displays the quartz–feldspar–lithic diagram for the Folk's sandstone classification; we observe that the studied sandstone contains an average quartz content of 67.6%, feldspar content of 16.2% and lithic content of 17.2%. The unstable feldspar particles lay a lithologic basis for the development of secondary dissolution pores in deep-buried sandstone reservoirs (Cao et al. 2017).

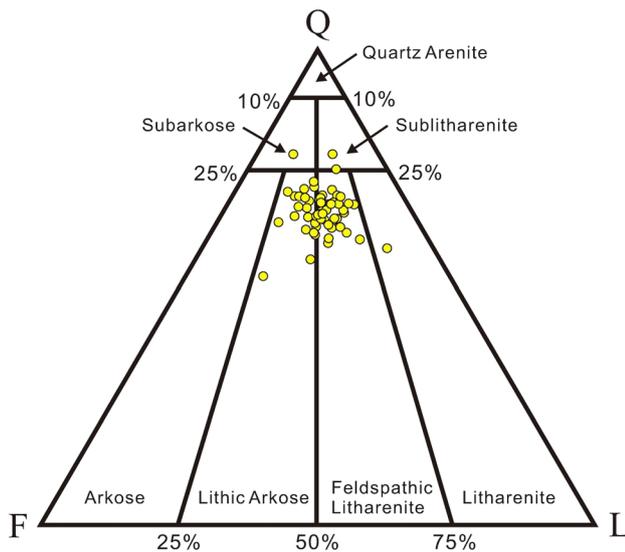


Fig. 1 Quartz–feldspar–lithic (QFL) diagram showing the mineral composition of deep-buried sandstone reservoirs. After Hu et al. (2013)

Figure 2 shows the pore types of deep-buried sandstone reservoirs by scanning electron microscope (SEM) observations. We observe three pore types, which are the primary intergranular pore, the secondary dissolution intra-granular and moldic pore, as well as a few microcracks.

The pore characteristics of rocks are closely related to the burial depth and diagenesis (Avseth et al. 2014). As the burial depth increases, the target sandstone layer in the study area has undergone complicated diagenesis, mainly including compaction, cementation and dissolution (Morad et al. 2010; Cao et al. 2017; Lai et al. 2018). Figure 3 displays the diagenesis and pore evolution mode of the target sandstone reservoirs in different stages. Stage I is the mechanical compaction. We see that the primary intergranular porosity decreases rapidly, and the contact pattern of mineral particles shows changes from the point contact to the suture line contact, accompanied by the quartz overgrowth phenomenon. In this stage, the siliceous cement content is low, with an average content of only 0.5% (Hu et al. 2013). As the burial depth increases in stage II, the calcite and clay mineral cements begin to form and fill the intergranular pores (Fig. 2a), which also decreases the porosity. With the further increase in the burial depth in stage III, part of the feldspar particles is dissolved by organic acid, forming secondary pores, including intragranular dissolved pores (Fig. 2b) and moldic pores (Fig. 2c), which greatly improves the reservoir pore space.

Methodology and model test

The extended Xu–White model

According to the above-mentioned geological analysis, the major mineral components of deep-buried low-porosity sandstone reservoirs in the study area are quartz, feldspar and clay. The pore types mainly include primary quartz intergranular semi-stiff pores, secondary feldspar dissolution stiff pores and wet clay soft pores.

Based on the Kuster–Toksöz (1974) and the differential effective medium theories, Xu and White (1995) proposed a clay–sand mixture rock-physics model considering both sand pores and wet clay pores. However, it inherits the defects of the Kuster–Toksöz model and the DEM model, which have low computational efficiency and are sensitive to the pore addition order (Keys and Xu 2002). Here, we extend the Xu–White model by introducing a third type of feldspar-related pores to establish a rock-physics model for deep-buried low-porosity gas-bearing sandstone reservoirs. The modelling process is mainly divided into three steps according to the Xu–White model framework, which includes the modelling of the solid rock matrix, dry rock, and fluid-saturated rock (Fig. 4).

Step 1. Solid rock matrix effective elastic moduli calculation

Following the Xu–White model (1996), we use the time-average equations (Wyllie et al. 1956) to calculate the effective moduli of the solid rock matrix composed of sand grains and clay particles. According to the clastic rock bulk volume formula (Guéguen and Palciauskas 1994),

$$V_S + V_C + \phi = 1, \quad (1)$$

where V_S and V_C represent the sand and clay volume of the whole rock; ϕ is the total porosity. The parameters V_C and ϕ are obtained from well-log data. Then, we can obtain the sand and clay volumes of the rock matrix which are given by:

$$V_C^m = \frac{V_C}{1 - \phi}, \quad (2)$$

$$V_S^m = \frac{1 - \phi - V_C}{1 - \phi}, \quad (3)$$

The P- and S-wave transit time of the rock matrix can be calculated using the time-average equations

$$T_m^p = V_S^m T_S^p + V_C^m T_C^p, \quad (4)$$

$$T_m^s = V_S^m T_S^s + V_C^m T_C^s, \quad (5)$$

and the density of the rock matrix is calculated by

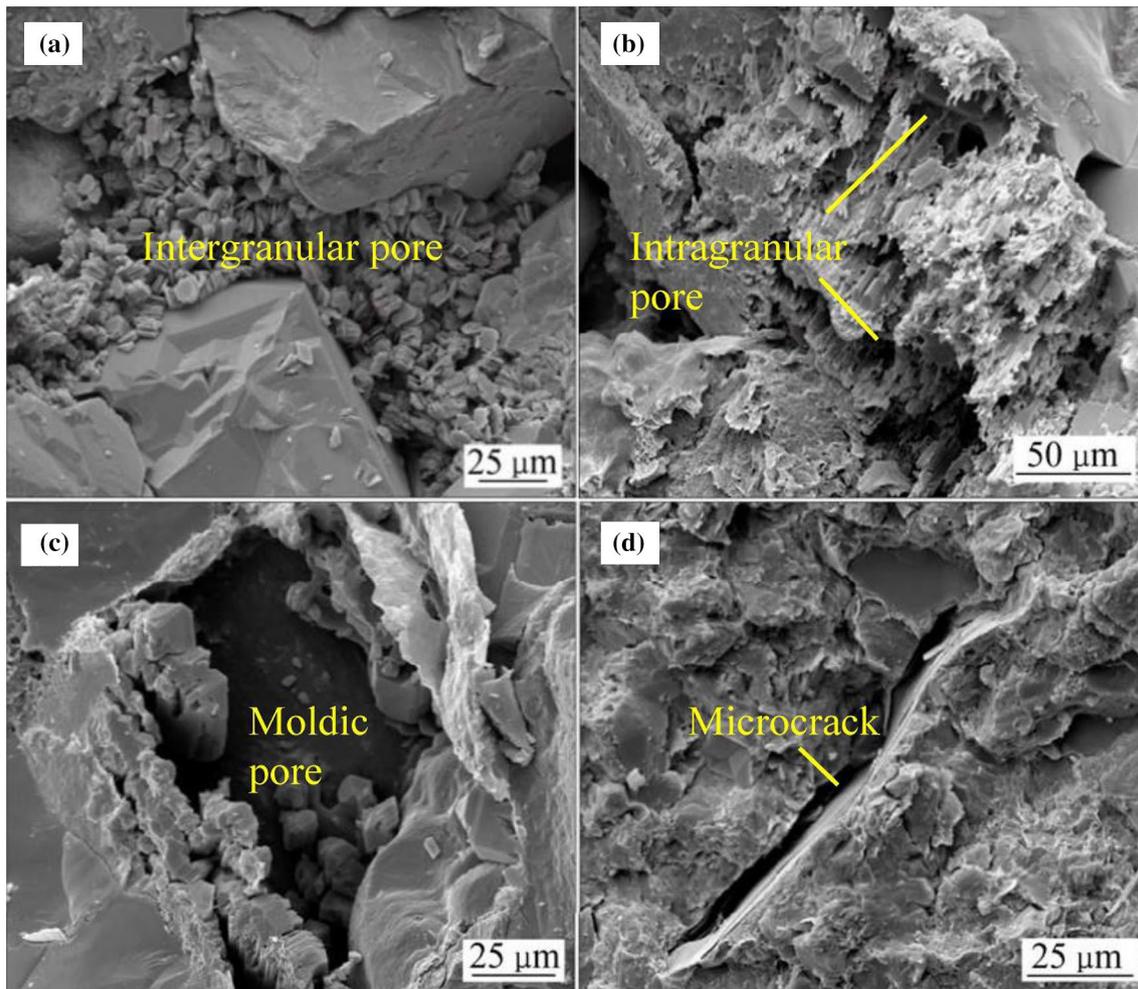


Fig. 2 Pore types of deep-buried sandstone reservoirs by SEM observations (after Cao et al. 2017): **a** primary intergranular pores filled with authigenic kaolinite; **b** secondary feldspar dissolution intragranular pores; **c** secondary feldspar dissolution moldic pores; **d** microcracks

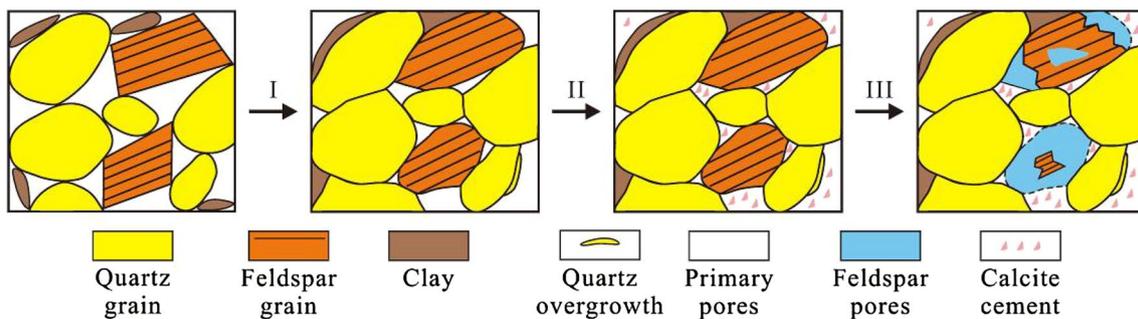


Fig. 3 Diagenesis and pore evolution mode of deep-buried sandstone reservoirs. Stage I represents mechanical compaction, stage II represents early cementation and stage III represents the dissolution of feldspar grains

$$\rho_m = V_S^m \rho_S + V_C^m \rho_C, \quad (6)$$

where T_m^p , T_S^p and T_C^p are the P-wave transit time of the rock matrix, sand and clay, respectively; T_m^s , T_S^s and T_C^s are the

corresponding S-wave transit time, respectively; and ρ_m , ρ_S and ρ_C represent the corresponding density, respectively.

Then, the effective bulk modulus K_m and shear modulus μ_m of solid rock matrix can be expressed as

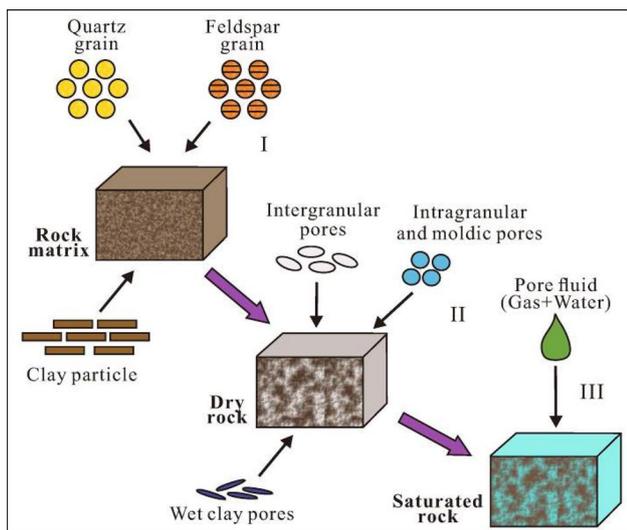


Fig. 4 Schematic diagram of the extended Xu–White model for deep-buried sandstone reservoirs (after Xu and Payne 2009). Step I represents the solid rock matrix modelling, step II represents the dry rock modelling, and step III represents the fluid-saturated rock modelling

$$K_m = \rho_m \left[\left(\frac{1}{T_m^p} \right)^2 - \frac{4}{3} \left(\frac{1}{T_m^s} \right)^2 \right], \tag{7}$$

$$\mu_m = \rho_m \left(\frac{1}{T_m^s} \right)^2. \tag{8}$$

Step 2. Multi-pore dry rock effective elastic moduli calculation

In a two-phase composite, assuming that the host material has elastic moduli K_m and μ_m , and the inclusion material has elastic moduli K_i and μ_i , with porosity ϕ , the effective bulk and shear moduli of this composite $K^*(\phi)$ and $\mu^*(\phi)$ can be obtained by solving the coupled system of ordinary differential equations (Berryman 1992)

$$(1 - \phi) \frac{d}{d\phi} [K^*(\phi)] = [K_i - K^*(\phi)] P_i^*, \tag{9}$$

$$(1 - \phi) \frac{d}{d\phi} [\mu^*(\phi)] = [\mu_i - \mu^*(\phi)] Q_i^*, \tag{10}$$

where P_i^* and Q_i^* are the geometric factors associated with Poisson’s ratio of the composite and the aspect ratio of the inclusion material, respectively (Keys and Xu 2002; Berryman et al. 2002), and i represents the inclusion phase. The expressions of geometric factors can be easily found in many

published papers (Cheng and Toksöz 1979; Xu and White 1995; Keys and Xu 2002).

For the dry rock, $K_i = 0$ and $\mu_i = 0$. Equations (9) and (10) become

$$\frac{d[K^*(\phi)]}{[K^*(\phi)]d\phi} = -\frac{P_i^*}{(1 - \phi)}, \tag{11}$$

$$\frac{d[\mu^*(\phi)]}{[\mu^*(\phi)]d\phi} = -\frac{Q_i^*}{(1 - \phi)}. \tag{12}$$

Assuming that the dry rock Poisson’s ratio is constant, the geometric factors P_i^* and Q_i^* can be approximated by P_i and Q_i , which are now related to Poisson’s ratio of the rock matrix and aspect ratio of the inclusion pores. In this case, the geometric factors are independent of the porosity ϕ , and the ordinary differential equations system can be decoupled (Keys and Xu 2002). Then, integrating Eqs. (11) and (12) within the porosity interval 0 to ϕ_i gives

$$\ln K^*(\phi_i) - \ln K^*(0) = P_i \ln(1 - \phi_i), \tag{13}$$

$$\ln \mu^*(\phi_i) - \ln \mu^*(0) = Q_i \ln(1 - \phi_i). \tag{14}$$

Combining the initial conditions $K^*(0) = K_m$ and $\mu^*(0) = \mu_m$ with Eqs. (13) and (14), the single-pore dry rock analytical solutions of the effective bulk and shear moduli can be written as

$$K^*(\phi_i) = K_m(1 - \phi_i)^{P_i}, \tag{15}$$

$$\mu^*(\phi_i) = \mu_m(1 - \phi_i)^{Q_i}. \tag{16}$$

For the multi-pore system of deep-buried sandstone reservoirs, we consider the total pore volume ϕ is composed of quartz intergranular semi-stiff pores ϕ_Q , feldspar dissolution stiff pores ϕ_F and wet clay soft pores ϕ_C ; then, we have

$$\phi = \phi_S + \phi_C = \phi_Q + \phi_F + \phi_C \tag{17}$$

Since the porosity of each pore type is proportional to the corresponding mineral content to the first-order approximation (Xu and White 1995), it can be denoted as $\phi_Q = aV_S^m \phi$, $\phi_F = bV_S^m \phi$, $\phi_C = V_C^m \phi$, where the coefficients a and b can be obtained from the analysis of rock slices. Finally, the multi-pore dry rock analytical approximation of effective bulk and shear moduli can be expressed as:

$$K^*(\phi) = K_m(1 - \phi_Q)^{P_Q}(1 - \phi_F)^{P_F}(1 - \phi_C)^{P_C} \tag{18}$$

$$\mu^*(\phi) = \mu_m(1 - \phi_Q)^{Q_Q}(1 - \phi_F)^{Q_F}(1 - \phi_C)^{Q_C} \tag{19}$$

Compared with the DEM Eqs. (9) and (10), the multi-pore dry rock analytical approximation Eqs. (18) and (19) are more computationally efficient and are not affected by the addition order of the different pore types.

Step 3. Fluid-saturated rock effective elastic moduli calculation

A fundamental assumption of Gassmann theory is that the pore pressures are equilibrated throughout the pore space (Mavko et al. 2009), which means it is appropriate for rocks with good fluidity and high porosity. However, due to the low porosity and clay–sand mixture background of deep-buried reservoirs in this study area, the pore fluid could have a patchy distribution behaviour (White 1975; Müller et al. 2010; Avseth et al. 2010).

Ignoring attenuation and dispersion caused by the wave-induced fluid flow in porous media, patchy-saturation model (Mavko et al. 2009) calculates the bulk modulus of each fluid phase locally using the Gassmann equation with the shear modulus remaining unchanged. Then, the effective moduli of the patchy-saturated rock can be approximately estimated by the Reuss average. Considering the gas–water two-phase fluid situation, the calculation process can be described by

$$K_{\text{sat},i} = K_{\text{dry}} + \frac{(1 - K_{\text{dry}}/K_m)^2}{\phi/K_{f,i} + (1 - \phi)/K_m - K_{\text{dry}}/K_m^2}, \quad (20)$$

$$\mu_{\text{sat}} = \mu_{\text{dry}}, \quad (21)$$

$$M_{\text{sat},i} = K_{\text{sat},i} + \frac{4}{3}\mu_{\text{sat}}, \quad (22)$$

$$M_{\text{sat}} = \left(\frac{S_g}{M_{\text{sat},g}} + \frac{S_w}{M_{\text{sat},w}} \right)^{-1}, \quad (23)$$

where $K_{f,i}$, $K_{\text{sat},i}$ and $M_{\text{sat},i}$ are the bulk modulus, saturated rock bulk modulus and saturated rock compressional modulus of fluid phase i , respectively; K_{dry} and μ_{dry} are the dry rock bulk and shear moduli obtained in step 2; S_g and S_w are gas saturation and water saturation, respectively; and M_{sat} and μ_{sat} are the effective compressional modulus and shear modulus of fluid-saturated rock, respectively.

Finally, from the saturated rock modulus we can calculate the P-wave velocity $V_p = \sqrt{M_{\text{sat}}/\rho_{\text{sat}}}$ and S-wave velocity $V_s = \sqrt{\mu_{\text{sat}}/\rho_{\text{sat}}}$, where the bulk density is given by $\rho_{\text{sat}} = (1 - \phi)\rho_m + \phi(S_w\rho_w + S_g\rho_g)$.

Theory model test

To analyse the applicability of the proposed extended Xu–White model and the influence of the pore aspect ratio on elastic parameters, theory model tests of single- and multi-pore type rocks are performed on water–quartz and gas–quartz systems, respectively. We take quartz as the host matrix, having $K_m = 36.6$ GPa, $\mu_m = 37$ GPa, $\rho_m = 2650$ kg/m³, and water or gas as the pore fluid,

having $K_w = 2.5$ GPa, $\rho_w = 1020$ kg/m³, $K_g = 0.05$ GPa, $\rho_g = 140$ kg/m³ (Mavko et al. 2009). The porosity range is set to 0–0.4. According to the average mean aspect ratios of different pore types defined by Xu and Payne (2009), the pore aspect ratios here are set to $\alpha = 0.02$ (soft pores), $\alpha = 0.12$ (semi-stiff pores) and $\alpha = 0.8$ (stiff pores).

Figure 5 displays the single-pore test results. The bulk and shear moduli of saturated rock are calculated using different rock-physics modelling methods with a single-pore aspect ratio. In both the water saturation and gas saturation models, the results of Hashin–Shtrikman bounds (black lines) and DEM model (blue lines) are considered as references to compare with the results of Kuster–Toksöz model (green points) and the proposed extended Xu–White model (red points). Note that the Kuster–Toksöz model results are close to the DEM model only when the porosity is smaller than the pore aspect ratio, and some points exceed the Hashin–Shtrikman lower bound. When the pore aspect ratio is close to 1, the Kuster–Toksöz model results are close to the Hashin–Shtrikman upper bound. The results obtained by the extended Xu–White model are sound, consistent well with the DEM results in the given porosity interval, and always within the Hashin–Shtrikman upper and lower bounds.

Figure 6 displays the multi-pore test result. The elastic moduli of saturated rock are calculated using the extended Xu–White model with different soft and stiff pore volumes along with the background semi-stiff pores. The red solid line represents rock moduli with only semi-stiff pores. The lower dashed green lines and the upper dashed blue lines indicate the increasing volume of soft pores and stiff pores, respectively. From both the water saturation and gas saturation models, we can observe that the bulk and shear moduli decrease with the increase in the soft pore volume and increase with the increase in the stiff pore volume. This rule is useful in the selection and adjustment of the rock-physics parameters in the real data application. In addition, the results are also consistent with the Hashin–Shtrikman upper and lower bounds within the given porosity range.

Application

Diagnostics of well-log data

Well-log data can effectively reflect the properties of rocks and pore fluids underground and are often used in the calibration and verification of rock-physics models. The quality of well logs has a significant influence on reservoir characteristics analysis and the establishment of rock-physics templates. Figure 7 displays the original well logs and interpreted log-facies for the target interval of a calibration

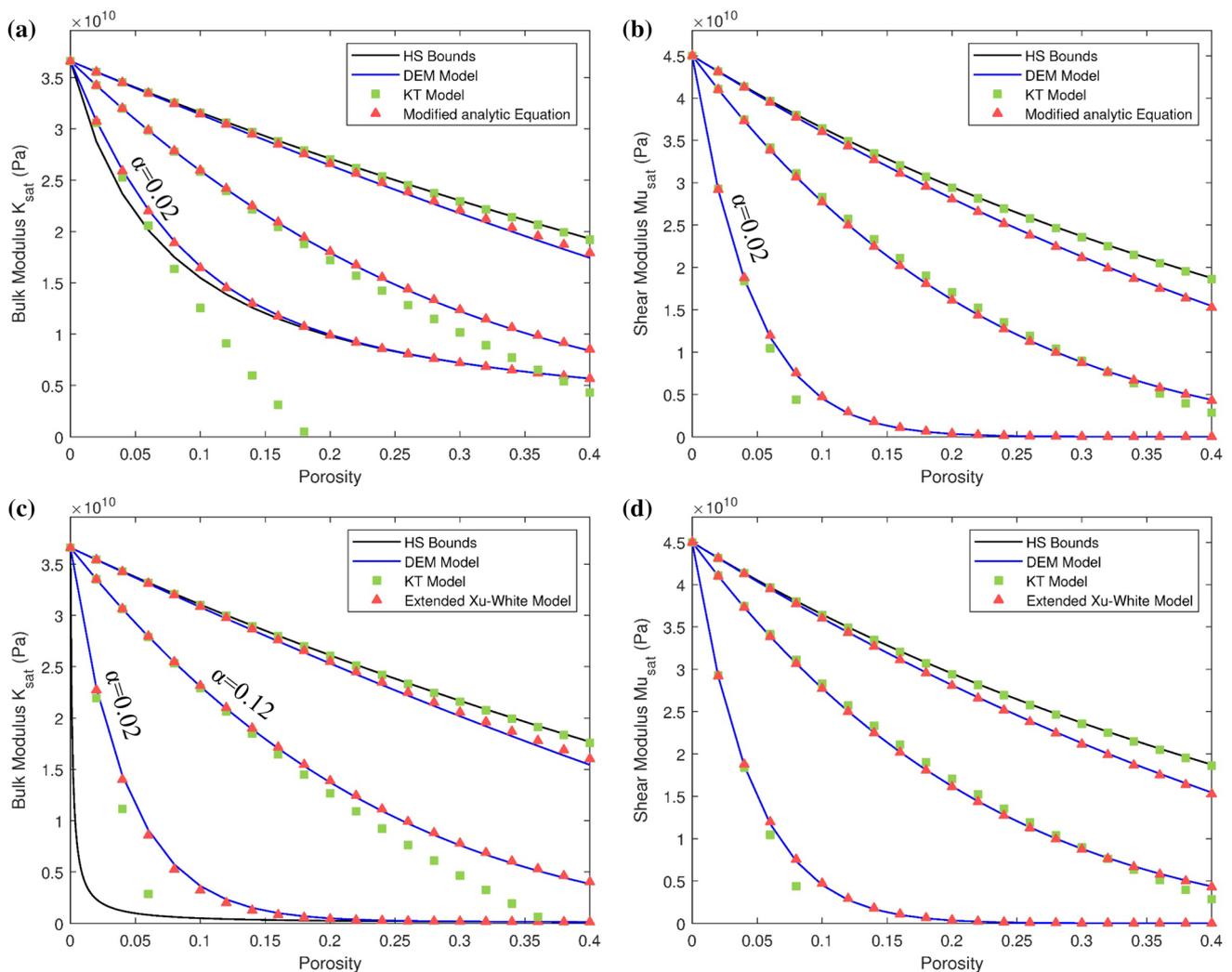


Fig. 5 Single-pore-saturated rock modelling with different pore aspect ratios, water-saturated bulk modulus (a) and shear modulus (b) versus porosity, gas-saturated bulk modulus (c) and shear modulus (d) versus porosity

well W1. The data set contains a complete set of well logs, including sonic logs (P- and S-wave velocities), a density log and petrophysical curves (porosity, shale volume and water saturation). Two main gas-bearing sandstone reservoirs are shown in yellow facies; their thicknesses are about 15 m and 20 m; and the average porosity inside the reservoir is approximately 13%. The original porosity curve in Fig. 7d is the effective porosity. Note that the shale layers have zero porosity and 100% shale volume values, as shown by the red arrows in Fig. 7. These values are not appropriate for rock-physics modelling. In addition, due to these zero porosity values, the crossplots of the effective porosity versus the P- and S-wave velocity in Fig. 8 show a messy distribution, especially the shale and siltstone points in the marked red ellipses. Thus, we need to estimate the total porosity before rock-physics modelling.

Since rock-physics models predict P- and S-wave velocities from the total porosity and shale volume, they can be used in an inversion mode to estimate a total porosity curve that matches the measured velocities. We use the extended Xu–White model to estimate the total porosity constrained by the measured P-wave velocity and compare the results with the conventional Xu–White model and density method. The estimated total porosity is then used in the prediction of the S-wave velocity. Because no constraint is imposed on the predicted S-wave velocity, the match between the predicted and measured S-wave velocities can reflect the reliability of the total porosity estimation result to some extent. The rock properties for the total porosity estimation using the extended Xu–White model and the conventional Xu–White model are listed in Table 1. Among them, the solid parameters (bulk modulus, shear modulus and density) and the pore aspect ratios are determined from the calibration of well W1;

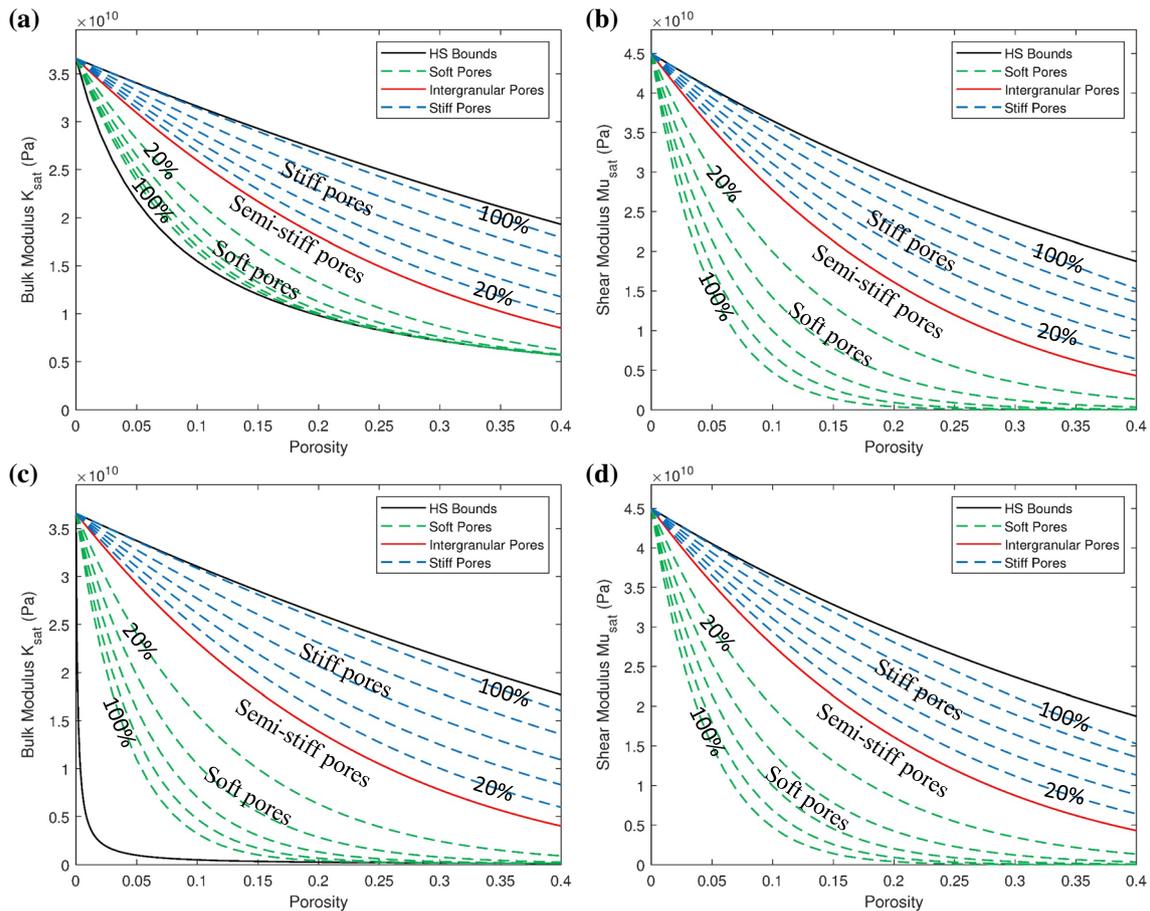


Fig. 6 Multi-pore-saturated rock modelling with different volumes of soft and stiff pores, water-saturated bulk modulus (a) and shear modulus (b) versus porosity, gas-saturated bulk modulus (c) and shear modulus (d) versus porosity

the fluid parameters are calculated by the Batzle–Wang formulas (1992) at a temperature of 170 °C and a pore pressure of 25 MPa for the target zone. In the extended Xu–White model, we consider three pore types, the intergranular semi-stiff pore, feldspar dissolution stiff pore and wet clay soft pore, while in the conventional Xu–White model, only two pore types, the semi-stiff pore and soft pore, are considered.

Figure 9a shows the corrected and original shale volume curves. The corrected curve is more reasonable as it does not contain 100% shale volume values, considering the existence of bound wet clay pores in shale layers. Figure 9b compares the estimated total porosity curves using the extended Xu–White model, conventional Xu–White model and density method. The density-estimated porosity (green) has values that exceed the general sandstone critical porosity; the extended Xu–White model result (red) is close to the conventional Xu–White model result (blue) overall but is more accurate at the sandstone reservoir locations, as the red arrows show. Figure 9c–e shows the comparisons of the predicted and measured P- and S-wave velocities and densities, using the total porosity estimated by the extended and conventional Xu–White models,

respectively. Since the measured P-wave velocity is used as a constraint in the total porosity estimation, the predicted P-wave velocity curves overlap with each other as expected. The S-wave velocity predicted by the extended Xu–White model (red) matches better with the measured S-wave velocity (black) than the conventional Xu–White model result (blue), which indicates that the total porosity estimated by the extended model is more reliable. In addition, the predicted density provides a more realistic value than the original density at the location where the calliper log shows an expansion in Fig. 9f.

Furthermore, we replace the original effective porosity of the crossplots in Fig. 8 with the estimated total porosity by the extended Xu–White model, as shown in Fig. 10. Note that the crossplots show clear rock-physics characteristics in this case. The scattered data points of different lithologies distributed regularly between the theoretical clean sandstone line and pure shale line, and the shale volume trend is consistent with the log interpretation results, as the black arrows indicate.

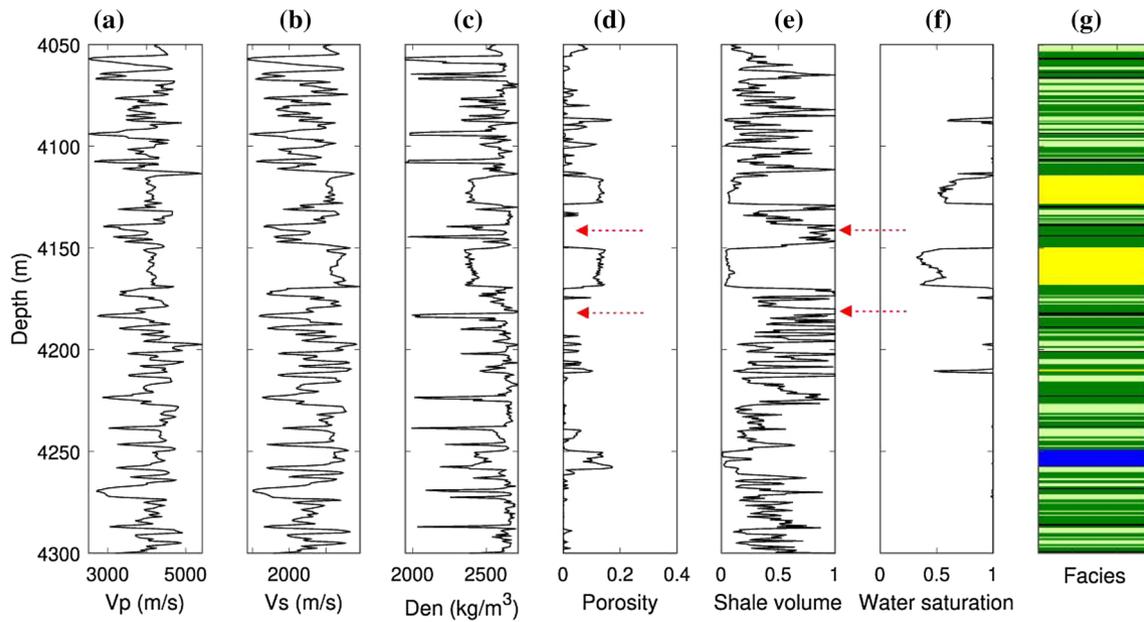


Fig. 7 Original log data of well W1. **a** P-wave velocity; **b** S-wave velocity; **c** density; **d** porosity; **e** shale volume; **f** water saturation; **g** facies defined by well-log interpretation. In the facies profile, shale

is dark green, siltstone is light green, the coal formation is black, the water-saturated sandstone is blue, and the gas-bearing sandstone is yellow

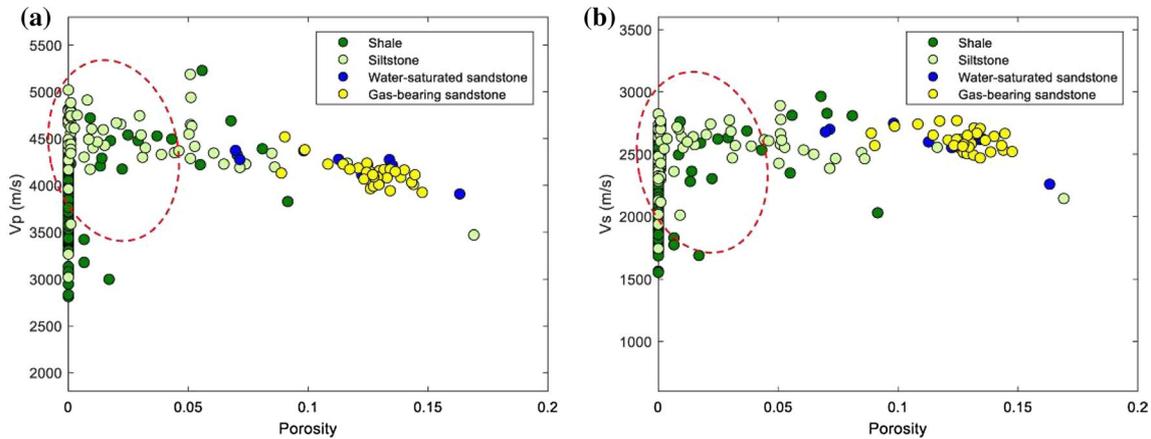


Fig. 8 Crossplots of P-wave velocity **(a)** and S-wave velocity **(b)** versus the original effective porosity from well W1

Table 1 Rock-physics model parameters for total porosity estimation

Lithology or fluid	Bulk modulus (GPa)	Shear modulus (GPa)	Density (g/cm ³)	Pore type	Pore aspect ratio
Sand	43.00	32.00	2.66	Semi-stiff	0.12
				Stiff	0.60
Clay	25.00	9.00	2.58	Soft	0.05
Water	2.33	0.00	0.95		
Gas	0.05	0.00	0.14		

The extended Xu–White model uses the semi-stiff, stiff and soft pores, while the conventional Xu–White model uses only the semi-stiff and soft pores

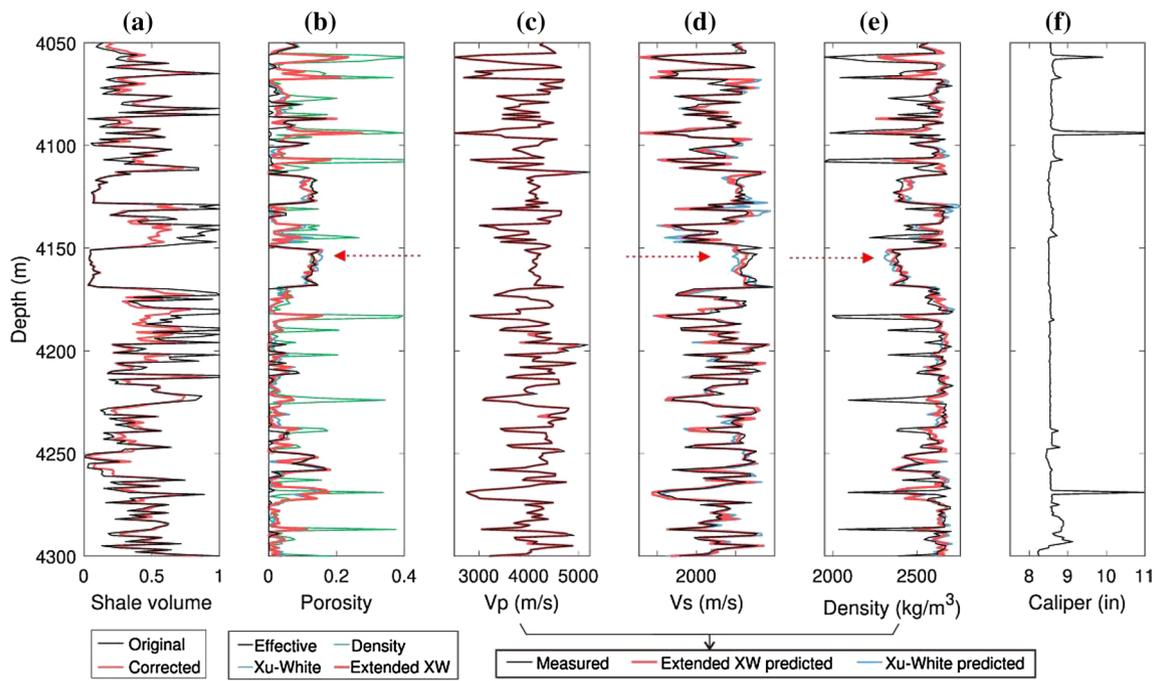


Fig. 9 Total porosity estimation-related curves. **a** Shale volume; **b** porosity; **c** P-wave velocity; **d** S-wave velocity; **e** density; **f** calliper log. The measured P-wave velocity is used as a constraint to invert

the total porosity, and the measured S-wave velocity is used as a reference to test the total porosity estimation result

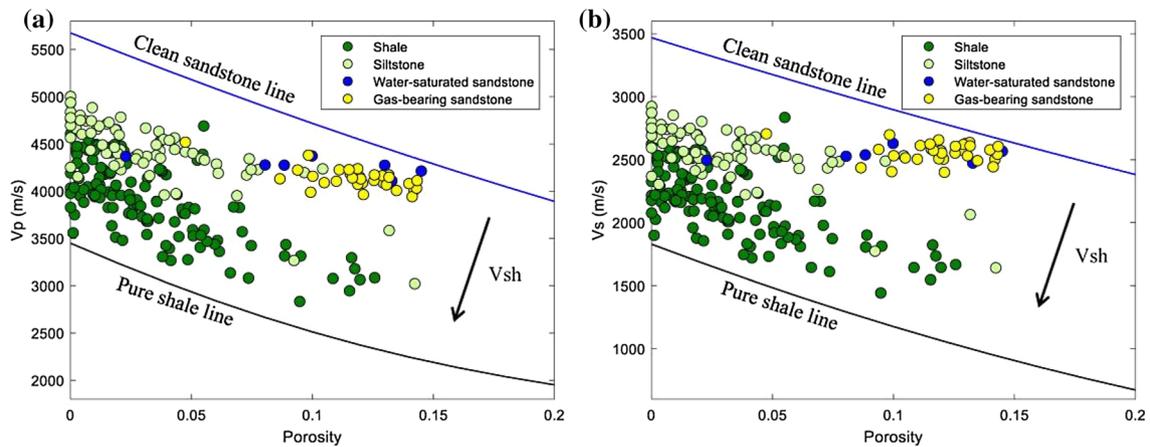


Fig. 10 Crossplots of P-wave velocity (**a**) and S-wave velocity (**b**) versus the estimated total porosity. The blue line and black line represent the theoretical clean sand line and the pure shale line, respectively. The black arrows indicate the increasing direction of the shale volume

Establishment and verification of rock-physics template

Rock-physics templates (RPTs) are charts or templates of rock-physics models for the interpretation of lithology and hydrocarbons (Ødegaard and Avseth 2004; Chi and Han 2009; Avseth et al. 2010). RPTs consist of crossplots of seismic parameters (e.g. acoustic impedance vs. V_p/V_s ratios),

which allow us to perform the rock-physics analysis of both well data and elastic inversion results.

Figure 11 displays the crossplots of acoustic impedance versus V_p/V_s ratios for the target zone of the calibration well W1, superimposed with RPT generated from the extended Xu–White model using the previous rock parameters in Table 1. The porosity range is set from 0 to 20%, and the shale volume and the water saturation ranges are set from 0 to 100%. Their variation trends are shown by arrows in each

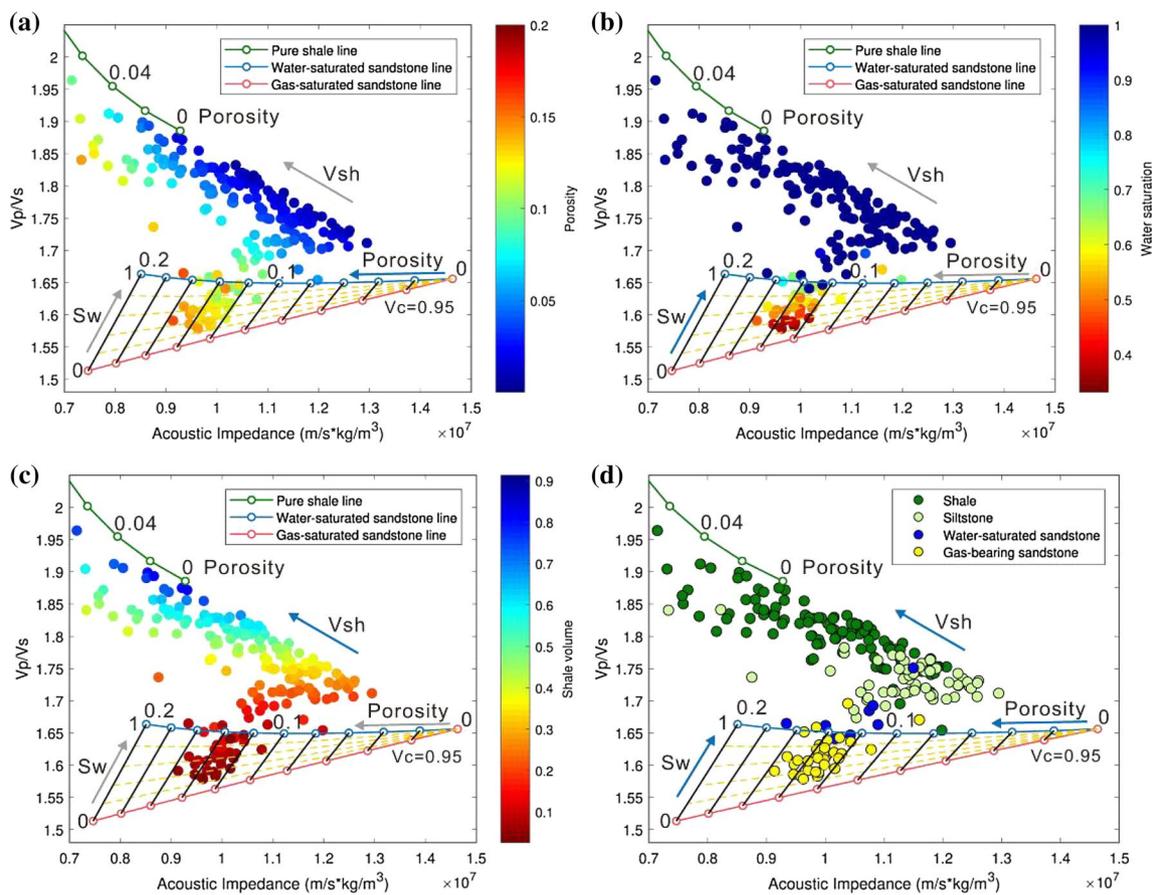


Fig. 11 Acoustic impedance versus V_p/V_s crossplots of well-log data from calibration well W1 superimposed with the modelled RPT and colour-coded by: **a** porosity, **b** water saturation, **c** shale volume and **d** well-log interpretation conclusions

crossplot of Fig. 11. The green solid line represents pure shale; the blue and red solid lines represent water- and gas-saturated sandstone with 5% clay content, respectively; and the black solid lines and the yellow dotted lines represent isolines of porosity and water saturation, respectively. Note that the RPT-indicated porosity for the data points in the reservoir zone is approximately 10–15%, which is consistent with the porosity shown by the colour-coded values of these points in Fig. 11a. The RPT-indicated water saturation and shale volume are also consistent with the corresponding colour-coded values in Fig. 11b, c. In addition, in Fig. 11d, we observe that the artificial log-interpreted gas-bearing sandstone points almost fall in the reservoir zone, and the water-saturated sandstone, siltstone and shale points sit just between the theoretical water-saturated sandstone line and pure shale line.

Our rock-physics model uses patchy distribution to calculate fluid-saturated rock moduli, as previously mentioned. For comparison, we here assume a uniform distribution of fluid in the RPT modelling. The result is shown in Fig. 12. Note that the main difference in Figs. 11b and 12 is the water saturation of the template. In the patchy distribution

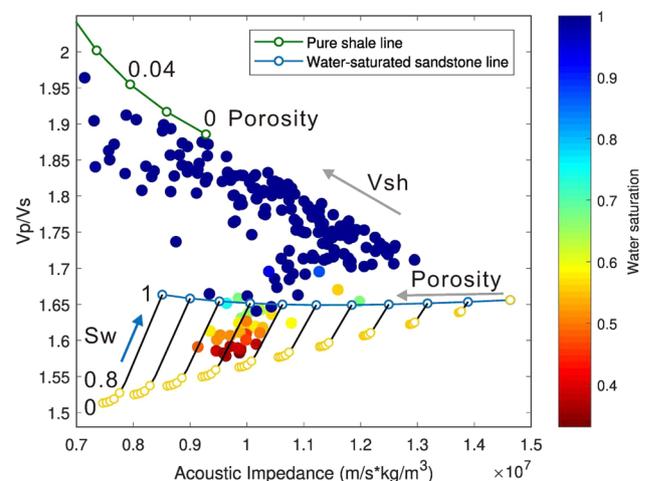


Fig. 12 Acoustic impedance versus V_p/V_s crossplots of well-log data from calibration well W1 superimposed with the comparative RPT using uniform fluid distribution and colour-coded by water saturation

mode, the elastic properties show a more linear change with increasing gas saturation, as shown by the yellow dotted lines in Fig. 11b, while in the uniform distribution mode,

a small amount of gas will cause almost the same elastic properties as the commercial amount of gas, as shown by the yellow circles in Fig. 12. In addition, the RPT-indicated minimum water saturation under the uniform distribution assumption is approximately 80%, which is not consistent with the minimum colour-coded values of data points, which is approximately 35%. This demonstrates that the patchy distribution assumption of the fluid saturation mode tends to be more suitable than the uniform distribution in our case.

To investigate the scale effect on the RPT, we apply a moving-window Backus averaging method (Lindsay and Van Koughnet 2001) to upscale the log curves to seismic scale in Fig. 13. The size of the upscaling window at a certain depth is dynamically determined by estimating the temporal wavelength from a chosen upscaling frequency and the velocity at that depth. The upscaling frequency here is set to 100 Hz according to the bandwidth of the seismic inversion results. Note that the ranges of the P- and S-wave velocity and density curves decrease substantially in Fig. 13 due to the filtering effect of the Backus average. After this, the scattered data points in Fig. 14 are transformed to the seismic scale. We observe that the best-fit template clay content value decreases from 0.95 to 0.85 when the data points upscale from the well-log scale to the seismic scale.

Then, a verification well, W2, is selected to verify the reliability of the template. Figure 15 displays the crossplots

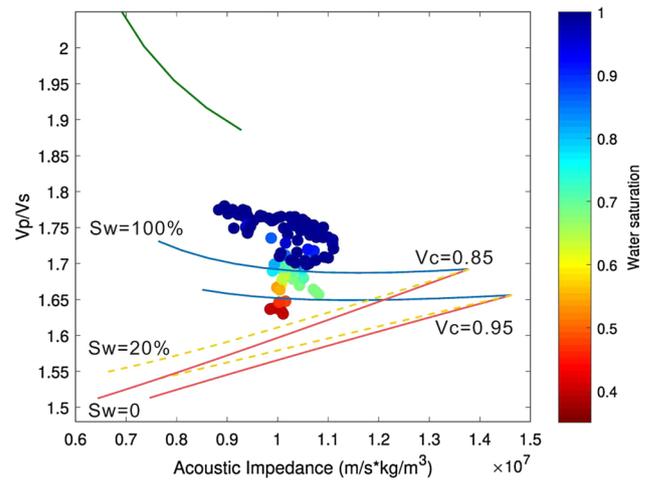
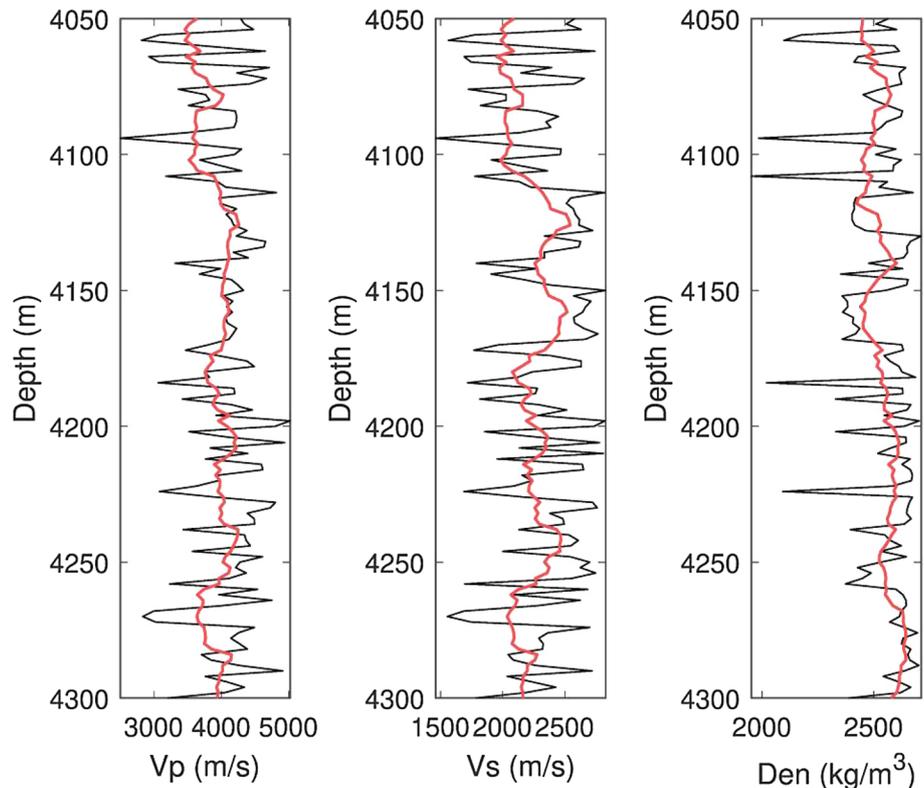


Fig. 14 Acoustic impedance versus Vp/Vs crossplots of seismic-scale log data in Fig. 13 superimposed with RPTs with different clay contents and colour-coded by water saturation

of the acoustic impedance versus Vp/Vs ratios for the target interval of the verification well W2, superimposed with the same RPT calibrated by well W1 at the well-log scale. Note that the template-indicated gas-bearing sandstone porosity range (10–14%), minimum water saturation (50%) and shale volume trend are consistent with the colour-coded values of scattered data points. In addition, the upscaled data points

Fig. 13 Backus average upscaling of well-log data; the black curves represent the original log data, and the red curves represent seismic-scale log data



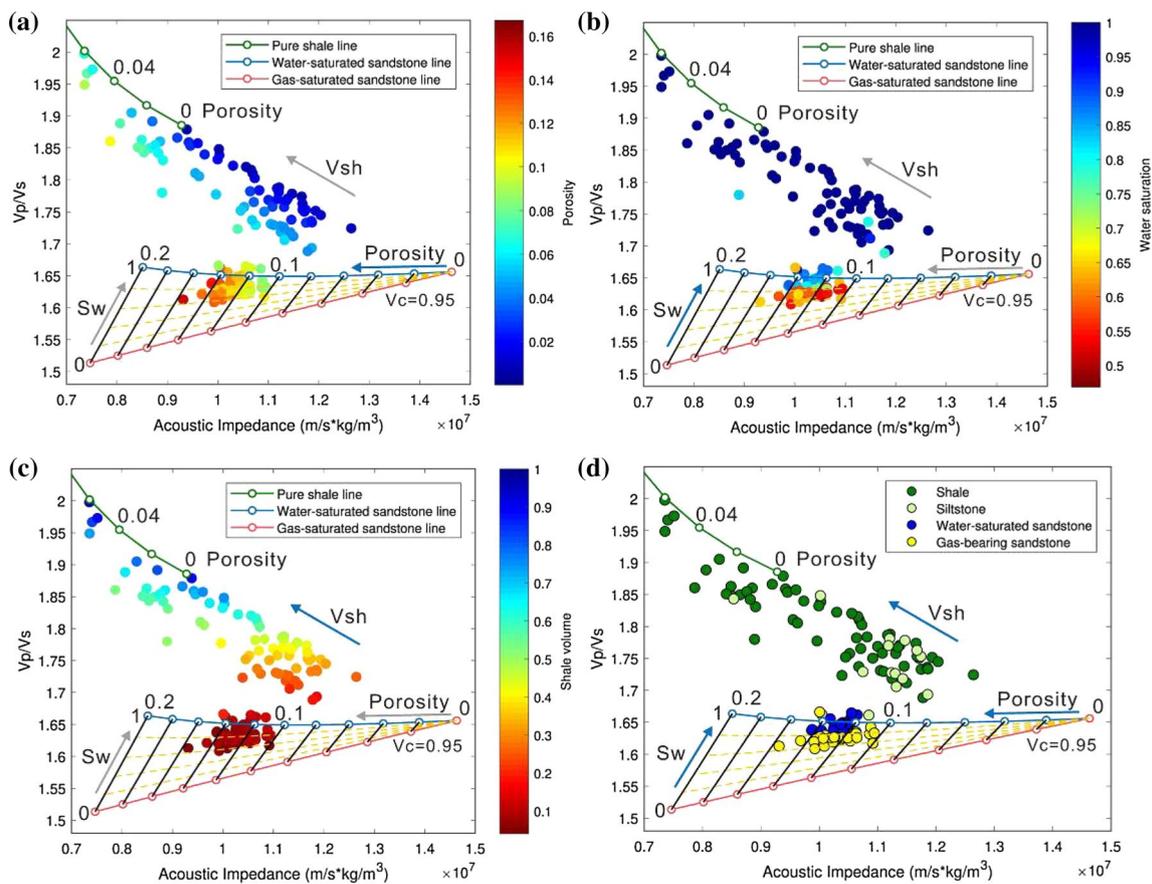


Fig. 15 Acoustic impedance versus V_p/V_s crossplots of well-log data from verification well W2 superimposed with the same rock-physics template calibrated by well W1 and colour-coded by: **a** porosity, **b** water saturation, **c** shale volume and **d** well-log interpretation conclusions

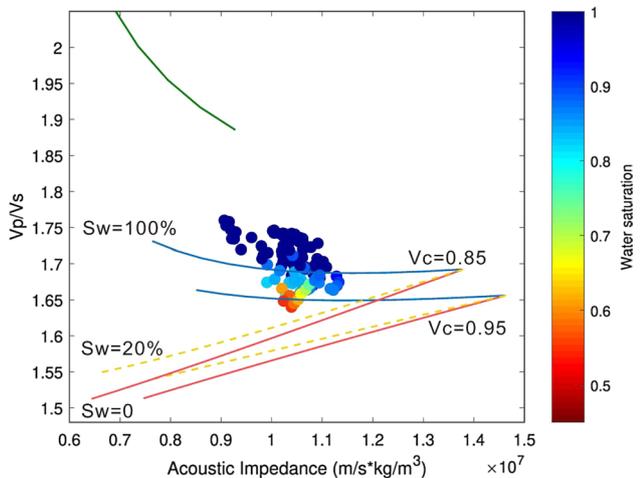


Fig. 16 Acoustic impedance versus V_p/V_s crossplots of seismic-scale log data from verification well W2 superimposed with RPTs with different clay contents and colour-coded by water saturation

in Fig. 16 also match well with the seismic-scale template in which the clay content is equal to 0.85. These demonstrate a good validity of the established RPT at both the well-log and the seismic scales.

Porosity and gas saturation prediction

In the following section, we apply the established RPT to predict the porosity and gas saturation of deep-buried sandstone reservoirs from the prestack seismic data. Figure 17 displays three partial angle stack seismic sections passing through well W1 which is located at CDP 1071. The inserted well synthetic seismic records in each section are calculated from the elastic parameters (P- and S-wave velocity and density) of well-log data and a 30-Hz zero-phase Ricker wavelet using the Aki and Richards equation (2002) with the corresponding average incident angles of 9°, 15°, 21°. Note that the synthetic seismic records match well with the real seismic data and has an obvious amplitude-versus-angle (AVA) response at the gas layers. The seismic amplitude decreases as the incidence angle increases, which can also be observed from the

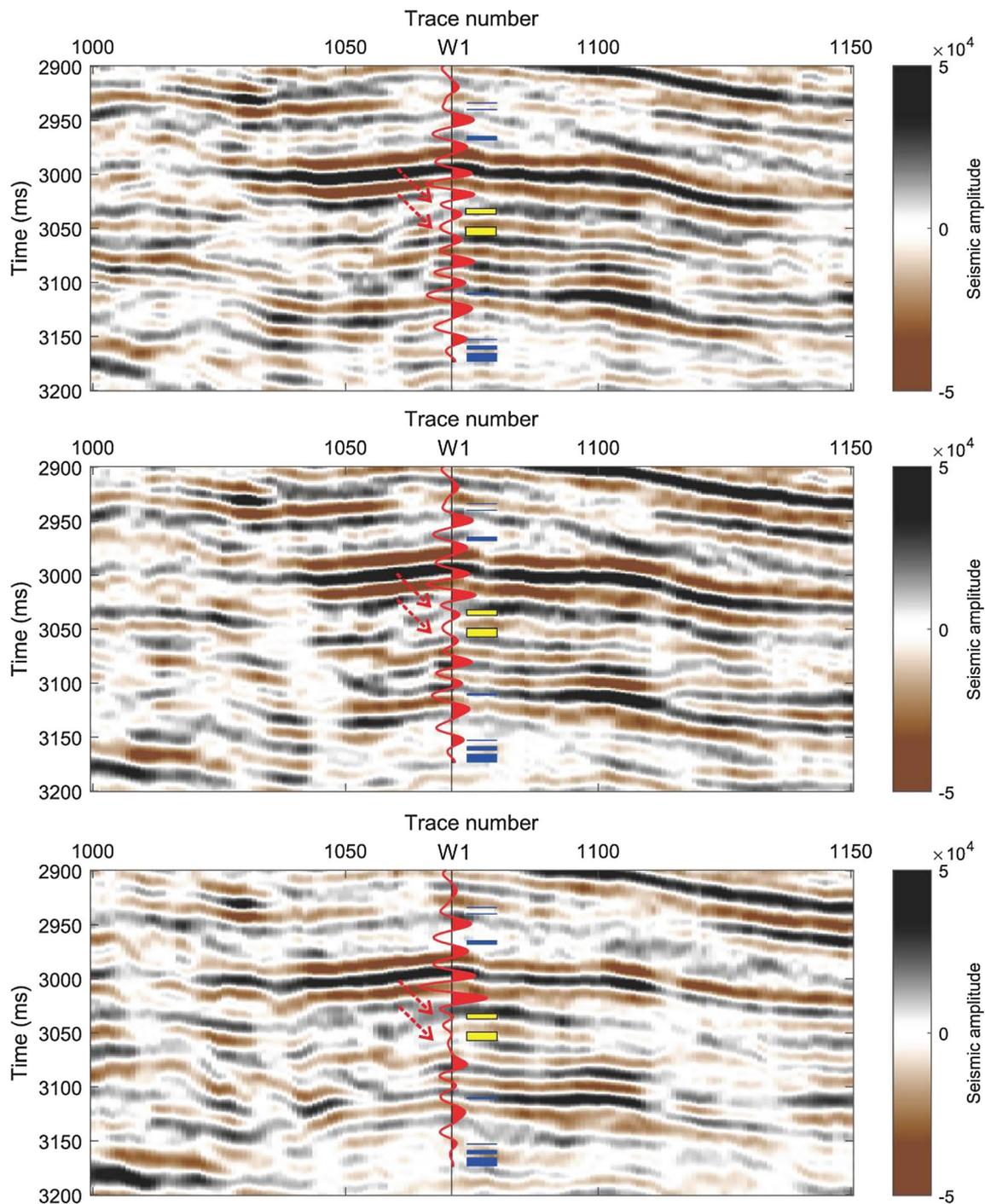


Fig. 17 Partial angle stack seismic sections crossing well W1 with average incident angles of 9° , 15° , 21° , respectively. The inserted well curve represents the calibrated synthetic seismic records generated by

the corresponding average incident angles; the reservoir interpretation results are displayed on the right side; yellow rectangle represents gas sand and blue rectangles represent brine sand

borehole-side seismic data. Figure 18 shows the acoustic impedance and V_p/V_s inversion results for the seismic line in Fig. 17 using the Bayesian linearized AVO inversion (Buland and Omre 2003). Arrows indicate two gas-bearing

sandstone reservoirs with medium acoustic impedance and low V_p/V_s values. Note that we can only roughly estimate the locations of two gas layers just from the inversion results. However, with the RPT, we can further estimate

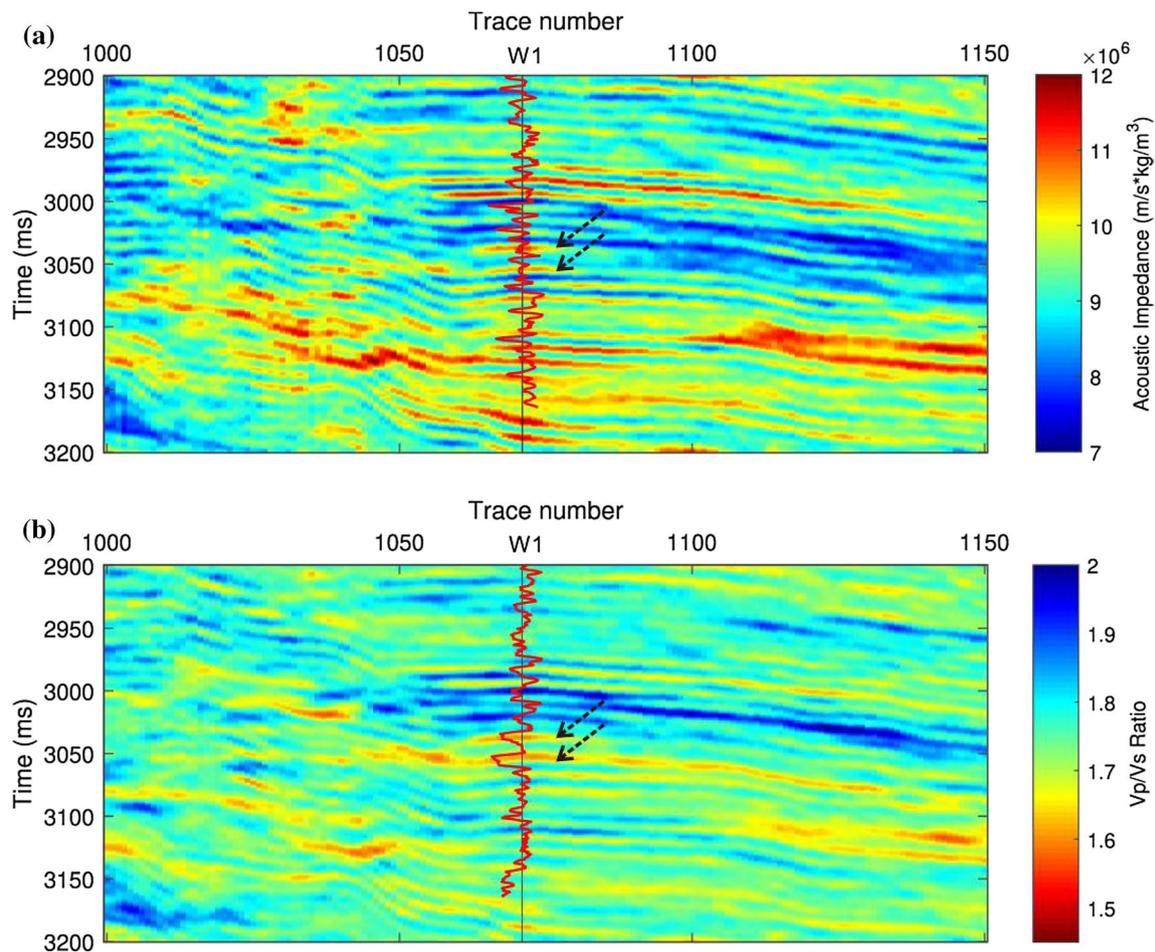


Fig. 18 Prestack seismic inversion results of acoustic impedance (a) and Vp/Vs (b) with the corresponding log cures inserted. Arrows indicate the locations of two gas reservoirs in well W1

the porosity and gas saturation of gas-bearing sandstone reservoirs.

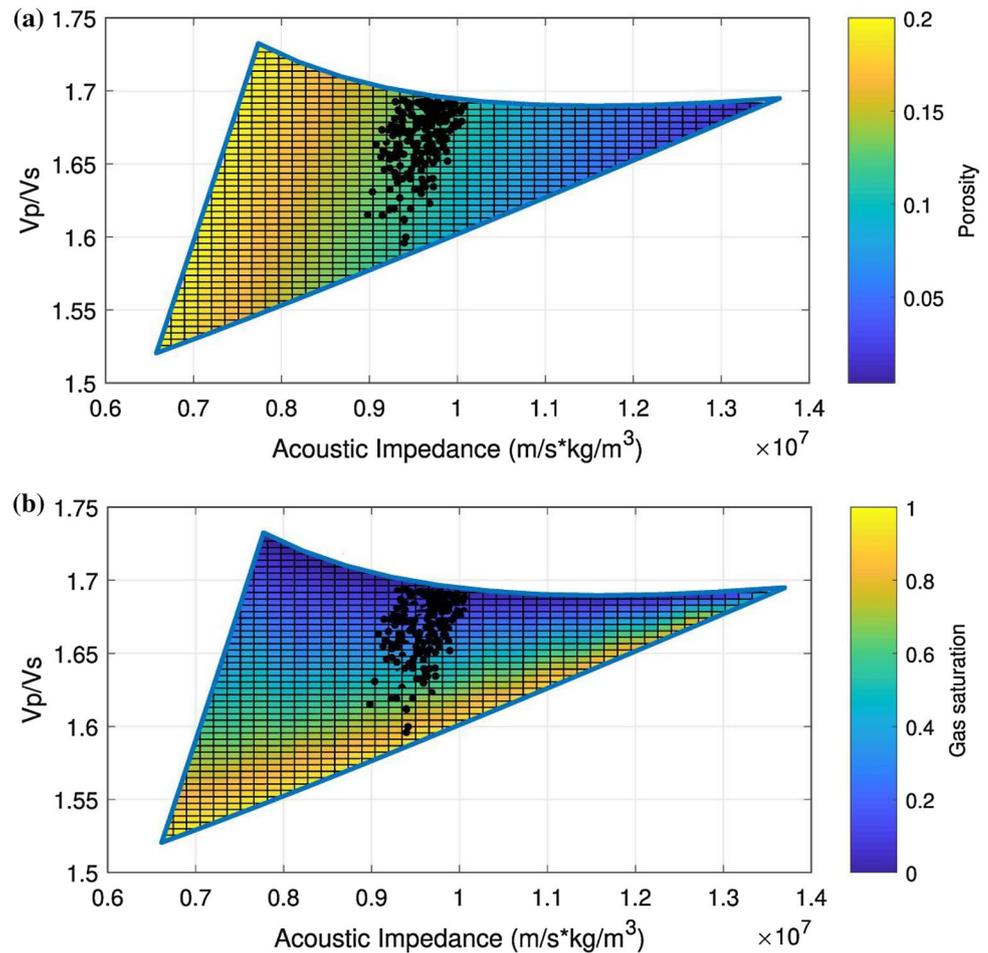
Figure 19 displays the projection of the acoustic impedance and Vp/Vs data pairs onto the seismic-scale RPT. These projection points are relatively concentrated due to the limited frequency bandwidth of the seismic inversion data. Points that fall in the reservoir zone are defined as sandstone reservoirs. In addition, based on the position of these projection data points, we can directly transform the acoustic impedance and Vp/Vs data pairs into the porosity and gas saturation data pairs. The porosity and gas saturation of points that fall outside the reservoir zone are set to zero. That is to say, we only focus on the porosity and gas saturation of sandstone reservoirs and ignore that of shale or siltstone. Figure 20 shows the prediction

results of the porosity and gas saturation. Note that the predicted porosity and gas saturation are in good agreement with the well-log data. These results demonstrate the validity of the improved rock-physics model and can be used as a guidance for the further deep-buried gas reservoir exploration in this area.

Discussion

A practical workflow for rock-physics modelling of deep-buried sandstone gas reservoirs with complex pore structures has been proposed. For clay–sand mixture layers, the Xu–White model can estimate both P- and S-wave

Fig. 19 Reservoir zone in seismic-scale RPT colour-coded by template-indicated porosity (a) and gas saturation (b); the projected black points represent acoustic impedance and the V_p/V_s data pairs from the seismic inversion



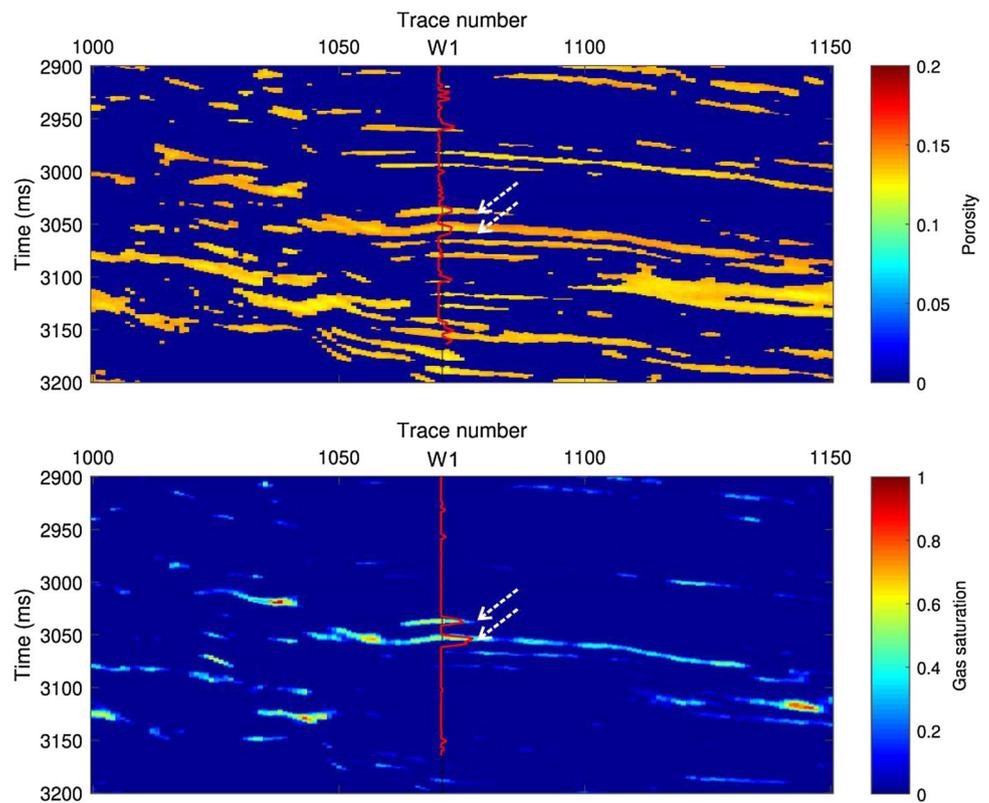
velocities over the full range of siliciclastic rocks from clean sandstones to pure shales, whether consolidated or moderately unconsolidated (Xu and White 1996). We extend the Xu–White model, originally accounts for two pore types (wet clay soft pore and quartz intergranular semi-stiff pores), to our case by introducing a third feldspar-related dissolution stiff pore. Our extended Xu–White model consists of three steps: (1) rock matrix modelling using time-average equations; (2) dry rock modelling using a multi-pore analytical approximation; and (3) fluid-saturated rock modelling using patchy saturation.

The input parameters required by the extended Xu–White model include the elastic moduli and densities of mineral particles and pore fluids, the porosity and shale volume of the rock frame, and three pore aspect ratios for three types of pores. For mixed minerals, the elastic moduli and density of clays are variable and not very well known, leading to uncertainties in the rock matrix modelling. The estimation of these parameters relies on careful calibration of well logs through trial-and-error experiments. In practice, the parameters calibration process has certain rules to follow; that is, the mineral moduli and density control the overall adjustment

of the predicted P- and S-wave velocity and density curves, while the pore aspect ratios act as a local adjustment, in which larger pore aspect ratios correspond to higher values of predicted velocities. Utilizing these rules can make the calibration process easier. When the predicted curves match well with the measured logs, we can determine these input rock-physics modelling parameters.

In dry rock modelling, by assuming a constant dry rock Poisson's ratio, we decouple the differential effective medium equations and then derive a multi-pore analytical approximation for the dry rock moduli calculation. Despite this assumption, we note that the Poisson's ratio derived from the multi-pore dry rock approximation in Eqs. (18) and (19) is not constant. In fact, the constant assumption is not a rigorous restriction to the dry rock Poisson's ratio, because we only apply this assumption in the calculation of geometry parameters P_i and Q_i , and the final expressions of dry rock moduli are still related to porosity. In addition, to force the dry rock Poisson's ratio to be constant will result in a poor approximation for the shear-wave velocity (Keys and Xu 2002). On the other hand, for rock-physics templates in Fig. 11, note that V_p/V_s varies with porosity, especially for

Fig. 20 Prediction results of porosity (a) and gas saturation (b) for deep-buried sandstone reservoirs with corresponding porosity and gas saturation log curves inserted



the gas-saturated sandstone model. However, also note that the increase in clay content will cause a more drastic variation of V_p/V_s , because the rock matrix is more sensitive to the clay content variation than the high-porosity rocks (Avseth et al. 2010). This means that the variation degree of V_p/V_s depends on the clay content of the rock matrix. Therefore, when sandstone contains clay, the dry rock Poisson's ratio will change with porosity, which also indicates that the dry rock Poisson's should not be forced to be constant.

The clay content in sandstone is also an important parameter related to the scale effect, reservoir heterogeneity and saturation distribution (Jakobsen et al. 2003b). For thin-interbedded sand and shale layers, upscaling means an increase in the intercalating shale, which, in turn, increases reservoir heterogeneity and thus affects the fluid saturation distribution pattern. At the well-log scale, the reservoir sandstone data in the studied wells match nicely with the template when the clay content is 0.95, which is basically consistent with the shale volume log curve values at sandstone locations. When upscaling to the seismic scale using the Backus average, the P- and S-wave velocity and density curves become smooth and concentrated, and the best-fit clay content value for the template changes. A different upscaling frequency corresponds to a different best-fit template, which leads to uncertainties in seismic interpretation. In our case, we set the upscaling frequency based on the frequency band of seismic inversion results. Both the

observations at the well-log scale and the seismic scale are consistent with a patchy-saturation behaviour. One possible explanation is that the heterogeneities caused by clay content in sandstone control the saturation pattern, even in seismic scale, the seismic wave may experience patchy saturation when the thickness of clay-sand mixture layer is relatively large (Avseth et al. 2010).

In addition to the two main uncertainties associated with the above-mentioned clay properties, the extended Xu-White model is also limited by several basic assumptions inherited from the DEM model such as the isotropic, linear, and elastic rock and idealized ellipsoidal pore geometries (Mavko et al. 2009). Besides, we have not considered uncertainties caused by shale-related anisotropy and fluid-flow-related dispersion and attenuation, which will be investigated in future research to obtain a better model. Despite these uncertainties, we have found that the extended Xu-White model is useful for modelling deep-buried sandstone reservoirs, and it can also be applied to other types of rocks which have complex pore structures, such as carbonate rock.

Conclusions

We have proposed an improved rock-physics model for deep-buried sandstone reservoirs based on the geological characteristics analysis. This model can handle complex pore structure rocks with high computational efficiency compared to the differential effective medium theory. It has been used in total porosity estimation and S-wave velocity prediction; the results show that our model is more accurate than the density method and the conventional Xu–White model, which considers only two pore types. The rock-physics template generated by the extended Xu–White model provides a very useful tool for the porosity and gas saturation prediction of deep-buried gas reservoirs. The template needs to be carefully calibrated and validated by well-log data while honouring local geologic factors. The reliability of the template depends on the input data quality and the model assumptions. Moreover, the scale effects on the template should be considered when we move from the well-log scale to the seismic scale. Nevertheless, the established rock-physics template shows good performance in the porosity and saturation prediction of gas reservoirs and can help reduce risk in the deep seismic exploration and prospect evaluation.

Acknowledgements This work was financially supported by the Natural Science Foundation of China (41674139) and the National Science and Technology Major Project (2016ZX05033-02). We thank the Sinopec Shanghai branch for data provided and permission to publish the results of this research. Yaneng Luo would also like to thank the Basin and Reservoir Studies (BRS) group at University of Bergen for hospitality and support during his one-year visit.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Aki K, Richards PG (2002) Quantitative seismology, 2nd edn. University Science Books, California
- Avseth P, Mukerji T, Mavko G (2005) Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk. Cambridge University Press, New York
- Avseth P, Mukerji T, Mavko G, Dvorkin J (2010) Rock-physics diagnostics of depositional texture, diagenetic alterations, and reservoir heterogeneity in high-porosity siliciclastic sediments and rocks—a review of selected models and suggested work flows. *Geophysics* 75(5):A31–A47
- Avseth P, Johansen TA, Bakhorji A, Mustafa HM (2014) Rock-physics modelling guided by depositional and burial history in low-to-intermediate-porosity sandstones. *Geophysics* 79(2):D115–D121
- Bachrach R, Avseth P (2008) Rock physics modeling of unconsolidated sands: accounting for nonuniform contacts and heterogeneous stress fields in the effective media approximation with applications to hydrocarbon exploration. *Geophysics* 73(6):E197–E209
- Bai J, Yue C, Liang Y, Song Z, Ling S (2013) Variable aspect ratio method in the Xu–White model for shear-wave velocity estimation. *J Geophys Eng* 10:1–6
- Batzle M, Wang Z (1992) Seismic properties of pore fluids. *Geophysics* 57:1396–1408
- Berge PA, Berryman JG, Bonner BP (1993) Influence of microstructure on rock elastic properties. *Geophys Res Lett* 20:2619–2622
- Berryman JG (1992) Single-scattering approximations for coefficients in Biot's equations of poroelasticity. *J Acoust Soc Am* 91:551–571
- Berryman JG, Pride SR, Wang HF (2002) A differential scheme for elastic properties of rocks with dry or saturated cracks. *Geophys J Int* 151:597–611
- Best A (2014) Physics of rocks for hydrocarbon exploration: introduction. *Geophys Prospect* 62:1203–1204
- Bosch M, Carvajal C, Rodrigues J, Torres A, Aldana M, Sierra JS (2009) Petrophysical seismic inversion conditioned to well-log data: methods and application to a gas reservoir. *Geophysics* 74(2):O1–O15
- Buland A, Omre H (2003) Bayesian linearized AVO inversion. *Geophysics* 68:185–198
- Buland A, Kolbjørnsen O, Hauge R, Skjæveland Ø, Duffaut K (2008) Bayesian lithology and fluid prediction from seismic prestack data. *Geophysics* 73(3):C13–C21
- Cao Q, Zhou W, Liu Y, Chen W, Ji A, Lu J, Wang Y (2017) Characteristics and origin of deep high-porosity zones in slope of Xihu Sag. *J Cent South Uni (Sci Technol)* 48:751–760
- Cheng C, Toksöz MN (1979) Inversion of seismic velocities for the pore aspect-ratio spectrum of a rock. *J Geophys Res* 84:7533–7543
- Chi X, Han D (2009) Lithology and fluid differentiation using rock physics template. *Lead Edge* 28:60–65
- Dutton SP, Loucks RD (2010) Diagenetic controls on evolution of porosity and permeability in lower Tertiary Wilcox sandstones from shallow to ultradeep (200–6700 m) burial, Gulf of Mexico Basin, U.S.A. *Mar Pet Geol* 27:69–81
- Dvorkin J, Nur A (1996) Elasticity of high-porosity sandstones: theory for two North Sea data sets. *Geophysics* 61:1363–1370
- Dyman TS, Wyman RE, Kuuskraa VA, Lewan MD, Cook TA (2003) Deep natural gas resources. *Nat Resour Res* 12:41–56
- Guéguen Y, Palciauskas V (1994) Introduction to the physics of rocks. Princeton University Press, Chichester
- Hu M, Shen J, Hu D (2013) Reservoir characteristics and its main controlling factors of the Pinghu Formation in Pinghu structural belt, Xihu depression. *Oil Gas Geol* 34:185–191
- Jakobsen M, Hudson J, Johansen TA (2003a) T-matrix approach to shale acoustics. *Geophys J Int* 154:533–558
- Jakobsen M, Johansen TA, McCann C (2003b) The acoustic signature of fluid flow in complex porous media. *J Appl Geophys* 54:219–246
- Johansen TA, Drottning Lecomte I, Ystidal HG (2002) An approach to combined rock physics and seismic modelling of fluid substitution effects. *Geophys Prospect* 50:119–137
- Keys RG, Xu S (2002) An approximation for the Xu–White velocity model. *Geophysics* 67:1406–1414
- Keys R, Matava T, Foster D, Ashabranner D (2017) Isotropic and anisotropic velocity model building for subsalt seismic imaging. *Geophysics* 82(3):S247–S258
- Kuster GT, Toksöz MN (1974) Velocity and attenuation of seismic waves in two-phase media. *Geophysics* 39:587–618
- Lai J, Wang G, Fan Z, Chen J, Wang S, Zhou Z, Fan X (2016) Insight into the pore structure of tight sandstones using NMR and HPMI measurements. *Energy Fuels* 30:10200–10214
- Lai J, Wang G, Chai Y, Xin Y, Wu Q, Zhang X, Sun Y (2017) Deep burial diagenesis and reservoir quality evolution of high-temperature, high-pressure sandstones: examples from Lower Cretaceous

- Bashijiqike Formation in Keshen area, Kuqa depression, Tarim basin of China. *AAPG Bull* 101(6):829–862
- Lai J, Wang G, Wang S, Cao J, Li M, Pang X, Zhou Z, Fan X, Dai Q, Yang L, He Z, Qin Z (2018) Review of diagenetic facies in tight sandstones: diagenesis, diagenetic minerals, and prediction via well logs. *Earth Sci Rev* 185:234–258
- Li H, Zhang J (2012) Analytical approximations of bulk and shear moduli for dry rock based on the differential effective medium theory. *Geophys Prospect* 60:281–292
- Lindsay R, Van Koughnet R (2001) Sequential Backus averaging: upscaling well logs to seismic wavelengths. *Lead Edge* 20:188–191
- Mavko G, Mukerji T, Dvorkin J (2009) *The rock physics handbook tools for seismic analysis of porous media*, 2nd edn. Cambridge University Press, New York
- Morad S, Al-Ramadan K, Ketzer JM, DeRos LF (2010) The impact of diagenesis on the heterogeneity of sandstone reservoirs: a review of the role of depositional facies and sequence stratigraphy. *AAPG Bull* 94(8):1267–1309
- Müller TM, Gurevich B, Lebedev M (2010) Seismic wave attenuation and dispersion resulting from wave-induced flow in porous rocks—a review. *Geophysics* 75(5):A147–A164
- Norris AN (1985) A differential scheme for the effective moduli of composites. *Mech Mater* 4:1–16
- Ødegaard E, Avseth P (2004) Well log and seismic data analysis using rock physics templates. *First Break* 23:37–43
- Pang X, Jia C, Wang W (2015) Petroleum geology features and research developments of hydrocarbon accumulation in deep petroliferous basins. *Pet Sci* 12:1–53
- Rezaee R, Saeedi A, Clennell B (2012) Tight gas sands permeability estimation from mercury injection capillary pressure and nuclear magnetic resonance data. *J Petrol Sci Eng* 88–89:92–99
- Ruiz F, Dvorkin J (2009) Sediment with porous grains: rock-physics model and application to marine carbonate and opal. *Geophysics* 74(1):E1–E15
- Sun L, Zou C, Zhu R, Zhang Y, Zhang S, Zhang B, Zhu G, Gao Z (2013) Formation, distribution and potential of deep hydrocarbon resources in China. *Pet Explor Dev* 40:687–695
- Vernik L, Kachanov M (2010) Modelling elastic properties of siliciclastic rocks. *Geophysics* 75(6):E171–E182
- Wang S, Yuan S, Wang T, Gao J, Li S (2018) Three-dimensional geosteering coherence attributes for deep-formation discontinuity detection. *Geophysics* 83(6):O105–O113
- White JE (1975) Computed seismic speeds and attenuation in rocks with partial gas saturation. *Geophysics* 40:224–232
- Wu J, Zhang L, Wan L, Zhao Q, Yang C, Wang Y (2017) Provenance analysis of Pinghu Formation in Xihu sag. *China Pet Explor* 22:50–57
- Wyllie MRJ, Gregory AR, Gardner LW (1956) Elastic wave velocities in heterogeneous and porous media. *Geophysics* 21:41–70
- Xu S, Payne MA (2009) Modelling elastic properties in carbonate rocks. *Lead Edge* 28:66–74
- Xu S, White RE (1995) A new velocity model for clays and mixtures. *Geophys Prospect* 43:91–118
- Xu S, White RE (1996) A physical model for shear-wave velocity prediction. *Geophys Prospect* 44:687–717
- Yuan S, Ji Y, Shi P, Zeng J, Gao J, Wang S (2019a) Sparse Bayesian learning-based seismic high-resolution time-frequency analysis. *IEEE Geosci Remote Sens Lett*. <https://doi.org/10.1109/lgrs.2018.2883496>
- Yuan S, Liu Y, Zhang Z, Luo C, Wang S (2019b) Prestack stochastic frequency-dependent velocity inversion with rock-physics constraints and statistical associated hydrocarbon attributes. *IEEE Geosci Remote Sens Lett* 16:140–144
- Zhao L, Nasser M, Han D (2013) Quantitative geophysical pore-type characterization and its geological implication in carbonate reservoirs. *Geophys Prospect* 61:827–841
- Zimmerman RW (1985) The effect of microcracks on the elastic moduli of brittle materials. *J Mater Sci Lett* 4:1457–1460
- Zou C, Zhu R, Liu K, Su L, Bai B, Zhang X, Yuan X, Wang J (2012) Tight gas sandstone reservoirs in China: characteristics and recognition criteria. *J Pet Sci Eng* 88–89:82–91



The influence of the frequency-dependent spherical-wave effect on the near-surface attenuation estimation

Yongzhen Ji^{1,2} · Shangxu Wang¹ · Sanyi Yuan¹ · Binpeng Yan¹

Received: 6 September 2018 / Accepted: 4 March 2019 / Published online: 6 March 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

The knowledge of Q is desirable for improving seismic resolution, facilitating amplitude analysis and seismic interpretation. The most commonly used methods for Q estimation are the frequency-spectrum-based methods. Generally, these methods are based on the plane wave theory assuming that the transmission/reflection loss is frequency independent. This assumption is reasonable in the far-field situation and makes the transmission/reflection coefficient irrelevant with the Q estimation result. However, in the near-surface context, this assumption is invalid because the seismic wave propagates in the form of spherical wave in the real seismic surveys and the spherical-wave transmission/reflection coefficient is frequency dependent. As a result, deviation will exist. In this paper, the influence of the spherical-wave effect on the Q estimation in the near-surface context was proved in both synthetic data and field data for the first time, and it was found that the deviation due to the spherical-wave effect is of order comparable to the intrinsic attenuation. The compensation method based on the forward modeling is then proposed to correct this deviation, and the effectiveness of the proposed method is proved by the reasonable estimated results of both synthetic data and field data example. These results raise caution for the interpretation of the extracted Q in the near-surface context if they do not account for the spherical-wave effect and point to the necessity of incorporating a frequency-dependent term in the frequency-spectrum-based method when applied to the Q estimation in the near surface.

Keywords Spherical-wave effect · Attenuation estimation · Near-field · Frequency-spectrum-based method

Introduction

Estimating the anelastic attenuation that is generally denoted by the quality factor Q is playing an important role in seismic imaging, data enhancement, reservoir characterization and other applications (Yuan et al. 2017). In general, Q is assumed to be frequency independent (Dasgupta

and Clark 1998; Zhang and Ulrych 2002; Li et al. 2015). The most commonly used methods for the Q estimation are the frequency-spectrum-based methods, such as the spectral ratio (SR) method (Tonn 1991), the frequency shifting methods (Quan and Harris 1997; Zhang and Ulrych 2002), the time–frequency methods (Hao et al. 2016) and the reflectivity spectrum method (Li et al. 2016a, b, c). These methods are based on the plane wave theory and assume that the transmission/reflection coefficient is frequency independent. This assumption makes the transmission/reflection coefficient irrelevant with the Q estimation result. As a far-field (when the propagation distance is very long or the frequency is very high, i.e., when KR is very large, where K is the wavenumber and R is the propagation distance) approximation of the spherical-wave coefficient, the plane wave transmission/reflection coefficient is indeed frequency independent (Li et al. 2017). As a result, these methods have been successfully applied on numerous field data sets, as the far-field approximation is valid in most seismic exploration contexts (Dasgupta and Clark 1998; Matsushima 2006; Matsushima et al. 2015). However, in

✉ Yongzhen Ji
jiyz232@sina.com
Shangxu Wang
wangsx@cup.edu.cn
Sanyi Yuan
yuansy@cup.edu.cn
Binpeng Yan
yanbp_cupb@163.com

¹ State Key Laboratory of Petroleum Resources and Prospecting, CNPC Key Laboratory of Geophysical Exploration, China University of Petroleum (Beijing), Changping, Beijing, China

² Sinopec Geophysical Research Institute, Nanjing, China

the near-field context, the seismic wave propagates in the form of spherical wave and the spherical-wave transmission/reflection coefficient is frequency dependent (Alulawi and Gurevich 2013; Li et al. 2017). Consequently, it is reasonable to infer that because of the difference between the frequency-dependent spherical-wave coefficient and the plane wave coefficient, deviation will exist in the Q estimation results derived by frequency-spectrum-based methods. Although many researchers have investigated the frequency-dependent spherical-wave coefficient in different aspects, for example, Li et al. (2017) addressed such issue for amplitude versus offset analysis, Wehner et al. (2018) investigated frequency-depth-dependent spherical reflection response from the sea surface, no attempt has been reported to investigate the influence on the attenuation estimation in the field data and to correct it.

In general, VSP (vertical seismic profile) data and crosswell data are used to investigate the near-surface Q , since the VSP measurement (including zero offset, offset, walkaway and reverse VSP) is considered to be best suited for attenuation studies and it enables sampling of the wavefield at various known depths (Matsushima 2006). To straightforward demonstrate the influence of the spherical-wave effect and minimize the influence of irrelevant factors, simple reverse VSP data with one interface are selected and synthetic data with similar geometry are designed to support the field data analysis. Because a similar geometry is used in both synthetic data and field data, the results can prove and confirm each other and keep consistency. In both synthetic data and field data, firstly, we demonstrated the influence of the spherical-wave effect in the near-field context: in synthetic data test, the deviation was proved based on the theoretical deduction, and in the practical VSP data test, the deviation was proved by a simple but novel experiment; secondly, we utilized the forward modeling-based compensation method to correct the deviation and derived the reasonable Q estimation.

Note that the influence of the spherical-wave effect on the attenuation estimation includes the frequency dependency of the spherical-wave transmission and reflection coefficient. However, in this paper, according to the available field data, we concretized the influence of the spherical-wave effect as the influence of the frequency-dependent spherical-wave reflection coefficient (FSRC) and adopted the reflection wave to conduct the attenuation analysis. Actually, using reflection wave to estimate Q in the near-surface context has several advantages: firstly, in the near-surface context, comparing first arrival traces that are too close together may lead to poor Q estimate and the reflection wave is an alternative to have more reliable estimation; secondly, the reflection wave can generally offer more useful areal coverage; thirdly, the reflection wave data in the VSP data are similar to the surface seismic data—a data type commonly used in the

seismic exploration, so the studies of the attenuation estimation using near-surface reflection wave is instructive and applicable to the attenuation estimation using surface seismic data (Montano 2015; Beckwith et al. 2016).

The reminder of the paper is organized in the following way: Firstly, we briefly reviewed the theory of the frequency-spectrum-based method and the FSRC; secondly, we proved and compensated the deviation caused by the FSRC in the Q estimation in the synthetic data and field data; thirdly, we discussed the potential difficulties of application and the further research directions; finally, some conclusions were drawn.

Theory

The frequency-spectrum-based method

There are many attenuation estimation methods, however, measurement of attenuation using frequency-spectrum-based methods may be a more convenient and stable procedure (Matsushima et al. 2015). The SR method is one of the most commonly used methods, and in this paper, we utilized SR method on behalf of these methods. The SR method is based on that the logarithm of the spectral amplitude of a seismic wave decays at a rate roughly proportional to the frequency. Seismic wave attenuation in the media of assumed constant Q can be modeled as the damped oscillating system as:

$$A(f, x_2) = B \cdot A(f, x_1) e^{-\pi f \Delta t / Q} \quad (1)$$

where $A(f, x)$ is the amplitude spectrum at the frequency f with the travel distance x , and Δt is the difference in receive time between the two waveforms with different travel distances. B is a parameter that takes into account the geometry function and frequency-independent factors, including the transmission/reflection coefficient (Li et al. 2016a, b, c). Then, take the logarithm, Eq. (1) can be rewritten as:

$$\ln \left[\frac{A(f, x_2)}{A(f, x_1)} \right] = -\pi f \Delta t / Q + \ln(B) \quad (2)$$

A linear regression is involved to estimate the slope of the left-hand side. The slope is equal to $-\pi \Delta t / Q$ (Matsushima 2006). Finally, the Q can be estimated from the slope.

The frequency-dependent spherical-wave reflection coefficient (FSRC)

Recently, researches reveal that, in near-field context, the reflection wave coefficient is frequency dependent (Tao et al. 2016; Li et al. 2017). Note that near-field-associated frequency-dependent reflectivity due to the sphere wave effect

is different from hydrocarbons-associated frequency-dependent reflectivity due to the fluid flow in the mesoscopic scale range (Yuan et al. 2019). Such spherical-wave frequency-dependent effect will affect the Q estimation related to the fluid flow, because if the reflection coefficient is frequency dependent, B in Eq. (2) will be frequency dependent. As a result, the linear regression will lead to an incorrect result.

The most important and fundamental methods to derive the FSRC are the Sommerfeld integral (Sommerfeld 1909) and the Weyl integral (Weyl 1919); the basic concepts of these methods are to decompose spherical waves into the superpositions of cylindrical waves and plane waves, respectively. In this paper, we achieved the FSRC based on the Sommerfeld integral (Li et al. 2016a, b, c; Li et al. 2017). For two different homogeneous semi-infinite acoustic media in contact with a plane, the frequency-dependent reflection coefficient can be represented as:

$$R_{pp}^{sph}(\omega, \varphi) = \frac{-i\omega \frac{1}{\alpha_1} \int_1^0 B(x) J_0\left(\omega r \frac{\sqrt{1-x^2}}{\alpha_1}\right) e^{i\omega \frac{x}{\alpha_1}(h+z_r)} dx - i^2 \omega \frac{1}{\alpha_1} \int_0^\infty B(ix) J_0\left(\omega r \frac{\sqrt{1+x^2}}{\alpha_1}\right) e^{-\omega \frac{x}{\alpha_1}(h+z_r)} dx}{e^{i\omega R/\alpha_1}/R}, \tag{3}$$

where $B(x) = \frac{\rho_2/\alpha_1 x - \rho_1/\alpha_2 \sqrt{1-(1-x^2)\alpha_2^2/\alpha_1^2}}{\rho_2/\alpha_1 x + \rho_1/\alpha_2 \sqrt{1-(1-x^2)\alpha_2^2/\alpha_1^2}}$ is the plane wave reflection coefficient (PRC) corresponding to the incident angle φ , $x = \cos(\theta)$ is the integral variable, i is the imaginary unit, ω is the angular frequency, J_0 is the zero-order Bessel function, r is the offset, h is the distance from the source to the interface, z_r is the distance from the receiver to the interface, α_1 and α_2 are the velocity of upper and lower media, respectively and ρ_1 and ρ_2 are the density of the upper and lower media, respectively. In anelastic situations (finite Q), velocities are frequency dependent and complex (Aki and Richards 1980): $\alpha(\omega) = \alpha_{ref} [1 + \ln(\omega_{cur}/\omega_{ref})/\pi Q - i/2Q]$, where ω_{cur} and ω_{ref} are the angular frequencies of current component and reference component.

As a result, for the near-surface context, B in Eq. (2) should be replaced by $B' \cdot R_{pp}^{sph}(\omega, \theta)$ and Eqs. (1) and (2) should be rewritten as:

$$A(f, x_2) = B' \cdot R_{pp}^{sph}(\omega, \theta) \cdot A(f, x_1) e^{-\pi f \Delta t / Q} \tag{4}$$

$$\ln \left[\frac{A(f, x_2)}{A(f, x_1) \cdot R_{pp}^{sph}(\omega, \theta)} \right] = -\pi f \Delta t / Q + \ln(B') \tag{5}$$

where the $R_{pp}^{sph}(\omega, \theta)$ represents the FSRC and B' represents the frequency-independent term excluding the transmission/reflection coefficient. As a result, the FSRC obtained by the forward modeling is added into the modified SR method,

and a linear regression is involved to estimate the slope of the left-hand side of Eq. (5) to obtain the estimation of Q .

Deviation due to the FSRC: synthetic data test

To show the deviation due to the FSRC, we designed a simple geometry synthetic data. The test was conducted for one model with a flat horizontal interface between two homogeneous isotropic half-spaces, as shown in Fig. 1. The upper half-space is described by the parameters $\alpha_1 = 1000$ m/s and $\rho_1 = 2000$ kg/m³, and the lower half-space is described by the parameters $\alpha_2 = 1600$ m/s, $\rho_2 = 2500$ kg/m³ and $Q=15$. The source is 10 m below the interface and the receivers are 40 m below the interface. The offset is 4 m. Note that the geometry and the parameter of the synthetic model were

designed approximately according to the field data to keep consistency. We exploited a Ricker wavelet of 50 Hz as the reference wavelet. Based on the convolution model (Robinson 1984), we first computed the original wavelet spectrum using Fourier transform and the reflection spectrum using the product of the original spectrum and the corresponding reflectivity spectrum; the spectrum of the PRC is calculated using B in Eq. (3) and the spectrum of the FSRC

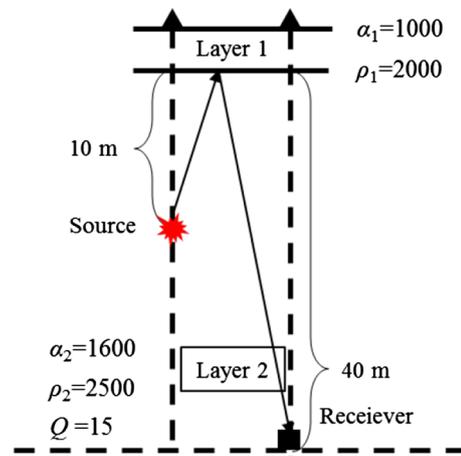


Fig. 1 Geometry of the synthetic model. The upper half-space is described by the parameters $\alpha_1 = 1000$ m/s and $\rho_1 = 2000$ kg/m³, and the lower half-space is described by the parameters $\alpha_2 = 1600$ m/s, $\rho_2 = 2500$ kg/m³ and $Q=15$. The source is 10 m below the interface and the receivers are 40 m below the interface. The offset is 4 m

is calculated using Eq. (3); Then, the reflection plane wave and the reflection spherical wave in the time domain were obtained using inverse Fourier transform. In the near-field context, the result of the convolution model using the FSRC is closer to the practical situation than that using the PRC, because such result is equivalent to the numerical simulation of the wave equation (Li et al. 2017).

Figure 2 shows the change of reflection coefficients calculated based on the spherical-wave theory and the plane wave theory with the frequency. The SRC changes with frequency, whereas the PRC remains invariant for different

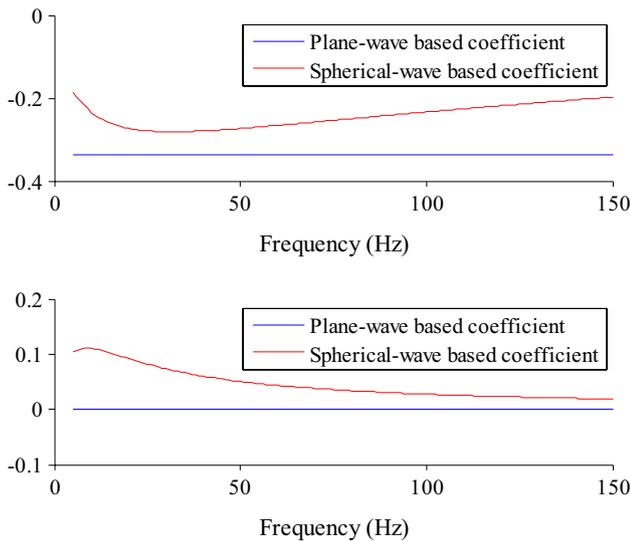


Fig. 2 Change of reflection coefficients calculated based on the spherical-wave theory and the plane wave theory with the frequency. The upper picture shows the real part of the reflection coefficients and the lower picture shows the imaginary part of the reflection coefficients

frequencies. Figure 3 shows the comparison between the reference wavelet (Fig. 3a) and the reflection plane wave and the reflection spherical wave received by the receiver (Fig. 3b). Both the phase of the reflection wave including the reflection plane wave and the reflection spherical wave shift, however, the shift of the reflection spherical wave is different from that of the plane wave. First, using the reflection spherical wave and the reference wave, based on the conventional SR method, we obtained the estimated $Q = 8.45$. As expected, the deviation exists, and the deviation is over 40%. In fact, one can predict the existence of such deviation from the frequency dependency of the coefficient as shown in Fig. 2. Then, we calculated the FSRC based on Eq. (3) and utilized it to modify the SR method based on Eq. (4). At last, a linear regression is involved to estimate the slope of the left-hand side of Eq. (5) to obtain the estimation of Q . The estimation is $Q = 15$.

The synthetic test explained and proved the source of the deviation again and it is the complement of the theoretical inference. The deviation was corrected to obtain the reasonable Q estimation result. However, the most persuasive test is the field data test.

Deviation due to the FSRC: field data

The description of the field data

The near-surface zone of interest is composed of unconsolidated soil and fine sand with a thickness of more than 40 m. Two boreholes, spaced 4 m apart, were drilled to the same depth of 40 m. The sources were fired from the bottom to the surface at an interval of 1 m in the shot hole. The receiver was placed at the bottom of the receiver hole. The near-surface zone can be divided into two layers, including

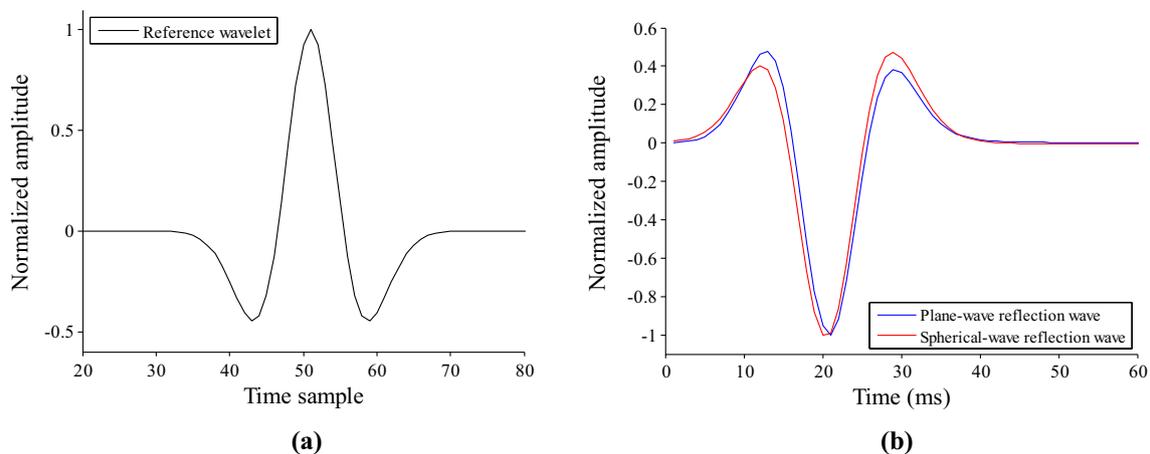


Fig. 3 Reference wavelet (a) and the reflection waveforms (b)

the weathering (layer 1) and subweathering layers (layer 2), with an interface at a depth of 2 m. Near-surface velocity model was established by the interpretation of the first breaks (Yuan et al. 2018). The velocity increases to 1618 m/s in the layer 2 from 340 m/s in the layer 1 (Li et al. 2016a, b, c). The geometry of the field data was shown in Fig. 4.

The idea of the experiment using the field data

The data are simple but valuable to eliminate other influencing factors, as shown in the following analysis, and focus on the spherical-wave frequency-dependent effect. To best utilize the data and prove our idea, we decided to conduct a novel experiment. First, according to the assumption of the conventional SR method, the reflection coefficient is frequency independent. As a result, it is irrelevant with the Q estimation result, which means that the Q estimation result using the reflection wave and the direct wave should be the same. As these two kinds of wavefields exist in these data, as shown in Figs. 4 and 5 where label ① represents the virtual ray path of the reflection wavefield and label ② represents the virtual ray path of the direct wavefield, we can estimate the Q using the direct wave and the reflection wave, respectively. Then, if we arrive at the same value (or similar value of less than $\pm 5\%$) for Q estimation, it is proved that there is no influencing factors about the reflection in the near-field Q estimation, and the conventional SR method is indeed suitable for near-surface context. If we arrive at different estimation results, as the direct wave experiences no reflection meeting

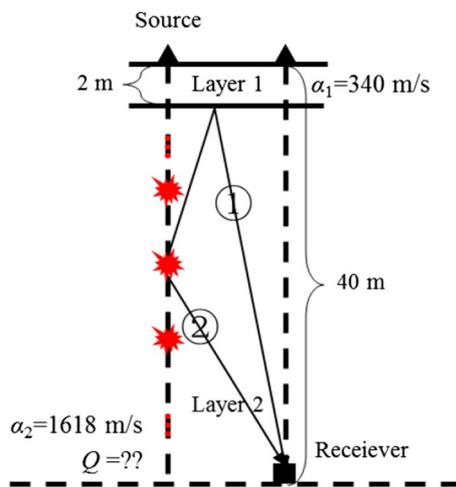


Fig. 4 Geometry of the field data. The sources were fired from the bottom to the surface at an interval of 1 m in the shot hole. The receiver was placed at the bottom of the receiver hole. The near-surface zone can be divided into two layers with an interface at a depth of 2 m. The velocity increases to 1618 m/s in the layer 2 from 340 m/s in the layer 1. Label ① represents the virtual ray path of the reflection wave and label ② represents the virtual ray path of the direct wave

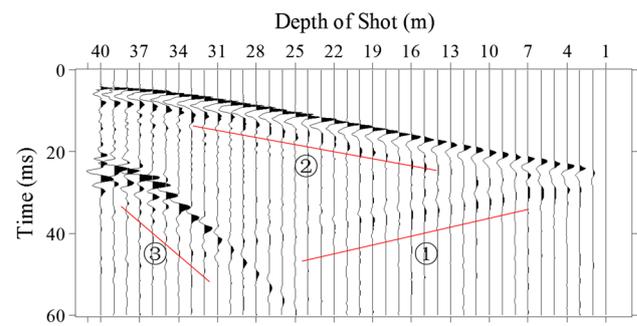


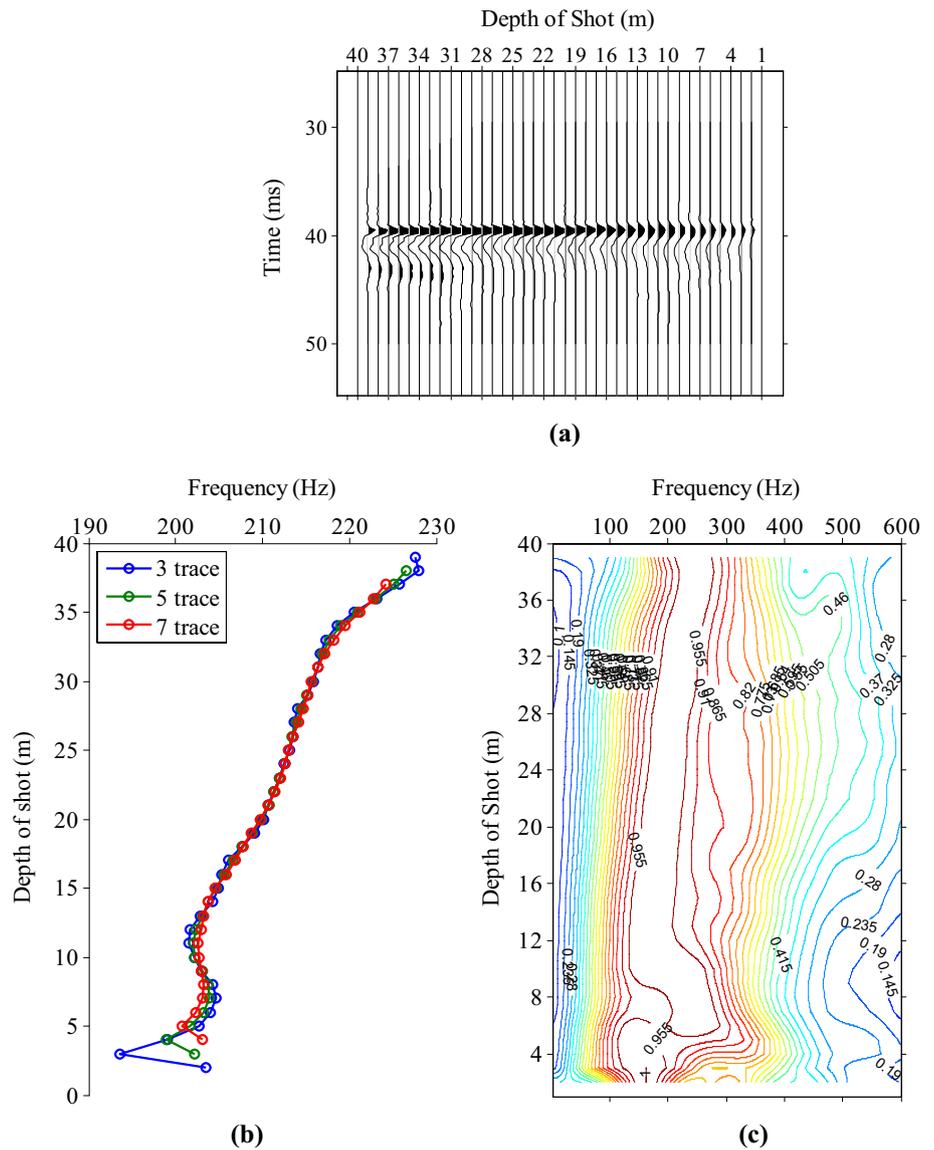
Fig. 5 Common receiver gather of the bottom receiver. Label ① represents the reflection wave, label ② represents the direct wave and label ③ represents the interference

the assumption of the SR method, the result using the direct wave can be treated as the reference; the result using the reflection wave should be corrected. Finally, we correct the reflection wave result using the proposed modified SR method, and the corrected result can be the judgment of our claim. If the result after the correction using the FSRC is similar with the reference Q estimated from the direct wave, the method considering FSRC effect is proved practicable.

Q estimation based on direct wave

To enhance the quality of the downgoing direct wave, preprocessing including a dip filtering and a spatial mix operation (Matsushima 2006) using different number of traces (one, three and five traces) was performed on the first arrivals aligned data. Then, primary downgoing waveforms were windowed with a hamming window of different window lengths (length = 20, 25 and 30 ms) centered at the first peak. The window length is estimated by averaging the length of the direct wave. Figure 6a shows the preprocessed data with three traces mix and 20 ms window. We checked the centroid frequency and relative Fourier amplitude spectrum for the quality control. Figure 6b shows the centroid frequency achieved over a bandwidth of 75–350 Hz ($f_c = \int_0^\infty fA(f)df / \int_0^\infty A(f)df$, where $A(f)$ is the spectral amplitude at frequency f) as a function of depth with different spatial mix lengths (Li et al. 2015). Figure 6b shows the 30 ms window case. Figure 6c shows the Fourier amplitude spectrum for the case of five traces mix with a 25 ms window. The anomalies appear under the depth of 10 m which may be caused by the interaction between the reflection and direct wave. The centroid frequency decreases with the increase in distance of propagation gently, and the relative Fourier amplitude spectrum also change gently. Because each trace in the common receiver gather has different sources, but the trend of centroid frequency and relative Fourier amplitude spectrum is gentle; we claim that after preprocessing the source signature is negligible. We

Fig. 6 Extracted direct wave (a), centroid frequency as a function of shot depth (b) and the relative Fourier amplitude spectrum as a function of shot depth (c). a The preprocessed data with three traces mix and 20 ms window. b The 30 ms window case. c The five traces mix with 25 ms window case



also conducted the similar quality control, including window length and number of mixed trace, and will explain the parameters in the *Q* estimation based on reflection wave section.

Because the result of *Q* estimation based on the direct wave will be the reference, in this section, three different techniques for the measurement of attenuation, SR, centroid frequency shift method and tartan triangle were applied.

The virtual ray path of the traces used for attenuation estimation is shown in Fig. 7. We used traces whose source depth is between 14 m and 36 m. The two traces used have different intervals of 1, 3 and 5 m. Figure 4 shows the 3 m case. Because we assumed that the *Q* of media is a constant, different trace pairs should give the same estimation of *Q*.

The mean value of *Q* estimation results with different intervals is shown in Table 1. The attenuation results show

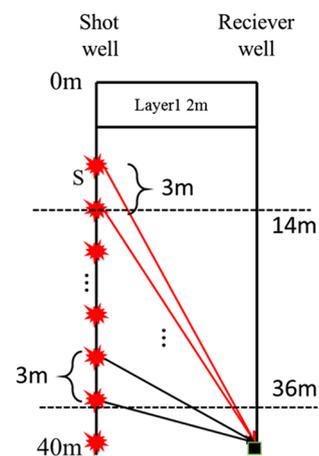


Fig. 7 Virtual ray path of the traces used for attenuation estimation with 3 m interval

Table 1 Estimated Q using different methods with different preprocessing parameters

Mean value of Q (23 pairs)	Spectral ratio method			Centroid frequency shift		
	Three traces mix	Five traces mix	Seven traces mix	Three traces mix	Five traces mix	Seven traces mix
Trace pairs with interval of 1 m						
30 ms window	17.03	12.49	12.21	20.41	13.48	13.12
25 ms window	18.17	12.89	12.59	21.24	13.81	13.45
20 ms window	20.99	13.73	13.38	22.93	14.49	14.10
Trace pairs with interval of 3 m						
30 ms window	13.19	12.70	12.82	14.16	13.59	13.86
25 ms window	13.63	13.12	13.26	14.53	13.96	14.25
20 ms window	14.58	14.00	14.17	15.30	14.69	15.00
Trace pairs with interval of 5 m						
30 ms window	13.62	14.16	14.07	15.09	16.91	16.27
25 ms window	14.18	14.85	14.69	15.06	17.82	16.97
20 ms window	15.30	16.31	15.97	16.52	19.36	18.18

that (1) the results of the centroid frequency shift method and the SR method are consistent. (The difference between results of same processing parameter is within 10%.) (2) The window length and the number of mix trace do not have too much influence on the results in the cases of 3 m interval and 5 m interval. (The average standard deviation is below 0.3.) (3) The interval of depth of shot pair of 1 m is not as good as other two sets, as comparing traces that are too close together may lead to higher estimated Q than the actual Q (Montano et al. 2015). (4) The estimated Q in the layer 2 is $\sim 13 - \sim 16$.

We employed the so-called tartan triangle (Bendeck 2017) to further comprehend the attenuation in the area. The tartan triangle (Fig. 8a) compares results from all possible shot pairs (preprocessed by 30 ms window and seven trace mix), allowing robust estimation in regions (judged according to the standard deviation on the estimated Q within a given 3×3 traces window). From the result of the standard deviation (Fig. 8b), we determine the area by the black line as the analysis area. The mean value of Q in the interest area is 15.05 and the standard deviation is 2.13. The result is consistent with the above estimation. Figure 8c is the velocity distribution derived by the ray path calculated according to the geometry and the picked arrival time. From Fig. 8b, c, we observed that the velocity in the layer 2 is almost constant and the standard deviation of the obtained Q , most of which are under 1.5, is small. As a result, we claim that the heterogeneity is small in the layer 2 and the scattering effect will be small enough to be negligible.

Here, brief conclusion about the Q estimation based on the downgoing direct wave is given. Three methods, including SR, centroid frequency shift and tartan triangle, have been used for the estimation of Q using the direct wave. The results derived under different preprocessing parameters showing the high quality and the low heterogeneity of the

data. The layer 2 has an estimated Q of $\sim 13-16$. The estimation of Q using the direct wave is well enough to be the reference.

Q estimation based on reflection wave

Figure 9 shows the virtual ray path of the traces used for attenuation estimation. Figure 10a shows the extracted reflection waveforms with the preprocessing parameter of three traces mix and 25 ms window. Figure 10b shows the relative Fourier amplitude spectrum of the extracted reflection waveforms. We observed that the reflection waveforms of depth of 40–21 m are distorted by the strong waveforms (see label © in Fig. 5) and the reflection waveforms of depth of 1–12 m are distorted by the interaction with the direct wave. The amplitude spectrum of such reflection waveforms is also distorted. As a result, the reflection waveforms whose source depth is from 19 m to 13 m are utilized for analysis. Trying to avoid too close comparison of waveforms, we picked the waveform pairs of 19 m with 17 m, 16 m, 15 m; 18 m with 16 m, 15 m, 17 m with 15 m.

Figure 11 shows the calculated $R_{pp}^{sph}(\omega, \theta)$ of different shot depths using the velocity of 1618 m/s, an estimated density ratio of 0.8 and an estimated Q of 13. Figure 12 shows the frequency spectrum of the two selected waveform pairs and their corrected frequency spectrum based on the calculated $R_{pp}^{sph}(\omega, \theta)$. Figure 12a shows the 19 m with 15 m case, and Fig. 12b shows the 19 m with 15 m. The results of Q estimation using the corrected (based on Eqs. 4 and 5) and uncorrected (based on Eqs. 1 and 2) spectrum with frequency band of 75–275 Hz are shown in Table 2. The mean value of the estimated Q obtained by the uncorrected waveform pairs deviates from the reference Q by over $\sim 50\%$, and the mean value of the estimated Q obtained by the corrected waveform pairs is within the reference range of Q value. The incident

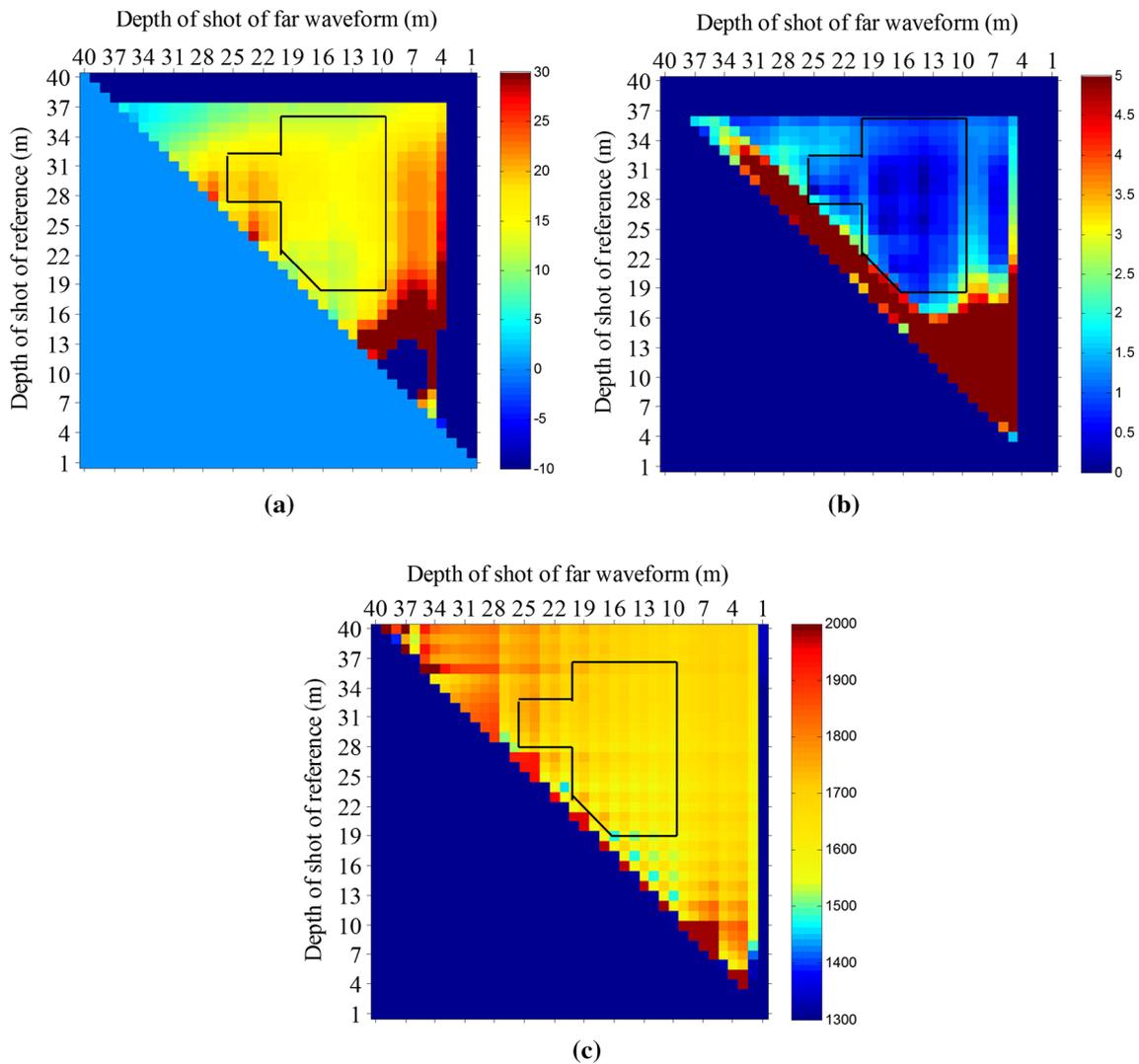


Fig. 8 Tartan triangle of the estimated Q (a), the standard deviation of estimated Q in a 3×3 traces window (b) and the velocity analysis (c)

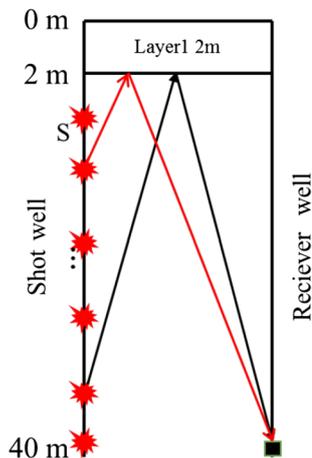


Fig. 9 Virtual ray path of the reflection wave used for attenuation estimation

angle of the reflection wave we used in the estimation is very small (within 5 degrees). As a result, although we use acoustic approximation, the result is reasonable.

Discussion

To improve the intrinsic attenuation estimation, many influencing factors were discussed by researchers, including the frequency dependency of Q , the scattering effect, and the near-field effect, etc., (Matsushima 2006; Haase and Stewart 2010; Mangriotis et al. 2011; Gurevich and Pevzner 2015), except for the influence of the spherical-wave effect. The confirmation of the deviation in the Q estimation caused by the FSRC is notably important to improve the understanding of influencing factors of the intrinsic attenuation estimation. In order to investigate the spherical-wave

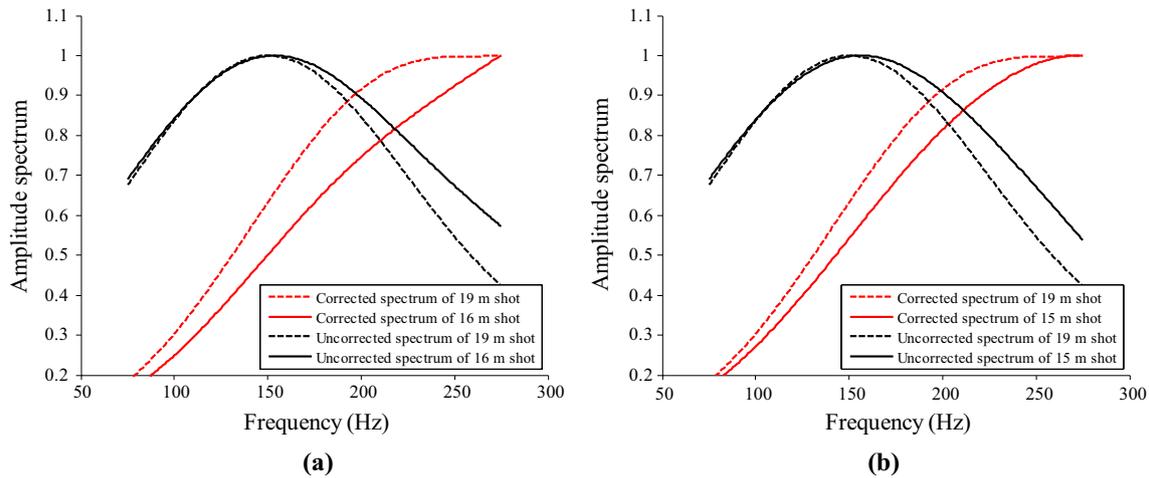


Fig. 12 Uncorrected and corrected spectrum of two selected waveform pairs. **a** is the 19 m with 16 m case; **b** is the 19 m with 15 m case

Table 2 Result of Q estimation using the corrected and uncorrected waveforms

Shot depth (m)	Corrected			Uncorrected		
	19	18	17	19	18	17
17	16.71	–	–	6.52	–	–
16	16.34	14.17	–	6.46	6.09	–
15	15.83	15.68	13.05	6.38	6.35	5.87
Mean value	15.30			6.28		
Standard deviation	1.40			0.25		

spherical-wave reflection coefficient (FSRC) is one of the reasons causing the deviation. Based on a special-geometry reverse VSP data, comprehensive analysis and estimation of attenuation in the near-surface zone is conducted. The influence of the FSRC is proved by estimating Q not only from the direct wave but also from the reflection wave. The estimation results using different methods and preprocessing parameters are consistent making the estimated Q using the direct wave well enough to be the reference. After the frequency-dependent correction, the estimation of Q using the reflection wave arrives at a reasonable result, which proved the theory inference and the compensation method to be effective. During the analysis, other influencing factors are evaluated and eliminated, and the main influencing factor in the research area is proved to be the frequency-dependent spherical-wave effect. It was found that the apparent attenuation due to the spherical-wave effect is of order comparable to the intrinsic attenuation. These results raise caution for the interpretation of extracted Q in the near surface and point to the necessity of incorporating a frequency-dependent term in Q estimation when applied to the near surface.

Acknowledgements This work was supported by the National Natural Science Foundation of China (41674127 and 41304108), the National Key Basic Research Development Program (2013CB228600), the

Major Scientific Research Program of Petrochina Science and Technology Management Department “Comprehensive Seismic Prediction Software Development and Applications of Natural Gas” (2016B-0603).

References

- Aki KT, Richards PG (1980) Quantitative seismology: theory and methods. W. H. Freeman, New York
- Alulawi B, Gurevich B (2013) Analytic wavefront curvature correction to plane-wave reflection coefficient for a weak contrast interface. *Geophys Prospect* 61(1):53–63
- Beckwith J, Clark R, Hodgson L (2016) Estimating frequency-dependent attenuation quality factor values from prestack surface seismic data. *Geophysics* 82(1):O11–O22
- Bendeck CDC, Clark RA, Booth AD, Wills W (2017) Constant and frequency-dependent attenuation from vertical seismic profiles in fractured granite and thinly layered sediment. In: 79th EAGE conference and exhibition
- Dasgupta R, Clark RA (1998) Estimation of Q from surface seismic reflection data. *Geophysics* 63(6):2120–2128
- Gurevich B, Pevzner R (2015) How frequency dependency of Q affects spectral ratio estimates. *Geophysics* 80(2):A39–A44
- Haase AB, Stewart RR (2010) Near-field seismic effects in a homogeneous medium and their removal in vertical seismic profile attenuation estimates. *Geophys Prospect* 58:1023–1032
- Hao Y, Wen X, Zhang B, He Z, Zhang R, Zhang J (2016) Q estimation of seismic data using the generalized S-transform. *J Appl Geophys* 135:122–134

- Li F, Zhou H, Jiang N, Bi J, Marfurt KJ (2015) Q estimation from reflection seismic data for hydrocarbon detection using a modified frequency shift method. *J Geophys Eng* 12(4):577–586
- Li F, Zhou H, Zhao T, Marfurt KJ (2016a) Unconventional reservoir characterization based on spectrally corrected seismic attenuation estimation. *J Seism Explor* 25(5):447–461
- Li G, Sacchi MD, Zheng H (2016b) situ evidence for frequency dependence of near-surface Q. *Geophys J Int* 204(2):1308–1315
- Li JN, Wang SX, Dong CH, Yuan SY, Wang JB (2016c) Study on frequency-dependent characteristics of spherical-wave PP reflection coefficient. *Chin J Geophys* 59(10):3810–3819
- Li JN, Wang SX, Wang JB, Dong CH, Yuan SY (2017) Frequency-dependent spherical-wave reflection in acoustic media: analysis and inversion. *Pure appl Geophys* 174(4):1759–1778
- Mangriotis MD, Rector JW III, Herkenhoff EF (2011) Case History Effects of the near-field on shallow seismic studies. *Geophysics* 76(1):B9–B18
- Matsushima J (2006) Seismic wave attenuation in methane hydrate-bearing sediments: vertical seismic profiling data from the Nankai Trough exploratory well, offshore Tokai, central Japan. *J Geophys Res Solid Earth*. <https://doi.org/10.1029/2005JB004031>
- Matsushima J, Ali MY, Bouchaala F (2015) Seismic attenuation estimation from zero-offset VSP data using seismic interferometry. *Geophys J Int* 204(2):1288–1307
- Montano M, Lawton DC, Margrave G (2015) Near-surface Q estimation: an approach using the up-going wave-field in vertical seismic profile data. In: SEG technical program expanded abstracts, pp 5600–5604
- Quan Y, Harris JM (1997) Seismic attenuation tomography using the frequency shift method. *Geophysics* 62(3):895–905
- Robinson EA (1984) Seismic inversion and deconvolution. Geophysical Press, Marbella
- Sommerfeld A (1909) Über die Ausbreitung der Wellen in der drahtlosen Telegraphie. *Ann Phys* 333(4):665–736
- Tao YH, Wang SX, Li JN, Dong CH, Yuan SY, Chen G (2016) The effect of frequency-dependent reflection coefficient on seismic attenuation estimation. In: 78th EAGE conference and exhibition 2016
- Tonn R (1991) The determination of the seismic quality factor Q from VSP data: a comparison of different computational methods. *Geophys Prospect* 39(1):1–27
- Wehner D, Landrø M, Amundsen L, Westerdahl H (2018) Frequency-depth-dependent spherical reflection response from the sea surface—a transmission experiment. *Geophys J Int* 214(2):1206–1217
- Weyl H (1919) Ausbreitung elektromagnetischer Wellen ueber einem ebenen Leiter. *Ann Phys* 365(21):481–500
- Yuan SY, Wang SX, Ma M, Ji YZ, Deng L (2017) Sparse Bayesian learning-based time-variant deconvolution. *IEEE Trans Geosci Remote Sens* 55(11):6182–6194
- Yuan SY, Liu JW, Wang SX, Wang TY, Shi PD (2018) Seismic waveform classification and first-break picking using convolution neural networks. *IEEE Geosci Remote Sens Lett* 15(2):272–276
- Yuan SY, Liu Y, Zhang Z, Luo CM, Wang SX (2019) Prestack stochastic frequency-dependent velocity inversion with rock-physics constraints and statistical associated hydrocarbon attributes. *IEEE Geosci Remote Sens Lett* 16(1):140–144
- Zhang C, Ulrych TJ (2002) Estimation of quality factors from CMP records. *Geophysics* 67(5):1542–1547



Study on logging interpretation of coal-bed methane content based on deep learning

Parhat Zunu¹ · Xiang Min¹ · Zhang Fengwei¹

Received: 10 October 2018 / Accepted: 2 January 2019 / Published online: 22 January 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

To solve quantitative interpretation problems in coal-bed methane logging, deep learning is introduced in this study. Coal-bed methane logging data and laboratory results are used to establish a deep belief network (DBN) to compute coal-bed methane content. Network parameter effects on calculations are examined. The calculations of DBN, statistical probabilistic method and Langmuir equation are compared. Results show that, first, the precision and speed of DBN calculation should determine the restricted Boltzmann machine's quantity. Second, the hidden layer neuron quantity must align with calculation accuracy and stability. Third, the ReLU function is the best for logging data; the Sigmoid function and Linear function are second; and the Softmax function has no effect. Fourth, the cross-entropy function is superior to MSE function. Fifth, RBMs make DBN more accuracy than BPNN. Furthermore, DBN calculation accuracy and stability are better than those of statistical probabilistic method and Langmuir equation.

Keywords Coal-bed methane · Geophysics logging · Deep learning · Restricted Boltzmann machine · Coal-bed methane content

Introduction

In the process of coal-bed methane (CBM) exploitation, the interpretation and evaluation of methane content are very important for the production. To date, the most accurate method for determining CBM content is coring and experimental analysis. However, these processes are too costly to conduct on a large project. Therefore, while exploring and developing CBM, it is important to identify an economical, effective, and universal means of CBM logging quantitative interpretation. These interpretations are used to establish the correlation between the logging parameters and methane content.

Hawkins et al. (1992) adopted the Langmuir isotherm adsorption theory (Langmuir 1918) in their proposal of the Langmuir equation, which is used to calculate the CBM content. Yang et al. (2005) studied the relationship between resistivity logging and the CBM reservoir. Bhanja and Srivastava (2008) established the relationships among the

lithology density, volume density, acoustic move-out and natural gamma logging, and CBM content. Wang (2009) presented the improved Langmuir equation to calculate CBM content. The approach achieved a good application effect. Zeliang et al. (2013) combined the Langmuir equation and multiple regressions to obtain the CBM content empirical formula and proved its effectiveness using logging data from the Qinshui Basin.

In general, traditional methods of CBM quantitative interpretation are predominantly statistical probabilistic approaches and the Langmuir equation. Statistical probabilistic approaches are usually established based on multiple linear regressions, which is difficult to use with a nonlinear problem, such as CBM content calculation. Furthermore, the use of the Langmuir equation to calculate CBM content is affected by the methane saturation. When CBM is unsaturated, the calculation can only show the relative content.

With the development of artificial intelligence technologies, neural networks have been introduced to logging data processing and interpretation. Currently, the most widely used method in CBM logging interpretation is the back-propagation neural network (BPNN) (Rumelhart et al. 1986). Pan and Liu (1997) applied a BPNN method to calculate CBM content. Junsheng and Ying (1999) obtained a good result by using

✉ Xiang Min
xjiexiangmin@163.com

¹ Xinjiang Institute of Engineering, Urumqi 830001, Xinjiang, China

BPNN to calculate coal-bed parameters, including CBM content. However, a local minimization problem exists in BPNN (Xiaofan and Tingkui 1994; Juanjuan and Hong 2006), which often makes it difficult to consistently obtain the global optimal solution; the result will change by repeated computation.

BPNN shortcomings are becoming increasingly recognized. Meanwhile, deep learning has emerged (Hinton 2012) and has been continuously developed. Deep learning is the latest achievement in artificial intelligence algorithms, which effectively discovers and describes the complex structural features of a problem. Moreover, it can improve computing performance (Krizhevsky and Sutskever 2012; Lecun et al. 2015; Silver and Huang 2016; Guo and Liu 2016). Deep learning has achieved a considerable success in many fields (Deng and Yu 2014; Alipanahi and DeLong 2015; Zhou and Troyanskaya 2015; Noda and Yamaguchi 2015; Liu and Luo 2015; Tomczak and Gonczarek 2017). Despite these advancements, it has rarely been applied to logging interpretation. In this paper, we introduce a deep belief network (DBN) method of deep learning to CBM logging data processing. We conducted experiments and determined the network parameters and a neural network that can be used to compute the CBM content. The proposed approach can be applied to the quantitative interpretation of CBM logging data.

Deep belief network

The DBN method of deep learning was proposed by Geoffrey Hinton, the ‘father of the neural network’, in 2006. DBN combines unsupervised and supervised learning, which involves several restricted Boltzmann machines (RBMs) and a BPNN.

Restricted Boltzmann machine

The RBM is the most important part of unsupervised learning in DBN. It is mainly used to extract the features of the inputs. An RBM includes two layers comprised of many neurons: one is the visual layer, which is used to receive the inputs, and the other is the hidden layer, which is used to detect the input features. Figure 1 depicts the structure of an RBM. The neurons in the same layer are independent of each other. The respective visual layer and hidden layer neurons are connected by the weight matrix, V , which can be represented as:

$$V = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1i} \\ v_{21} & v_{22} & \cdots & v_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ v_{j1} & v_{j2} & \cdots & v_{ji} \end{bmatrix} \tag{1}$$

where i is the quantity of the visual layer neurons and j is the quantity of the hidden layer neurons. The RBM turns the neurons on or off by means of V .

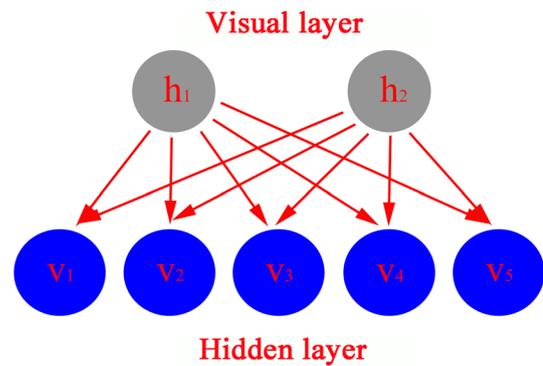


Fig. 1 Structure of an RBM

If the input vector in the visual layer is $P = (p_1, p_2, \dots, p_i)^T$, we first calculate the incentive value of the neurons in the hidden layer:

$$h = VP \tag{2}$$

where $h = (h_1, h_2, \dots, h_i)^T$. Then, we calculate the probability that the hidden layer neurons will be turned on (represented by 1) by the Sigmoid function:

$$P(h_m = 1) = \frac{1}{1 + e^{-h_m}} \tag{3}$$

where $h_m \in h$.

Thus, the probability that the hidden layer neurons will be turned off (represented by 0) is:

$$P(h_m = 0) = 1 - P(h_m = 1) \tag{4}$$

where $h_m \in h$.

Finally, we compare the probability that the hidden layer neurons will be turned on with a random value u ($u \in [0, 1]$) from the uniform distribution by the function as follows:

$$h_m = \begin{cases} 1, & P(h_m = 1) \geq u \\ 0, & P(h_m = 1) < u \end{cases} \tag{5}$$

The above procedure will determine whether the hidden layer neurons will be turned on.

The training of the RBM essentially establishes the probability distribution that can reflect the features of the samples by turning on or off the hidden layer’s neurons.

Back-propagation neural network

Proposed by Rumelhart and McClelland in 1986, the BPNN is a multilayer feedforward neural network (MLFNN) based on the back-propagation algorithm. The BPNN is the part of supervised learning in the DBN that is mainly used to calculate the final output. It is widely used in function approximation, pattern recognition, classification, data compression, data mining, and other areas. The BPNN can be divided into

three layers: input, hidden, and output. There is only one input layer and one output layer, while there may be one or several hidden layers. Figure 2 depicts the structure of the BPNN. In each layer, there are many neurons. The neurons in the same layer are not connected with each other. The neurons in different layers are connected by weight w_i . The input layer receives the data, and the hidden and output layers process the data. The data processing model is usually the Sigmoid function:

$$y_i = \sum_{i=1}^n x_i w_i - b \tag{6}$$

$$y_o = f(y_i) = \frac{1}{1 + e^{-y_i}} \tag{7}$$

where y_i is the input of a certain layer, y_o is the output of the layer, and x_i denotes the output of the previous layer. In addition, b is the threshold, which means that, if the input is greater than b , it will change the output; moreover, if the input is less than b , it will not affect the output.

Principal of the deep belief network

The characteristics extracted by the RBM can be used as new input to the BPNN. Then, the DBN will be successfully trained. The single BPNN is a typical ‘shallow’ neural network. The calculation is not ideal when the hidden layer is greater than two. However, in a DBN, a few additional RBMs exist. That is, the DBN determines the range of connection weights by RBMs and then computes the result by BP training. Compared with the single BPNN, the DBN training speed and convergence time are faster.

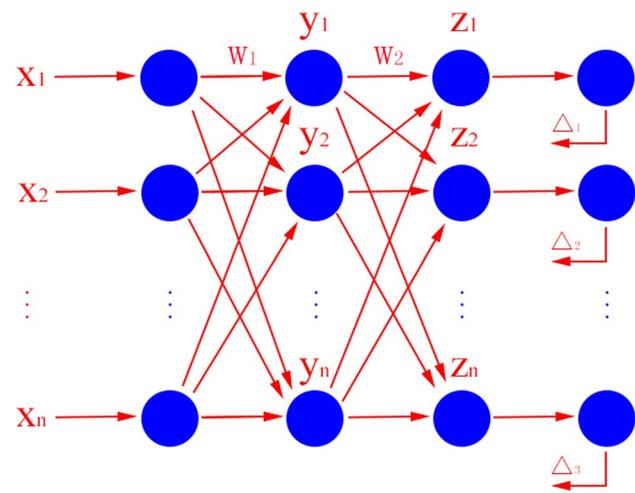


Fig. 2 Structure of a BPNN

Deep belief network parameter selection

Normalized data

Each logging technique has a different physical principle. Each of the respective measured physical parameters is significantly different with respect to the dimension and order of magnitude. However, the computer cannot distinguish these differences. Therefore, the logging data must be pre-processed to have the same dimensions and numerical ranges. To this end, normalization is generally a good method.

$$x_{nor} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{8}$$

where x_{nor} is the data after normalization, x comprises the original logging data, and x_{max} and x_{min} are the respective maximum and minimum of the different logging data. Through normalization, all the data change to dimensionless quantities, and their intervals are [0,1].

In order to optimize the performance of DBN, it is necessary to select suitable logging data to train the network. In this paper, the relationship between logging data and CBM contents is analysed by grey correlation analysis.

In this paper, nine conventional logging curves (calliper, natural gamma ray, spontaneous potential, deep lateral resistivity, shallow lateral resistivity, micro-spherically focused, density, acoustic and neutron) are represented as $X_i(h)=[x_1(h), x_2(h), \dots, x_9(h)]$ and CBM contents are represented as $Y(h)$, where h is the depth.

$$\gamma(Y, X_i) = \frac{\frac{1}{n} \sum_{h=1}^n \min_i \min_k |Y(h) - x_i(h)| + \xi \max_i \max_k |Y(h) - x_i(h)|}{|Y(h) - x_i(h)| + \xi \max_i \max_k |Y(h) - x_i(h)|} \tag{9}$$

The result of grey relational analysis is as follows.

Table 1 shows the grey relational grades between logging data and CBM contents. The greater the grade is, the stronger the correlation is. Generally speaking, in this paper for logging interpretation, the appropriate grey relational grade is greater than 0.8. (0.8 is the most commonly used standard.) So calliper, natural gamma ray, deep lateral resistivity, shallow lateral resistivity, density, acoustic and neutron are selected as the input parameters of DBN.

RBM quantity influence

The logging data in this study were obtained from Songliao Basin of Northeast China. First, we selected 300 coal seams for laboratory analysis data as the DBN samples. Second, we selected seven logging curves—calliper, deep lateral resistivity, shallow lateral resistivity, density, acoustic, neutron,

Table 1 Grey relational grades between logging data and CBM contents

	Calliper	Natural gamma ray	Spontaneous potential	Deep lateral resistivity	Shallow lateral resistivity
CBM contents	0.82	0.85	0.69	0.80	0.81
		Micro-spherically focused	Density	Acoustic	Neutron
CBM contents		0.75	0.84	0.82	0.83

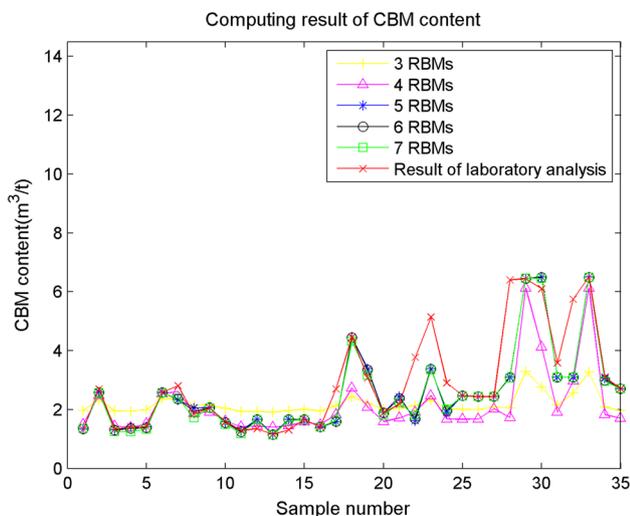


Fig. 3 Comparison of different quantities of RBMs

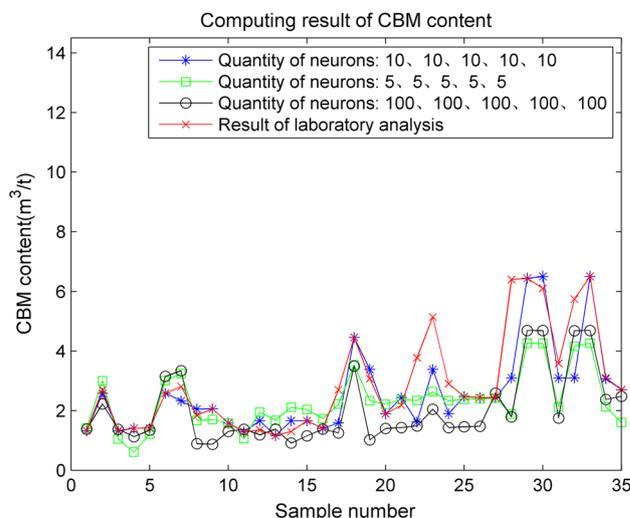


Fig. 4 Comparison of different quantities of neurons in hidden layers in each RBM

and natural gamma ray—as the DBN input parameters. Methane content was used as the DBN output parameters. After repeated experiments, we, respectively, explored the influence on the calculations by the quantity of RBMs, quantity of hidden layer neurons, and activation functions. Finally, we computed the CBM content of 35 coal seams in a certain well, ‘JQ1 Well’, by a statistical probabilistic method, the Langmuir equation, and a DBN. We then analysed the effects of different methods.

The DBN consists of several RBMs and a BPNN. The RBMs are mainly used to extract the features of logging data, and their quantities have an obvious influence on the calculation. Theoretically, the higher the quantity of RBMs, the more effectively the RBMs can extract the features of the logging data. However, the corresponding computing speed will also be significantly reduced. Therefore, the accuracy and speed should both be considered in practice.

Figure 3 presents the results of applying DBN to process CBM logging data with different quantities of RBMs. When there are three, four, or five RBMs, the methane content discrepancy between the computing result and laboratory analysis dramatically decreases with the increased quantity of RBMs. When there are more than five RBMs, the calculation accuracy is not obviously improved over the case involving exactly five RBMs. However, the operation is significantly

more complex, which leads to a decreased computing speed. Thus, through the comparison, it is obvious that five RBMs maximize the computing accuracy and speed.

Influence of neuron quantity

As described in “Restricted Boltzmann machine” section, each RBM consists of a visible layer and a hidden layer. The visual layer is mainly used to receive the logging data, and its quantities of neurons are the same as the quantities of the logging curves of input, i.e. seven. The hidden layer is mainly used to extract the features of the logging curves, and its neuron quantities have a considerable influence on extracting the features of the logging curves. Usually, the quantity of neurons is less than the sample. In this paper, we selected 300 coal seams as the samples. So we chose 5, 10, and 100 neurons as the experimental object. Figure 4 shows the computation results with different quantities of neurons in the hidden layers. According to “RBM quantity influence” section, five RBMs is the best choice. Therefore, when there are 100 neurons in the respective RBM hidden layers, the discrepancy of methane content between the computing results and laboratory analysis is the maximum and the most time is used in all three cases. This means

that a greater quantity of neurons in the hidden layer does not improve the effect. When there are five neurons, the fitting degree between the computing results and laboratory analysis is higher than 100 neurons. However, with too few neurons, the logging data features cannot be fully extracted, which makes the computing unstable by BPNN training of the DBN. Furthermore, repeated calculations are needed to provide the best outputs. When there are ten neurons, the discrepancy of methane content between the computing results and laboratory analysis is minimal and the calculation is stable. Therefore, for the logging data used in this paper, the best choice is ten neurons in each RBM hidden layer.

Activation function influence

In an artificial neural network, the activation function is used to add nonlinear factors, thereby enabling the neural network to more effectively solve the nonlinear problem. Currently, the frequently used activation functions are the following:

ReLU function:

$$f(x) = \max(0, x) \tag{10}$$

Sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{11}$$

Linear function:

$$f(x) = ax + b \tag{12}$$

Softmax function:

$$f(x) = \frac{e^{x_j}}{\sum_{k=1}^K e^{x_k}} \quad j = 1, 2 \dots K \tag{13}$$

Figure 5 shows the results of CBM content when there are five RBMs and ten neurons in each RBM hidden layer. When the activation function is Softmax, the computing result of the methane content is very poor and it is scarcely able to provide useful information. In comparing the computing results with the activation function of Sigmoid and Linear in the figure, the Sigmoid function is more suitable for the Nos. 12 and 15 coal seams, and the Linear function is more effective for the No. 27 coal seam.

In addition, the results with the two activation functions are very close. When the activation function is ReLU, compared with the other three activation functions, the computing result of CBM content has the fewest discrepancies with the laboratory analysis results, which means that the fitting degree is the best. It can be observed that the ReLU function is the most appropriate activation function for the logging data used in this paper, and the Sigmoid function and Linear function are second, while the Softmax function should not be considered.

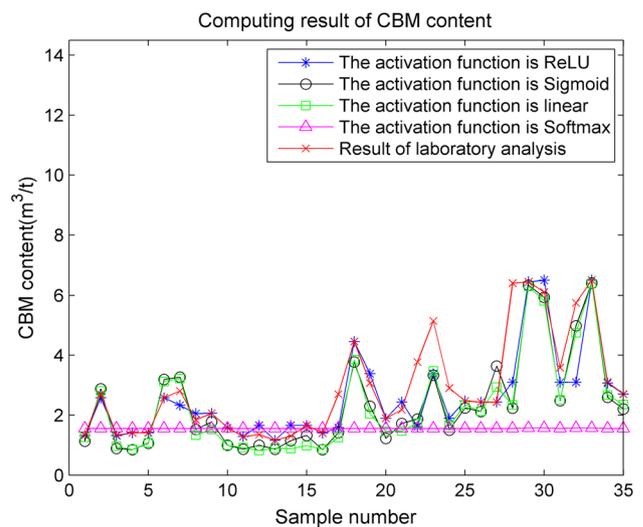


Fig. 5 Comparison of different activation functions

Loss function influence

The loss function is used to measure the difference between the calculation results and the expected values. At present, cross-entropy function and mean square error (MSE) function are most frequently used.

Cross-entropy function:

$$H(p, q) = - \sum_x p(x) \log q(x) \tag{14}$$

MSE function:

$$\text{MSE}(p, q) = E(p(x) - q(x))^2 \tag{15}$$

where $p(x)$ and $q(x)$, respectively, present the calculation results and the expected values. $E(x)$ is used to get the mathematical expectation.

Figure 6 shows the results of CBM content with the cross-entropy function and MSE function. When using cross-entropy, compared with MSE function, the results of the Nos. 7, 14, 15, 20, 31, and 33 coal seams are slightly accurate. But in general, both the two loss functions are acceptable.

Comparison with BPNN

Figure 7 presents the results of CBM content with DBN and BPNN. The computing results' accuracies of DBN are generally better than BPNN. Either DBN or BPNN is built by gradually determining the connection weights. The difference between DBN and BPNN is that there are several RBMs in DBN. For BPNN, the connection weights are randomly initialized. And for DBN, the computing process of

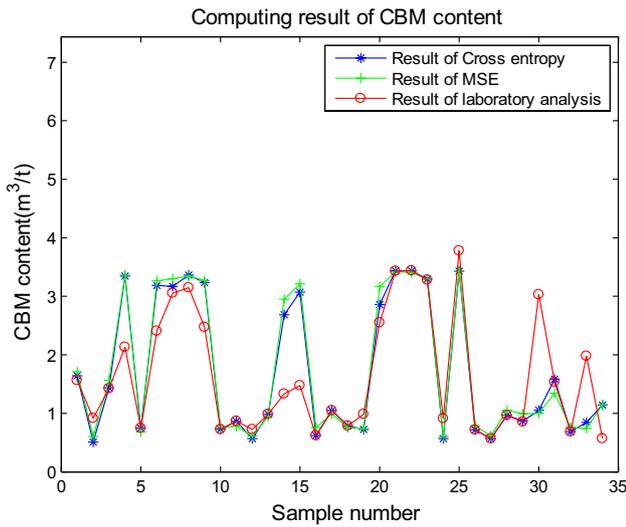


Fig. 6 Comparison of different loss functions

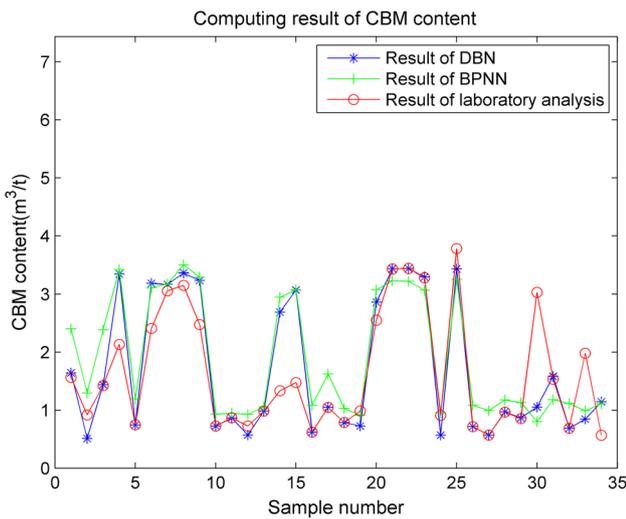


Fig. 7 Comparison of DBN and BPNN

RBM is essential in determining the range of connection weights, which improves the accuracy of the calculation.

Comparison of related methods

The statistical probabilistic method and Langmuir equation are commonly used for calculating CBM content. The statistical probabilistic method supposes that CBM content and logging data obey a certain statistical probabilistic distribution. Thus, through multiple regressions, the regression equation of the relation between CBM content and logging data can be obtained. The equation can then be used to estimate the CBM content in the same region. The

Langmuir equation is a widely used isothermal adsorption equation with which CBM content can also be estimated.

In this paper, we compare the statistical probabilistic method, Langmuir equation, and DBN in the calculation of CBM content. DBN is comprised of five RBMs and a BPNN. There are ten neurons in the hidden layer of each RBM. The ReLU function is selected as the activation function. In Fig. 8, the most accurate methods are DBN, followed by the statistical probabilistic method, and then the Langmuir equation. The computing result of the latter is the CBM content from when the methane saturation is 100%. However, in most cases, the methane is unsaturated; thus, the computing result is only the relative CBM content in the coal seam. If there is no information about the methane saturation, we cannot obtain the accurate methane content by the Langmuir equation. Therefore, the CBM content of the Langmuir equation is generally higher than those of the other two methods.

The statistical probabilistic method and DBN are fundamentally similar. That is, through the samples, we can obtain the relationship between the CBM content and the logging data. Accordingly, we can estimate the CBM content in other coal seams. The difference between the approaches is the means of reflecting the relationship between the CBM content and logging data. The statistical probabilistic method supplies a regression equation (usually a linear equation), and DBN provides a neural network. As shown in Fig. 8, DBN has a better effect in the calculation of CBM content, which is a typical nonlinear problem.

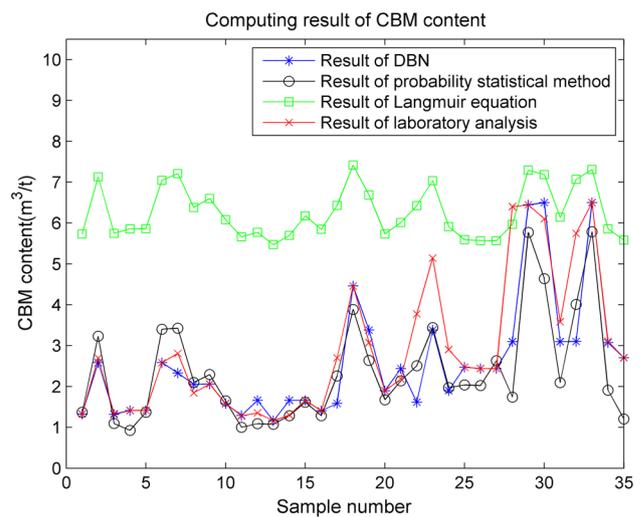


Fig. 8 Comparison between DBN and BPNN

Conclusion

Determining a means of combining logging data and CBM content for exploring a CBM content estimation method without coring has been a continual focus of CBM logging. This paper introduced deep learning into the quantitative interpretation of CBM logging. We discussed aspects of DBN parameter selection, as well as the advantage of DBN over the statistical probabilistic method and the Langmuir equation.

Several experiments were conducted, and the following conclusions were drawn. First, increasing the quantity of RBMs is helpful for extracting the features of the logging data. However, it requires additional time. Therefore, although the layer is helpful for finding features, a higher quantity of RBMs is not better and the calculation accuracy and speed should thus also be considered.

Second, too many RBM hidden layer neurons can significantly reduce the computing precision and speed, whereas too few hidden layer neurons can reduce the calculation accuracy and stability of the results. Therefore, the quantity of implicit layer neurons should be moderate.

Third, the ReLU function is the best for logging data, and the Sigmoid function and Linear function are second. The Softmax function has no effect on the selection of activation functions.

Fourth, the cross-entropy function is superior to MSE function.

Fifth, RBMs are used to determine the range of the connection weights. Compared with BPNN, it improves the accuracy of the calculation.

Sixth, the calculation results of the Langmuir equation are greatly influenced by the methane saturation. The statistical probabilistic method has a limited ability to deal with nonlinear problems. DBN is the best choice for calculation of CBM logging.

Using DBN to predict the CBM content can be generally applied in different regions. But it must be noted that, in order to ensure accuracy, we cannot train DBN by the cross-regional samples. Furthermore, the network parameters are very important to the calculation results for DBN. The network parameters were determined through repeated experiments. In the future, we intend to examine the relationship between the network parameters and CBM logging data. Accordingly, a rapid and accurate method of determining the network parameters will be explored. In addition, DBN can be introduced in other aspects of logging interpretation, such as lithology identification, porosity calculation, and saturation calculation. These problems still need further study.

Acknowledgements This research was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (2017D01B08), the Scientific Research Planning Project of

Xinjiang Uygur Autonomous Region (XJEDU2017S056 and XJEDU2017S057), the Ph.D. Research Startup Foundation of the Xinjiang Institute of Engineering (2016xgy341812), *Tianchi* Doctor Research Project of Xinjiang Uygur Autonomous Region (BS2017001) and Xinjiang Uygur Autonomous Region key specialty of Geological Engineering.

References

- Alipanahi B, Delong A (2015) Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat Biotechnol* 33(8):831–838. <https://doi.org/10.1038/nature14539>
- Bhanja AK, Srivastava OP (2008) A new approach to estimate CBM gas content from well logs. *SPE115563:1–5*. <https://doi.org/10.2118/115563-ms>
- Deng L, Yu D (2014) Deep learning: methods and applications. *Found Trends Signal Process* 7(3):197–387. <https://doi.org/10.1561/20000000039>
- Guo Y, Liu Y (2016) Deep learning for visual understanding: a review. *Neurocomputing* 187(C):27–48. <https://doi.org/10.1016/j.neucom.2015.09.116>
- Hawkins JM, Schraufnagel RA, Olszewski AJ (1992) Estimating coalbed gas content and sorption isotherm using well log data. *Phys Rev Lett* 97(7):1143–1238. <https://doi.org/10.2523/24905-MS>
- Hinton GE (2012) A practical guide to training restricted boltzmann machines, vol 7700. Springer, Berlin, pp 599–619
- Juanjuan L, Hong C (2006) Researching development on BP neural networks. *Control Eng China* 13(5):449–451
- Junsheng H, Ying W (1999) Interpretation of well logging data for coalbed methane using BP neural network. *Geol Prospect* 35(3):41–45
- Krizhevsky A, Sutskever I (2012) ImageNet classification with deep convolutional neural networks. *Int Conf Neural Inf Process Syst* 25(2):1097–1105. <https://doi.org/10.1145/3065386>
- Langmuir I (1918) The adsorption of gases on plane surfaces of glass, mica and platinum. *J Am Chem Soc* 40:1361–1370. <https://doi.org/10.1021/ja02242a004>
- Lecun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
- Liu Z, Luo P (2015) Deep learning face attributes in the wild. In: *Proceedings of the IEEE international conference on computer vision*, pp 3730–3738. <https://doi.org/10.1109/iccv.2015.425>
- Noda K, Yamaguchi Y (2015) Audio-visual speech recognition using deep learning. *Appl Intell* 42(4):722–737. <https://doi.org/10.1007/s10489-014-0629-7>
- Pan H, Liu G (1997) Applying back-propagation artificial neural networks to predict coal quality parameters and coal bed gas content. *Earth Sci J China Univ Geosci* 22(2):210–214
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning Represent Back Propag Errors. *Nature* 323:533–536. <https://doi.org/10.1038/323533a0>
- Silver D, Huang A (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529(7587):484–492. <https://doi.org/10.1038/nature16961>
- Tomczak JM, Gonczonek A (2017) Learning invariant features using subspace restricted boltzmann machine. *Neural Process Lett* 45(1):173–182. <https://doi.org/10.1007/s11063-016-9519-9>
- Wang Z (2009) Logging methods evaluation of the gas content in Coal-bed methane reservoir. Jilin University, Changchun, pp 55–60
- Xiaofan Y, Tingkui C (1994) Inherent advantages and disadvantages of artificial neural networks. *Comput Sci* 2:23–26

- Yang Y, Cloud T, Kirk CV (2005) New application of well log parameters in coalbed methane (CBM) reservoir evaluation at the Drunkards Wash Unit, Uinta Basin, Utah. In: SPE Eastern regional meeting, 1–9. <https://doi.org/10.2523/97988-ms>
- Zeliang J, Haifei X, Haibin G (2013) Technology for evaluation of CBM reservoir logging and its application. *Coal Geol Explor* 41(2):42–45
- Zhou J, Troyanskaya OG (2015) Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods* 12(10):931–934. <https://doi.org/10.1038/nmeth.3547>



Effect of water–air heat transfer on the spread of thermal pollution in rivers

Monika Barbara Kalinowska¹

Received: 14 November 2018 / Accepted: 21 January 2019 / Published online: 20 February 2019
© The Author(s) 2019

Abstract

While working on practical problems related to the spread of thermal pollution in rivers, we face difficulties related to the collection of necessary data. However, we would like to predict the increase in water temperature at the best accuracy to forecast possible threats to the environment. What level of accuracy is necessary and which processes that influence the water temperature change have to be taken into account are usually problematic. Those problems, with special stress on water–air heat exchange in practical applications in the so-called mid-field region in rivers, which is very important for the environmental impact assessment, constitute the main subject of the present article. The article also summarises the existing knowledge and practice on water–air heat exchange calculations in practical applications.

Keywords Thermal pollution modelling · Water–air heat exchange · Heat budget · Mixing in rivers · RivMix model · Heat transport equation

Introduction

In the aquatic environment, thermal pollution is understood as a result of any process that changes ambient water temperature. In rivers, such change in water temperature that goes beyond the natural range of temperature variation may be caused by discharged heated water coming from different industrial facilities using water for the cooling purpose (thermal electric power plants, chemical plants, etc.). We are aware that relatively small changes in natural ambient temperature might create substantial environmental problems; for instance, they may influence the dissolved oxygen concentration (Rajwa-Kuligiewicz et al. 2015), population of fish and other aquatic organisms and plants (Brett 1956; Coutant 1999; Currie et al. 1998; Murray 2002). The list of potential environmental impacts of thermal pollution is very long and widely discussed in the literature (see, e.g., Caissie 2006; Hester and Doyle 2011; Webb et al. 2008). Therefore, the prediction of possible water temperature increase and assessment of its environmental impact, for example, before constructing industrial facilities become mandatory.

While trying to predict the potential increase of natural water temperature below the heated water discharge, one might have problems with making assumptions and relevant simplifications to obtain a desirable solution. Heat transport models generally consist of nonlinear differential equations with no analytical solutions under real conditions, and therefore, they must be solved numerically. Choosing a proper solution to a problem is a challenge. It requires not only a compromise between the computation time and cost needed to obtain the results and the desired accuracy of the solution but also depends on the availability of input data and their accuracy. The most obvious simplifications of the general three-dimensional (3D) problem pertain to the reduction of the dimension to the two- (2D) or even one-dimensional (1D) problem. That corresponds to the description of so-called near-field (3D); mid-field (2D); or far-field region (1D) (see Table 1). In each, particular region, vertical, horizontal and longitudinal mixing prevails, respectively. While the vertical mixing is relatively fast, the transverse mixing may continue far away from the discharge point. Therefore, the mid-field region is very important from the environmental point of view. To describe the heat transport in the 2D case in the mid-field region, the 2D depth-averaged equation has to be used (see, e.g., Kalinowska and Rowinski 2015; Kalinowska and Rowiński 2008; Rutherford 1994; Szymkiewicz 2010). Further simplifications of heat transport

✉ Monika Barbara Kalinowska
Monika.Kalinowska@igf.edu.pl

¹ Institute of Geophysics Polish Academy of Sciences,
Warsaw, Poland

Table 1 Schematic description of characteristic regions for mixing in rivers

3D region near-field	2D region mid-field	1D region far-field
Vertical mixing prevails	Horizontal mixing prevails	Longitudinal mixing prevails
It begins at the discharge point and finishes after full vertical mixing. Rapid process with maximum dimensions of few tens of the water depth (Jirka and Weitbrecht 2005), of the order of 100 H (Endrizzi et al. 2002). Approximated distance to the location of complete vertical mixing: $L_{mv} \approx 50 H^1$	It begins after complete mixing over the depth and finishes after full lateral mixing. May continue as long as several hundreds of river width (Endrizzi et al. 2002). For typical river ($B/H = < 10, 100 >$), the complete mixing requires from 100 B to 1000 B . Approximate distance to the location of complete horizontal mixing: $L_{mh} \approx 8 (B/H)$	It begins after complete mixing over the depth and width and stretches down the river
3D heat transport equation: $\frac{\partial T}{\partial t} = \nabla \cdot [(\mathbf{D}_M + \mathbf{D}_T) \cdot \nabla T] - \nabla[\mathbf{v} \cdot T] + Q$	2D depth-averaged heat transport equation: $h \frac{\partial T}{\partial t} = \nabla(h\mathbf{D} \cdot \nabla T) - \nabla(h\mathbf{v} \cdot T) + Q$	1D cross-sectionally averaged heat transport equation: $\frac{\partial T}{\partial t} = \frac{1}{A} \frac{\partial}{\partial x} \left(AD \frac{\partial T}{\partial x} \right) - v_x \frac{\partial T}{\partial x} + Q$
high/slow ←	computational cost/simulation time	→ low/fast
huge ←		amount of input data
Result: 3D temperature filed	Result: 2D depth-averaged temperature field	Result: 1D cross-sectionally averaged temperature values

H averaged river depth, B river width, t time, T water temperature (time, depth and cross-sectionally averaged, respectively, in 3D, 2D and 1D), \mathbf{x} position vector ($\mathbf{x} = (x, y, z)$ in 3D, $\mathbf{x} = (x, y)$ in 2D and $\mathbf{x} = x$ in 1D, x longitudinal coordinate, y transversal (lateral) coordinate, z vertical coordinate), \mathbf{v} velocity vector, \mathbf{D}_M molecular heat diffusion tensor, \mathbf{D}_T turbulent heat diffusion tensor, \mathbf{D} dispersion tensor, D longitudinal dispersion coefficient, $A = BH$, Q source function describing additional heating or cooling processes (for details, see Kalinowska and Rowinski 2015)

¹ Assuming discharge point at the bottom (distance will be reduced, for example, when the discharge point will be in the middle depth or other)

equation are associated with different equation terms. Those terms are often difficult to estimate like, for example, the dispersion tensor components (see, e.g., Kalinowska and Rowiński 2008; Rowinski and Kalinowska 2006) in 2D equation, longitudinal dispersion coefficient (see, e.g., Guymer 1998; Piotrowski et al. 2006; Rowiński et al. 2005; Wallis and Manson 2004) in 1D equation or source functions that may include the information on heat exchange between river water and its surrounding.

The description of the heat exchange between river water and river surrounding is usually very complicated. The involved processes often depend on various local and temporal factors. It is often taken for granted that the more factors are taken into account in the model/report, the better this model/report is. At the same time, the exact results are expected despite the availability of an insufficient amount of input data. An obvious question arises what result is necessary and what accuracy of the solution is really required. In most cases, modelling of the whole 3D temperature field is not feasible and needed. On the other extreme, we may ask whether 1D models can be sufficient tools for environmental impact assessments for rivers. Usually, it is not the case since vertical mixing takes a lot of time and cross-sectionally averaged temperature values may be much lower than the point values within the cross section. Nevertheless, due to the lack of input data allowing to run 2D models, 1D approaches are often used, but one has to realise that they may introduce serious errors. In case of applying 1D approach, the resulting predicted temperature increase at the beginning

(cross-sectionally averaged values) is very low compared to the predicted maximum temperature that may appear at comparable distances from the discharge point in case of 2D approach. The focus of the present paper is, therefore, the modelling of the spread of thermal pollutions in the river mid-field zone, very crucial from the environmental point of view, using the 2D depth-averaged models. For such models within the relatively short timescale (compared to the 1D models), the issue related to the additional heat sources that eventually should be taken into account is of crucial importance. To be very accurate, one may obviously try to include all known possible processes that can potentially affect the river temperature (see, e.g., Evans et al. 1998; Hannah et al. 2004; Johnson 2004; Webb and Zhang 1999; Xin and Kinouchi 2013): heat exchange with the atmosphere and with the river bottom and banks; rainfall and groundwater flow, heat production from biological and chemical processes or friction. However, very often data necessary to calculate those heat fluxes are not easily, if at all, available. For instance, the interactions between riverbed and the stream water are complex and depend on many local and temporal factors like groundwater-stream water flux, shading, bed morphology, geologic heterogeneity. It is difficult to assess them reliably. Most of those heat fluxes usually insignificantly affect the final results in the mid-field zone, especially when we are interested only in temperature difference caused by the artificial heat source (not the actual water temperature itself). Whether their omission is admissible or not strongly depends on the considered case. Note

that in some cases, for example, in long timescale or specific situation, some of them may be significant. Generally, in the case of thermal pollution modelling, the issue of omitting different heat exchange terms with the environment such as heat exchange with the bottom, banks and sediment is very common, but the heat exchange with the atmosphere is usually taken into account in various heat transport models and desired by users (even if in 2D models other unavoidable errors committed during the calculation affect the final results to a much larger extent). Often its estimation is very problematic and doubtful, especially in practical applications for thermal pollution spreading. However, the issue of excluding the heat exchange with the atmosphere is usually a subject of discussion. The role of the heat exchange between water and air in case of thermal pollution modelling in rivers in the mid-field region is the main subject of the presented study. This article also summarises existing knowledge on water–air heat exchange calculations, with the special emphasis on practical applications.

Water–air heat exchange

Concerning the heat exchange between a river and its environment, the heat exchange between water and atmosphere is the most significant (Evans et al. 1998; Webb and Zhang 1999). Its intensity depends on water temperature and external meteorological and hydrological conditions, and it is influenced by many processes (Chapra 2008; Edinger 1974; Rutherford et al. 1993) which may be divided into two groups:

Processes independent of water temperature, i.e.,

- Shortwave solar radiation—radiation emitted by the sun (also called shortwave radiation or solar radiation);
- Longwave atmospheric radiation—radiation that water receives from the atmosphere (also called thermal radiation); sometimes, it is calculated together with other sources of the longwave radiation from surrounding terrain like, for example, radiation from vegetation;

Processes dependent on water temperature, i.e.,

- Longwave water back radiation—radiation emitted by the water surface (also called longwave back radiation);
- evaporation and condensation—there are processes with matter changes from one state to another; evaporation is the loss of water to the atmosphere in the form of water vapour, and it is associated with the heat loss from the water surface, while the condensation is the reverse process;
- conduction and convection—processes that take place at the border of water and air if they have different temperatures.

Those processes have been schematically presented in Fig. 1. Note that other processes may also be taken into account but their role usually become significant only in some specific cases.

Finally, the net heat flux Q_A [W m^{-2}] that results from the energy balance at the water–air interface is defined as follows:

$$Q_A = q_s + q_a - q_b \pm q_e \pm q_h, \quad (1)$$

where q_s [W m^{-2}]—net flux of shortwave solar radiation, q_a [W m^{-2}]—net flux of longwave atmospheric radiation, q_b [W m^{-2}]—longwave back radiation flux, q_e [W m^{-2}]—evaporation and condensation flux, q_h [W m^{-2}]—conduction and convection flux. Many different formulae are available in the literature that can be used to compute each term of Eq. (1), but the selection of an appropriate one is a debatable issue. Generally, the net heat flux is calculated based on the water temperature (T_w) and various meteorological data, primarily on the basis of air temperature (T_a), air humidity (R_h), air pressure (p_a), measured shortwave solar radiation (q_{si}) and wind speed (u) (although many others may be taken into account). Figure 2 presents examples of values of the calculated heat fluxes for a selected case over the day.

The practice of calculations of water–air heat exchange

An estimate of the net heat flux at the water–air interface turns out to be a challenge, especially when it has to be based on historical data. The first problem is related to obtaining of necessary input data for the analysed site. Intuitively, the best source in case of meteorological data is the nearest meteorological station. Unfortunately, in many cases, the nearest station does

Fig. 1 Most important processes affecting the heat exchange at the water–air interface

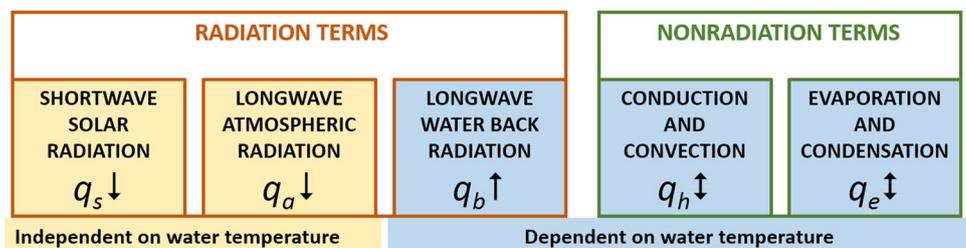
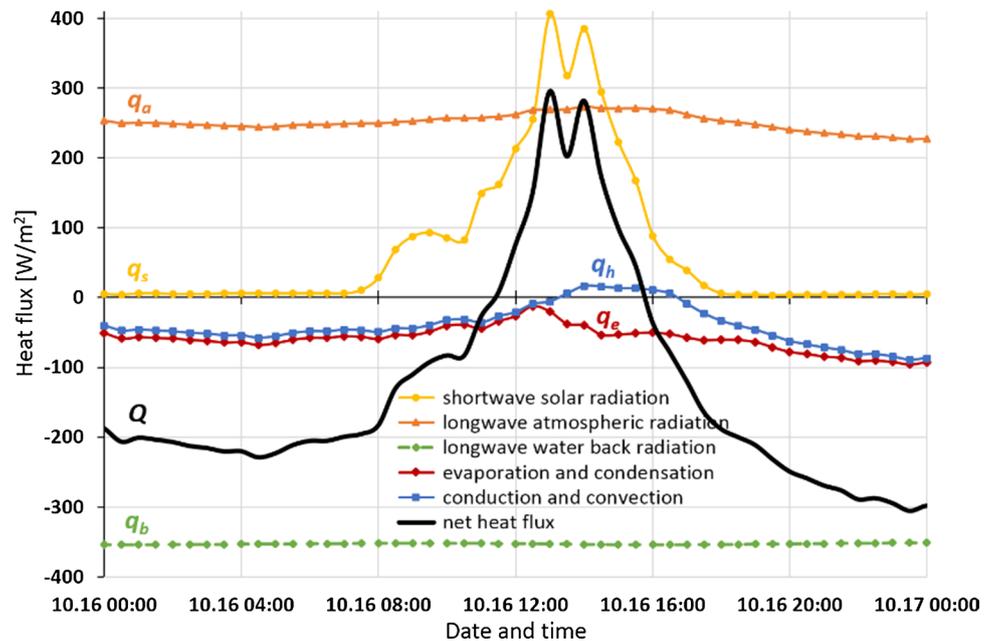


Fig. 2 Heat fluxes exemplary values calculated for a selected case over the day (Narew River, SET I, *Meteo NPN*, see Table 3)



not provide all the necessary data or is still too far from the considered river reach. Another problem is associated with the differences in data obtained from the neighbouring stations. Even in conducive situations when the necessary meteorological data are available in the vicinity of the river reach, we still may receive uncertain results related to the location of the station, for which conditions such as shading or wind speed are often considerably different from those at the river channel. The problem is widely discussed in the literature (see, e.g., Benyahya et al. 2010; Garner et al. 2014; Johnson 2004). Moreover, some measured quantities may vary along the river channel (or even across the river width). Finally, very different values of net heat flux may be obtained. Johnson (2004), for example, measured the heat fluxes for the same stream in case of different conditions. The results showed that the final sum of heat fluxes measured at the same time was 580 W m^{-2} towards the stream in full sun and 149 W m^{-2} away from the stream under the shade. Additionally, each term in Eq. (1) is sensitive to the chosen computation method; i.e., different formulae may lead to varying results since they often depend on not well-defined parameters, factors or coefficients that are site specific. Moreover, in the “competition” for the best formula, increasingly “more accurate” formulae take into account more and more factors and thus require more and more input data, which again in practical applications are rarely available or costly.

Study sites and input data

To illustrate the calculations of the net heat flux at the water–air interface and to discuss the possible problems that may be encountered in practical applications, the measured

water temperatures in two different lowland rivers in Poland, namely Narew and Świder (see Fig. 3) having different characteristics (see Table 2), were used. Although both river sections are of similar width, Świder River is significantly shallower. The detailed description of both sites may be found in (Rajwa et al. 2014) and (Rajwa-Kuligiewicz et al. 2015). The relevant historical meteorological data necessary to compute the net heat flux Q_A were collected from different available meteorological stations. Additionally, special in situ measurements on Świder River using the portable meteorological station were performed to execute additional analyses. The station was located on the riverbank to capture the river microclimate conditions.

All data series mentioned in the paper are listed in Table 3. Water temperature has been measured every 5 or every 1 min. Most of the meteorological data have been measured every minute or more often, for some datasets every 5 min. Thirty-minute backwards averaged values were used for the calculations. Some example sets of meteorological input data used in the article may be found in (Kalinowska et al. 2018, Figs. 1–4). Figure 3 presents the location of measurement points of water temperature in the Narew and Świder Rivers, the nearby meteorological stations providing data and the approximate distance from the measurement points to the considered meteorological stations. Investigations in both rivers were carried out under different conditions, and data from different meteorological stations were taken (for details see Table 3). Some datasets taken into account at the beginning were useless because of existing significant gaps or unexplained outliers, which is sometimes a serious problem in practical applications based on historical data. The discrepancy between data from

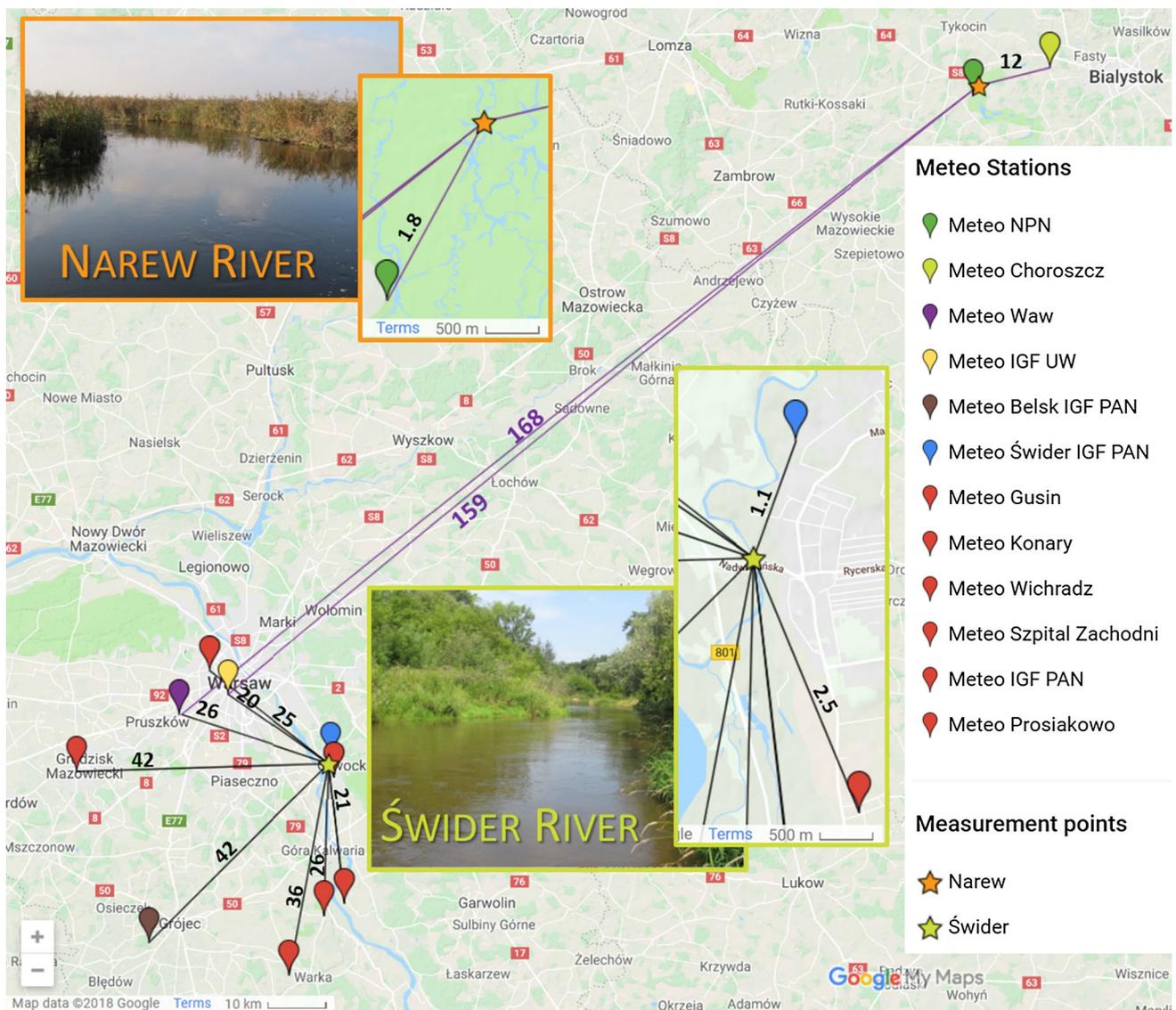


Fig. 3 Map with the marked water temperature measuring points on the Narew (53°04'38"N 22°49'09"E) and Świder (52°06'26"N 21°13'39"E) Rivers (orange and green stars) and selected nearby

meteorological stations. Approximate distances [km] from the particular measurement points to meteorological stations have also been indicated. Map has been created with the use of Google Maps

Table 2 Considered rivers' sections characteristics (during the measurement campaign in 2013)

Name	Width B [m]	Averaged depth H [m]	Slope S [–]	Hydraulic radius R [m]	Discharge Q [m ³ s ⁻¹]	Shear velocity U^* [m s ⁻¹]	Mean velocity U [m s ⁻¹]
Narew River	20	1.8	0.0001	1.16	11.0	0.03	0.3
Świder River	20	0.3	0.0023	0.26	3.0	0.07	0.5

Table 3 List of input data sets mentioned in the paper

Data set	Measurements time range	Meteo station	Distance ¹ [km]	Data available
Narew River				
SET I	From 16.10 00:00 to 17.10.2013 00:00 T_w (every 5 min) ²	<i>Meteo NPN</i> ³	1.8	$T_a, R_{hr}, p_a, q_{si}, u$
		<i>Meteo Choroszcz</i> ⁴	12	T_a, R_{hr}, p_a, q_{si}
		<i>Meteo Waw</i> ⁵	159	$T_a, R_{hr}, p_a, q_{si}, u$
		<i>Meteo IGF UW</i> ⁶	168	$T_a, R_{hr}, p_a, q_{si}, q_{ai}, u$
Świder River				
SET II	From 26.09 00:00 to 27.09.2013 00:00 T_w (every 5 min)	<i>Meteo Świder IGF PAN</i> ⁷	1.1	T_a, u
		<i>Meteo IGF UW</i> ⁶	20	$T_a, R_{hr}, p_a, q_{si}, q_{ai}, u$
		<i>Meteo Waw</i> ⁵	26	$T_a, R_{hr}, p_a, q_{si}, u$
		<i>Meteo Belsk IGF PAN</i> ⁸	42	T_a, R_{hr}, p_a, u
SET III	From 02.05 00:00 to 03.05.2016 00:00 T_w (every 1 min)	<i>Meteo Local</i> ⁹	0.1	$T_a, R_{hr}, p_a, q_{si}, u, C$
		<i>Meteo Prosiakowo</i> ¹⁰	2.5	T_a, R_{hr}, p_a, u
		<i>Meteo Waw</i> ⁵	26	$T_a, R_{hr}, p_a, q_{si}, u$
SET IV	From 30.07 00:00 to 31.07.2016 00:00 T_w (every 1 min)	<i>Meteo Local</i> ⁹	0.1	$T_a, R_{hr}, p_a, q_{si}, u, C$
		<i>Meteo Prosiakowo</i> ¹⁰	2.5	T_a, R_{hr}, p_a, u
		<i>Meteo IGF PAN</i> ¹⁰	25	$T_a, R_{hr}, p_a, q_{si}, u$
		<i>Meteo Gusin</i> ¹⁰	21	T_a, R_h
		<i>Meteo Konary</i> ¹⁰	26	T_a, R_h
		<i>Meteo Waw</i> ⁵	26	$T_a, R_{hr}, p_a, q_{si}, u$
		<i>Meteo Wichradz</i> ¹⁰	36	T_a, R_{hr}, p_a, u
		<i>Meteo Szpital Zachodni</i> ¹⁰	42	$T_a, R_{hr}, p_a, q_{si}, u$

T_w water temperature [°C], T_a air temperature [°C], R_h air humidity [%], p_a air pressure [mb=hPa], q_{si} measured shortwave solar radiation [$W m^{-2}$], u wind speed [$m s^{-1}$], C cloudiness [octans]

¹The distance between the measurement point and meteo station; ²water temperature was measured with the Handheld Optical Dissolved Oxygen Meter (ProODO, YSI) equipped with water temperature sensor, barometric pressure sensor and an optical dissolved oxygen sensor; ³*Meteo NPN*—Narew National Park Weather Station (<http://meteo.npn.pl/>), 53°06'17.56"N 22°47'51.84"E, 114 m a.s.l.; ⁴*Meteo Choroszcz*—Choroszcz Weather Station, 53°08'33.1"N 22°59'16.7"E, 148 m a.s.l.; ⁵*Meteo Waw*—Warsaw Meteo Station (<http://www.meteo.waw.pl/>), 52°10'53"N 20°52'13"E 110 m a.s.l.; ⁶*Meteo IGF UW*—IGF UW Meteorological observatory (<http://metobs.igf.fuw.edu.pl/>), 52°12'42.7"N 20°58'59.5"E, 148 m a.s.l.; ⁷*Meteo Świder IGF PAN*—Geophysical Observatory IGF PAS in Świder, 52°06'55"N 21°14'15"E, 96 m a.s.l.; ⁸*Meteo Belsk IGF PAN*—Central Geophysical Observatory of the Institute of Geophysics in Belsk, 51°50'10"N 20°47'34"E, 180 m a.s.l.; ⁹*Meteo Local*—potable meteorological station: Davis Vantage Pro2 Weather Station (Wireless); ¹⁰*Meteo Prosiakowo*, *Meteo IGF PAN*, *Meteo Gusin*, *Meteo Konary*, *Meteo Wichradz* and *Meteo Szpital Zachodni*—stations that are part of the Davis Instruments network, data are provided online on the: <http://www.weatherlink.com/>

different stations was also observed (see, e.g., the air temperature and air humidity in Fig. 4). The sensitivity analyses for different input meteorological parameters may be found in Kalinowska et al. (2018).

Results and discussion

Each term in Eq. (1) was analysed and computed taking into account various calculation methods and different available input data sets for both rivers, preceded by existing knowledge summary. Details of calculations, different options, various empirical or semi-empirical formulae and the problems encountered are discussed below.

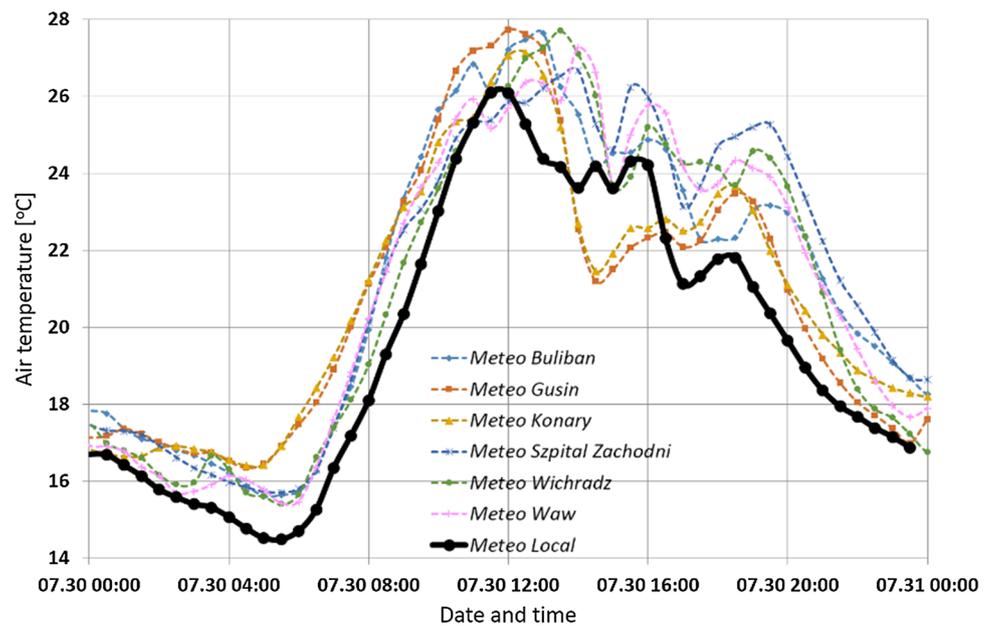
Processes independent on water temperature

Shortwave solar radiation

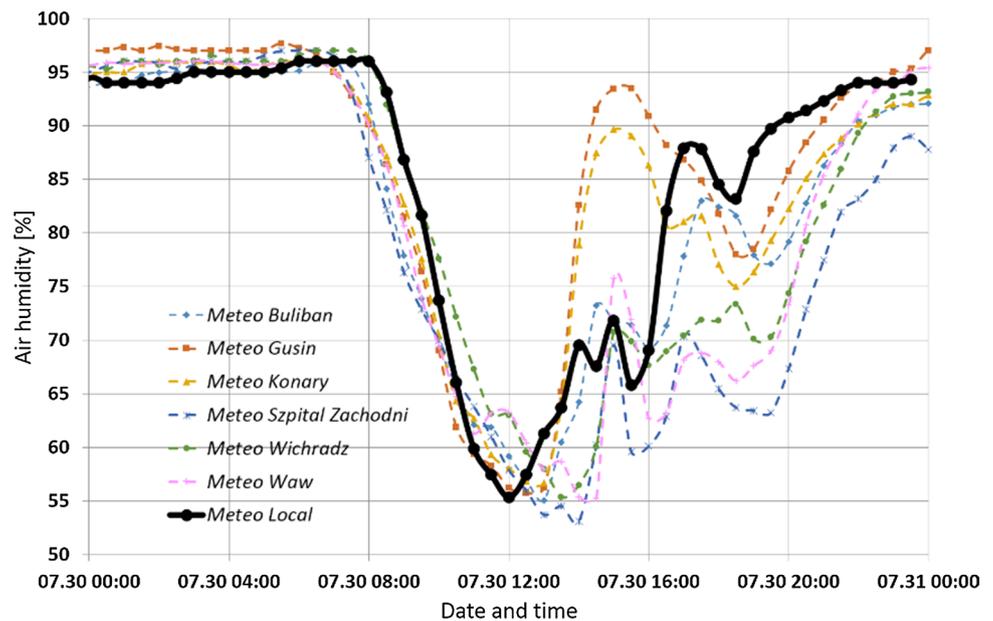
The value of the shortwave solar radiation can vary considerably over time. It depends on many factors such as:

- sun's position—varies depending on the date and time of day, site location and elevation above the sea level;
- scattering and absorption—some amount of solar radiation is absorbed and dissipated by the atmosphere, reflected by clouds or absorbed by atmospheric gases and dust;
- reflection—upon reaching the water surface, a part of the solar radiation is reflected by the surface;
- shading—for example, from the riverbank or the riparian vegetation.

Fig. 4 Discrepancy that exists between data from different available meteorological stations for Świder River (SET IV): **a** air temperature; **b** air humidity



(a)



(b)

The heat flux value may range between 800 and 1000 W m^{-2} for a sunny day and between 100 and 300 W m^{-2} for a very cloudy day.

The net flux of shortwave radiation is a result of the difference between incoming (q_{si}) and reflected solar radiation (q_{sr}) from the water surface, and it is usually approximated by (Magnusson et al. 2012)

$$q_s = q_{si}(1 - A), \quad (2)$$

where A is the albedo (or reflection coefficient for shortwave radiation), $[-]$ is a measure of reflectivity of the water surface. Albedo depends mainly on the colour of the surface (the darker the surface is, the smaller is the albedo), and for water, it ranges from 1 to 10% (depending on the clarity of the water, the water surface conditions—wavelength, chlorophyll concentration, the water depth, etc.). It also depends on the sun angle (water has a higher albedo for low sun angle, Katsaros et al. 1985). In heat budget studies in rivers, its value is usually assumed to be constant and it ranges from

0.02 to 0.07, for example, 0.03 (Benyahya et al. 2010; Caisie et al. 2007); 0.05 (Garner et al. 2014; Magnusson et al. 2012); 0.06 (Xin and Kinouchi 2013); 0.07 (Alcântara et al. 2010). In the present study, the albedo has been assumed to be constant at $A = 0.06$.

Equation (2) may be additionally corrected, taking into account the shade, by $(1 - S_F)$ term, where S_F is the shading factor. The S_F factor ranges from 0 (no shading) to 1 (complete shading). Many researchers try to consider the influence of the riverbank or riparian vegetation on the net shortwave solar radiation (see, e.g., Garner et al. 2017; Glose et al. 2017; Sinokrot and Stefan 1993), but then additional information about the parameters describing the shading effect is required. Since all those parameters may vary spatially and seasonally, the shading effect is usually not considered in practical applications. It is worth pointing out that not only the density of the vegetation but also the orientation of the channel plays an essential role in controlling solar radiation inputs; refer to Garner et al. (2017), Lee et al. (2012) and Li et al. (2012) for details.

The incoming shortwave radiation q_{si} may be measured directly by a pyranometer, and in practical applications, the historical data from the nearest meteorological station are usually used. However, as mentioned earlier, the “nearest” meteorological station may be sometimes far away from the measurement site. The example results for the net shortwave solar radiation heat flux q_s , calculated based on measured q_{si} from different meteorological stations for the Narew River (SET I): close to the site (*Meteo Choroszcz* ~ 12 km, *Meteo NPN* ~ 2 km) and far away from the site (*Meteo IGF UW* ~ 159 km, *Meteo WAW* ~ 168 km), are presented in Fig. 5a. The discrepancy in the results for the close and faraway stations is clearly visible. In the second example for the Świder River (SET IV), the measured values for q_{si} from three different stations (*Meteo Waw* ~ 26 km, *Meteo IGF UW* ~ 25 km and *Meteo Szpital Zachodni* ~ 42 km) were available. The results of q_s (see Fig. 5b) differ between each other but also they differ significantly, especially in the morning, from the q_s computed based on the q_{si} measured directly on site (*Meteo Świder Local* ~ 0.1 km). This difference may be the effect of the riverbank vegetation. Note that even if we measure the value of q_{si} on site, it is not without significance whether the station is located on the riverbank, and on which one, or in the middle of the channel.

In case q_{si} is not measured, it may be estimated for the given geographical location and time of the year and day, using one of the available formulae for the so-called theoretical incoming clear-sky solar radiation, based on the value of the solar constant (see, e.g., Allen et al. 1998; Carmona et al. 2014; Flerchinger et al. 2009; Lhomme et al. 2007). During the calculations, it is crucial not to forget the difference between the local time and solar time (see, e.g., Allen et al. 1998; Khatib and Elmenreich 2015; Lhomme et al. 2007). Otherwise, an unnecessary error may be introduced. Figure 5 presents the results of theoretical q_{si} calculations for the Narew and Świder rivers for

the selected data sets (SET I, SET III and SET IV) (the orange lines). However, such formulae do not include additional effects that reduce the shortwave solar radiation reaching the water surface like, for example, cloud cover. Therefore, a discrepancy between the results based on the theoretically calculated and measured values of q_{si} is readily seen. It may be huge for cloudy days. Indeed, the discrepancy may be reduced using the formula that includes cloud cover. An example is shown in Fig. 5c (pink line) for the Świder River (SET III). The result is much closer to the result determined based on the measurements from the local meteorological station (black line). However, one has to face the problem of rare availability of the information on cloud cover, and even if it is available, it may be a subject of the judgemental estimate of the observer. Note that cloud cover, also called cloudiness, is the portion of the sky cover that is attributed to clouds and measured in eighths (oktas) or per cents. More exact and devoid of subjunctives is the estimate of cloudiness with use of all-sky cameras.

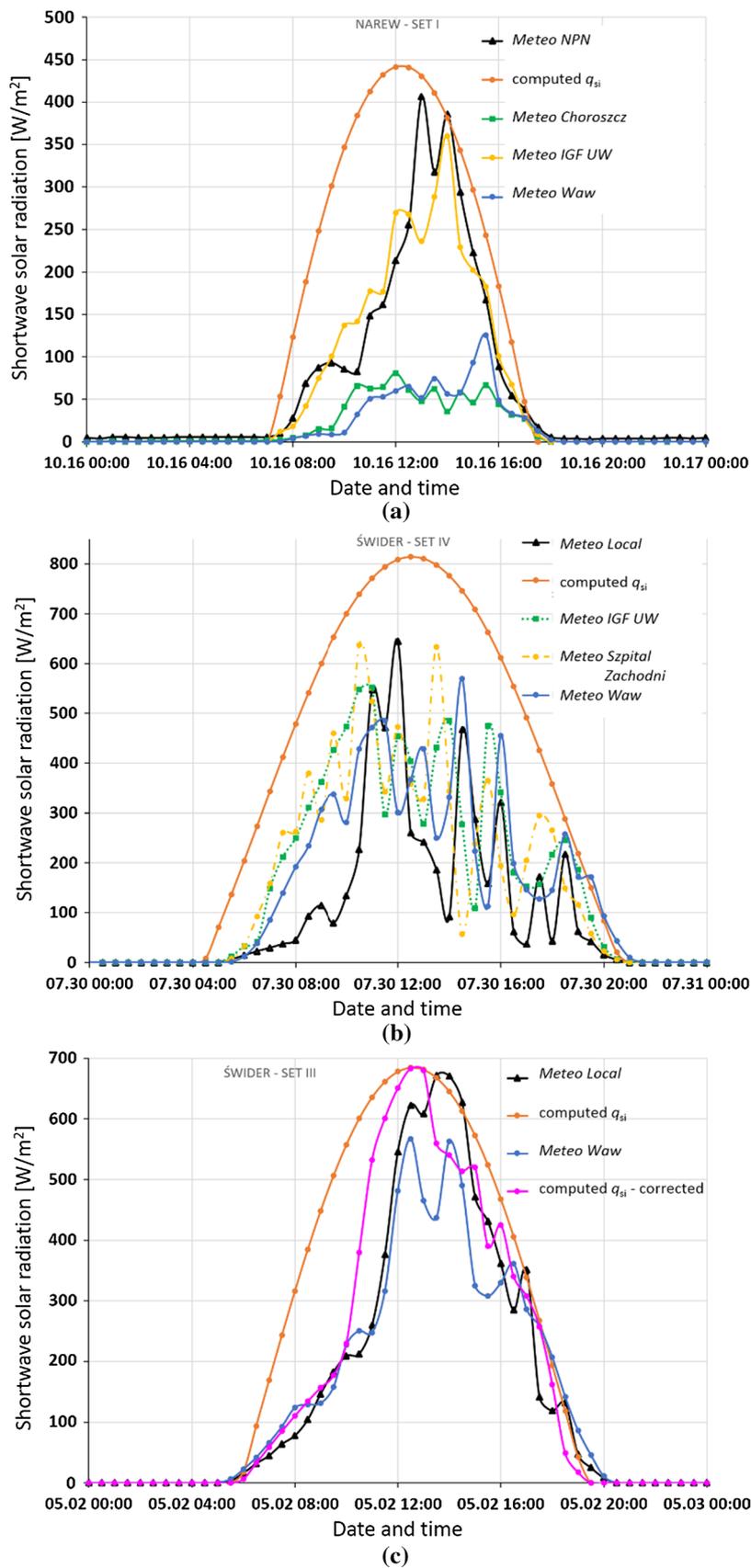
Consequently, the net shortwave solar radiation values obtained from measured or computed values of q_{si} may differ significantly. Obviously, the best will be the results based on the value of q_{si} measured directly on (or close to) the river site. In the case when the meteorological station is far distant from the site, it might be better to use the calculated value of q_{si} , but only if the cloudiness information is available. In other cases, all received values of the net flux will be very uncertain.

It is worth to note that other methods to estimate hourly solar radiation are also in use, pertaining to machine learning techniques or using a variety of empirical formulae. For example, Khatib and Elmenreich (2015) have proposed a model for predicting hourly solar radiation data using daily solar radiation averages. They have also made an overview of existing empirical formulae.

Longwave atmospheric radiation

The atmosphere as all terrestrial objects emits longwave radiation. The value of atmospheric longwave radiation mostly depends on the air temperature and varies between 30 and 450 W m⁻² (Wunderlich 1972). It could be measured directly by pyrgeometer, but while the measurements of shortwave solar radiation are relatively easily available from various meteorological stations, the measurements of longwave atmospheric radiation are unique. For the Narew River analysed case, the nearest meteorological station with such measurements has been found only at a distance of about 159 km away from the measurements site (*Meteo IGF UW*). The pyrgeometers are also more expensive than pyranometers, and their measurements have usually more significant errors (Choi et al. 2008). The net longwave atmospheric radiation heat flux (q_a) is therefore generally calculated using one of the formulae that are based on more readily available measured meteorological

Fig. 5 Net shortwave solar radiation heat flux calculated for: **a** the Narew River (SET I), based on the measured q_{si} data from different meteorological stations (*Meteo Choroszcz*, *Meteo NPN*, *Meteo IGF UW* and *Meteo Waw*), and based on computed theoretical incoming q_{si} ; **b** the Świder River (SET IV), based on the measured q_{si} data from different meteorological stations (*Meteo Waw*, *Meteo IGF UW*, *Meteo Szpital Zachodni*, *Meteo Świder Local*), and based on computed theoretical incoming q_{si} ; **c** the Świder River (SET III), based on the measured q_{si} data from *Meteo Waw* and *Meteo Świder Local* meteorological stations and based on computed theoretical incoming q_{si} with and without the inclusion of the correction for cloud cover



data. However, a large number of the proposed formulae for calculation of q_a make the choice of the best one very difficult.

All objects with a temperature higher than 0 K emit radiation. The total radiation energy sent by unit of the object surface is proportional to the fourth power of the absolute object temperature. Stefan Boltzmann's law (Boltzmann 1884) describes this relation:

$$\phi = \varepsilon \sigma T^4, \quad (3)$$

where ϕ [W m^{-2}]—the heat flux, ε [—]—the emissivity of the radiating body ($\varepsilon = 1$ for true blackbody), $\sigma = 5.67 \times 10^{-8}$ [$\text{W m}^{-2} \text{K}^{-4}$]—the Stefan–Boltzmann constant, T [K]—the absolute temperature in kelvin ($^\circ\text{C} + 273.15$).

The incoming longwave atmospheric radiation (q_{ai}) reaching the water surface can be determined for the clear-sky conditions using the modified Stefan–Boltzmann law as follows (see, e.g., Chapra 2008; Ji 2008):

$$q_{ai} = \varepsilon_a \sigma (T_a + 273.15)^4. \quad (4)$$

Then, the net longwave atmospheric radiation is equal to

$$q_a = q_{ai}(1 - r), \quad (5)$$

where ε_a —emissivity of atmosphere [—], r —reflection coefficient for longwave radiation (for water $r = 0.03$), T_a [$^\circ\text{C}$]—air temperature [$^\circ\text{C}$].

The biggest problem in applying Eqs. (4) and (5) is how to determine the value of atmospheric emissivity. The first empirical relation was proposed by Ångström (1915). It is usually calculated based on dew point temperature of air (Tang et al. 2004) or most commonly based on measurements of the air temperature and/or the actual vapour pressure (also called water vapour pressure in the air or air vapour pressure) e_a [mb = hPa] (see e.g., Anderson 1954; Brunt 1932; Brutsaert 1975; Crawford and Duchon 1999; Idso 1981; Idso and Jackson 1969; Iziomon et al. 2003; Prata 1996; Satterlund 1979; Swinbank 1963), measured close to the water surface (usually two metres above the ground):

$$\varepsilon_a = f(e_a, T_a). \quad (6)$$

Different relationships may be used here: empirical formulae, for example, the formula proposed by Brunt in 1932:

$$\varepsilon_a = a_1 + a_2 \sqrt{e_a}; \quad (7)$$

or physically based formulae, for example, the Brutsaert formula (1975):

$$\varepsilon_a = a_3 \left(\frac{e_a}{T_a + 273.15} \right)^{a_4}; \quad (8)$$

where a_1 [—], a_2 [$\text{mb}^{-1/2}$], a_3 [$\text{mb}^{-a_4} \text{K}^{a_4}$] and a_4 [—] are empirical coefficients. Originally proposed values by Brant are: $a_1 = 0.55$, $a_2 = 0.065$, and by Brutsaert: $a_3 = 1.24$, $a_4 = 0.14$. The actual vapour pressure may be expressed as:

$$e_a = e_{\text{sat}} \frac{R_h}{100}, \quad (9)$$

where R_h —air humidity [%], and e_{sat} is a saturation vapour pressure in the air [mb]. The saturation vapour pressure is commonly calculated using the relatively highly accurate empirical Magnus formula (Lawrence 2005; Magnus 1844):

$$e_{\text{sat}} = r_1 \exp \left(\frac{r_2 T_a}{T_a + r_3} \right), \quad (10)$$

where r_1 [mb], r_2 [—], r_3 [$^\circ\text{C}$] are empirical coefficients. Two sets of r coefficients are usually used in heat budget studies: 6.12, 17.27 and 237.3 (e.g., Chapra 2008; Raudkivi 1979) and 6.112, 17.67, 243.5 (e.g., Bolton 1980; Ji 2008), but many others had been proposed (refer to Alduchov and Eskridge 1996). However, different sets of r coefficients provide similar results for the considered air temperature values and do not affect the final values of the q_a as much as, for example, a_1 or a_2 coefficients. The second set of coefficients have been used in the paper.

Besides Eqs. (7) and (8), there are many other formulae for atmospheric emissivity calculations. The new methods are usually juxtaposed with oldest formulae, see, for example, Alados et al. (2012), Duarte et al. (2006), Flerchinger et al. (2009), Iziomon et al. (2003), Marthews et al. (2012) and Santos et al. (2011). The selected, used herein, formulae are summarised in Table 4. Note that the coefficients in particular relationships may also take different values and they may vary significantly by site. For example, Table 5 presents the selected values of a_1 and a_2 coefficients presented in formula (7). Since formula (7) is commonly used in water quality modelling, it has been calibrated for many sites around the world; see, for example, Berdahl and Martin (1984), Berger et al. (1984), FAO (1990), Heitor et al. (1991), Iziomon et al. (2003), Kjaersgaard et al. (2007), Monteith (1961), Sellers (1965), and Swinbank (1963). In summary, according to different studies (Alados et al. 2012; Elsasser 1942; Iziomon et al. 2003; Jiménez et al. 1987), the values for a_1 and a_2 coefficients range from 0.34 to 0.7 and from 0.023 to 0.110, respectively. Since the clear night-time sky is more emissive than the daytime, some authors, for example, Berdahl and Fromberg (1982) and Li et al. (2017), suggest using different coefficients for the daytime and night-time, or propose a correction depending on the hour of the day (Alados-Arboledas and Jimenez 1988). Above all, it is very important to pay attention to the units in which the vapour pressure is given, and then, correctly convert the values of the coefficients if necessary. It is surprising how many models suffer from simple conversion error.

Different values may be obtained by applying available formulae for the longwave atmospheric radiation heat flux (see example in Fig. 6). It is worth to point out

Table 4 Selected formulae for atmospheric emissivity

Reference/source	Formula	Remarks
Clear-sky conditions		
Brunt (1932), Monteith (1961), Swinbank (1963), Sellers (1965), Berger et al. (1984) and Heitor et al. (1991)	$\epsilon_a = a_1 + a_2 \sqrt{e_a}$	With different values for a_1 and a_2 coefficients presented in Table 5
Swinbank (1963)	$\epsilon_a = a_8 T^2$	$a_8 = 9.365 \times 10^{-6}$
Brutsaert (1975)	$\epsilon_a = a_3 \left(\frac{e_a}{T_a + 273.15} \right)^{a_4}$	$a_3 = 1.24, a_4 = 0.14$
All-sky conditions		
Abramowitz et al. (2012)	$\epsilon_a = a_5 e_a + a_6 (T_a + 273.15) - a_7$	$a_5 = 3.1, a_6 = 2.84, a_7 = 522.5$
Sridhar et al. (2002)	$\epsilon_a = a_3 \left(\frac{e_a}{T_a + 273.15} \right)^{a_4}$	$a_3 = 1.31, a_4 = 0.14$
Cloudy sky conditions (corrections)		
Crawford and Duchon (1999)	$\epsilon_a^{cld} = (1 - s) + s \epsilon_a$	$s = q_{si}/q_{si0}$ —solar index, $c = (1 - s)$ —fractional cloud cover, q_{si} —measured solar radiation, q_{si0} —estimated theoretical solar radiation for clear-sky condition
Lhomme et al. (2007)	$\epsilon_a^{cld} = (1.37 + 0.34s) \epsilon_a$	
Wunderlich (1972)	$\epsilon_a^{cld} = (1 + 0.17c^2) \epsilon_a$	

Table 5 Values of a_1 and a_2 coefficients in Brunt's formula (7) for clear-sky atmospheric emissivity used in Figs. 6 and 7

Reference	a_1 []	a_2 [mb ^{-1/2}]
Brunt (1932)	0.55	0.065
Monteith (1961)	0.53	0.065
Swinbank (1963)	0.64	0.037
Sellers (1965)	0.61	0.048
Berger et al. (1984)	0.66	0.040
Heitor et al. (1991)	0.59	0.044

the difference between results calculated for two different datasets from two metrological stations (close to and far away from the site). Note the difference between the calculated and measured (violet dots) values of longwave atmospheric radiation that was available for one of the considered meteorological stations (Fig. 6b). This difference may be larger or smaller depending on the considered part of the day which is the result of varying sky conditions. In the case of no-clear-sky conditions, the value of atmospheric longwave radiation increases.

Similarly to the shortwave solar radiation, the mentioned formula for atmospheric emissivity may be corrected, to better predict the longwave atmospheric radiation, by considering the cloud cover. Many correction algorithms have been proposed (see, e.g., summary presented by Flerchinger et al. (2009), and many authors claim that those algorithms improve the results. Since they usually require the information on cloud cover which is rarely available, this makes the calculation even harder. Surprisingly, Abramowitz et al. (2012) showed that such correction factors might be redundant since generally

they did not improve the results. They also proposed a new formula for the emissivity of atmosphere for all-sky conditions as follows:

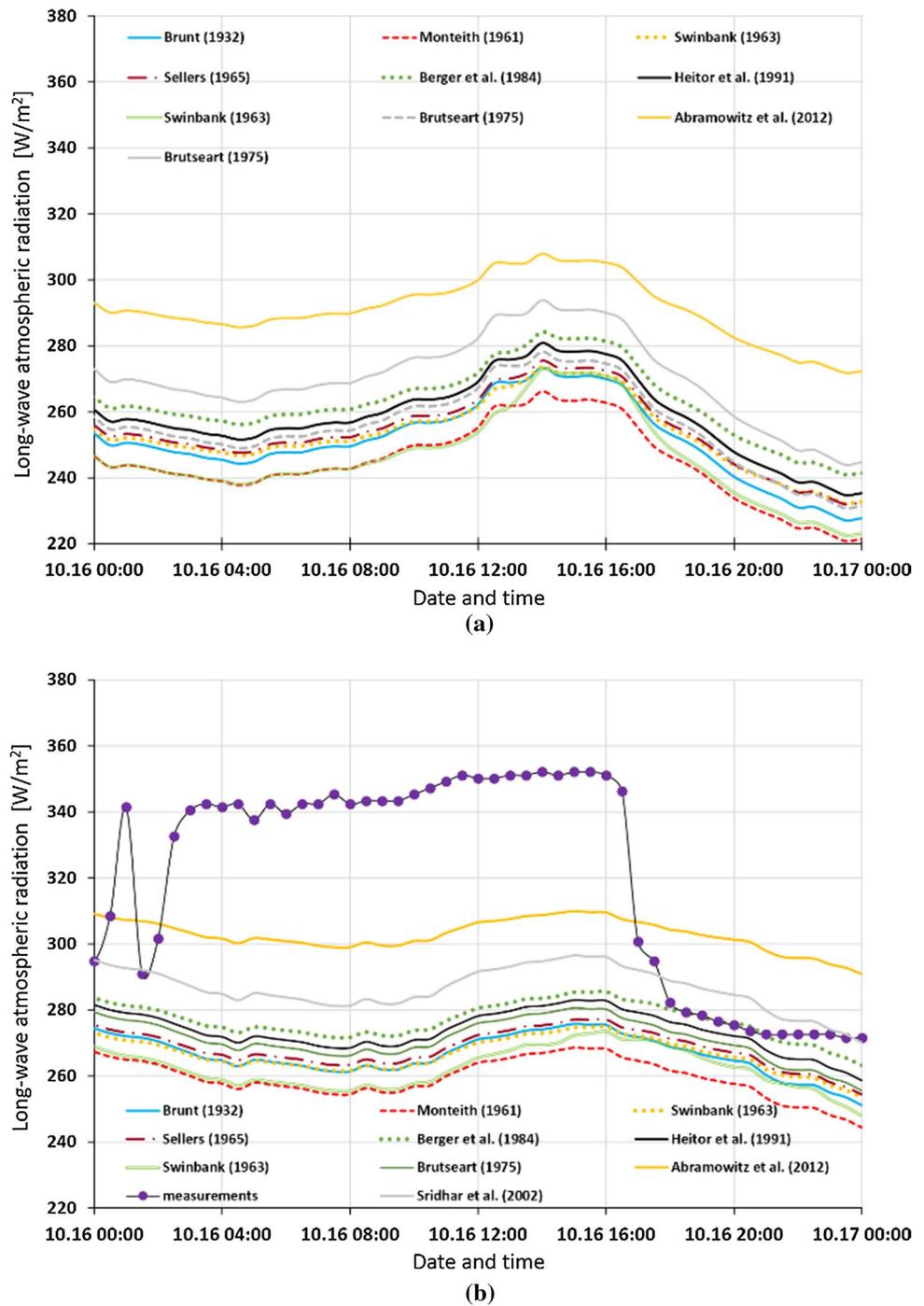
$$\epsilon_a = a_5 e_a + a_6 (T_a + 273.15) - a_7, \quad (11)$$

which actually in case of the considered data gives the results closer to the measured ones under cloudy sky condition (see Fig. 6b, yellow line). a_5 , a_6 and a_7 are empirical coefficients, originally equal to 3.1, 2.84 and 522.5, respectively. A similar formula for all-sky condition (without the necessity to provide information about the cloudiness) based on the Brutsaert formula (8) has been proposed by Sridhar and Elliott (2002). He suggested the value for the a_3 coefficients to be equal to 1.31 instead of 1.24 to obtain the better fit to the large range of all cloud condition data. One can see that for the considered case, the Sridhar formula is better than all formulae for clear-sky conditions, but the Abramowitz formula works much better than the formula proposed by Sridhar (Fig. 6b). In cases, when solar radiation measurements are available (which is much more common than the availability of cloudiness information), the problem with cloud cover estimation may be avoided using the solar index defined as:

$$s = \frac{q_{si}}{q_{si0}}, \quad (12)$$

where q_{si} is the measured shortwave solar radiation, and q_{si0} is the theoretically calculated (estimated) value of the solar radiation for clear-sky condition (see the previous section). Crawford and Duchon (1999) and Lhomme et al. (2007) proposed the formulae for the emissivity of the atmosphere under cloudy sky conditions based on the solar index as follows:

Fig. 6 Longwave atmospheric radiation heat flux calculated for Narew River (SET I) with different formulae for atmospheric emissivity: **a** based on data from the nearest meteorological station *Meteo NPN* (**b**); based on data from *Meteo IGF UW*



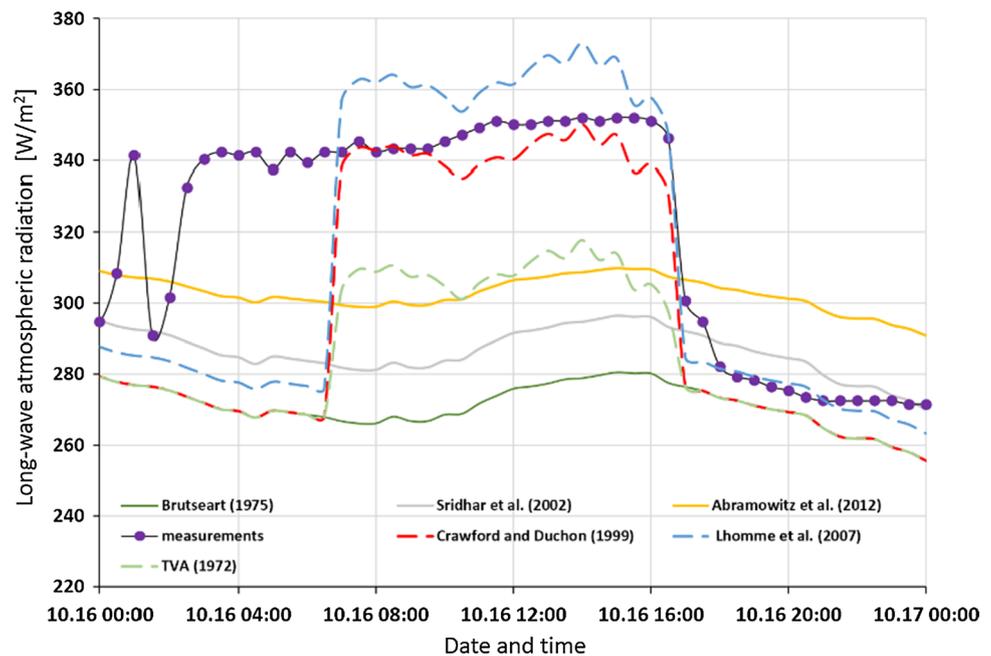
$$\epsilon_a^{cld} = (1 - s) + s\epsilon_a, \tag{13}$$

$$\epsilon_a^{cld} = (1.37 + 0.34s)\epsilon_a, \tag{14}$$

where the ϵ_a^{cld} is the emissivity of the atmosphere under cloudy sky conditions and ϵ_a is the emissivity of the atmosphere under clear-sky conditions. Both of them have been applied to correct the results presented in

Fig. 6b. As could be seen in Fig. 7 (red line—Crawford and Duchon (1999); blue line—Lhomme et al. (2007)), they improve the results for the cloudy sky conditions during the daytime, but since the information about the solar radiation during the night is not available, they cannot be used to correct the results during the night. In the face of lack of information about the solar index during

Fig. 7 Longwave atmospheric radiation heat flux calculated for the Narew River (SET I) with different algorithms for cloudiness correction with the meteorological data from *Meteo IGF UW*. Formulae for cloudy sky conditions are based on Brutsaert formula (Brutsaert 1975)



night-time, Lhomme et al. (2007) proposed to use for the night period, the averaged value of the solar index that appeared between 14:00 and 16:30 the day before. In some studies, an average value for the whole previous day is used. The average values of the solar index, from a day or the end of the day measurements, may help to improve the results. For a particular case, however, with clear-sky conditions during the night (as in the analysed case), the results may be even worse. While using the end of the day measurements of the solar radiation, note that the calculation of the solar index may be troublesome due to the errors of measuring small values of solar radiation. To sum up, when the information about the cloudiness is lacking, the best is to use the all-sky conditions algorithms to calculate the atmospheric emissivity.

Additionally to the cloud sky correction, other information, for example, information about site elevation, may be considered (e.g., Deacon 1970; Marks and Dozier 1979), but it insignificantly affects the results (Flerchinger et al. 2009). Also, since all objects emit the longwave radiation, the longwave radiation from topography and vegetation (see, e.g., Leach and Moore 2010) may be taken into account. A detailed review of various relationships based on different input parameters, with various possible corrections, may be found in (Vall and Castell 2017). But it is worth to remember that more corrections and parameters are taken into account, more input data are necessary. Overall, if no measurements for incoming longwave radiation and no additional information are available, the clear-sky or all-sky condition algorithms are the first and the only possible assumption for incoming longwave radiation in practical applications.

Processes dependent on water temperature

Longwave water back radiation

Similarly to all terrestrial objects, water also emits longwave radiation, which can provide a significant contribution to heat loss from the water surface. The value of outgoing water longwave radiation heat flux usually varies from 300 to 500 W m^{-2} (Deas and Lowney 2000), and assuming that water temperature is known, it is relatively easy to estimate. Similarly to longwave atmospheric radiation, it may also be computed using the Stefan Boltzmann law (see Eq. (3)):

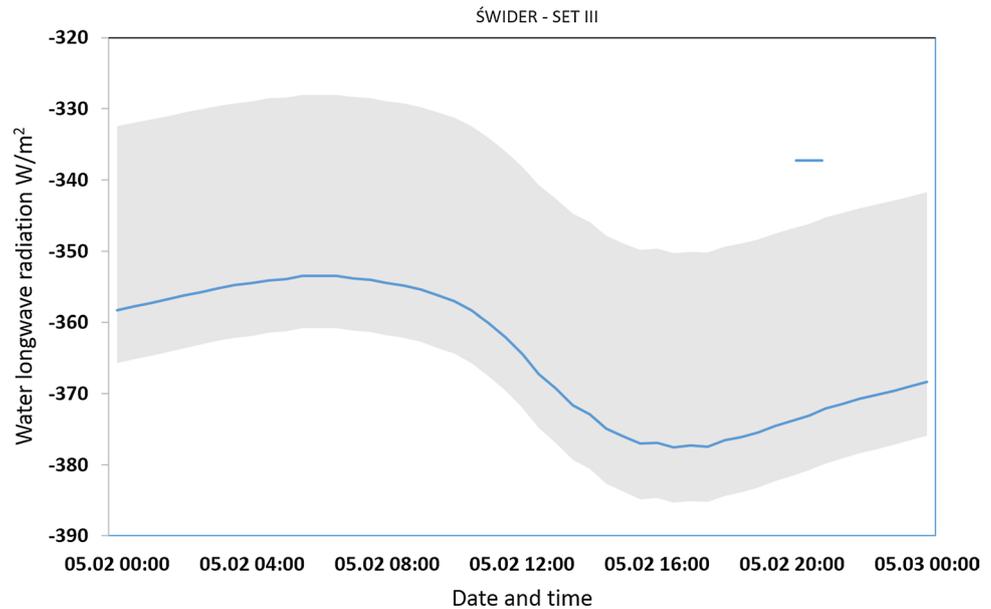
$$q_b = \varepsilon_w \sigma (T_w + 273.15)^4, \quad (15)$$

where ε_w —emissivity of water [–]. The water emissivity depends on water transparency and surface smoothness. Often, its value is assumed to be equal to 0.97 (e.g., Chapra 2008; Ji 2008) or to 0.96 (e.g., Brewster 1992), but the values between 0.9 and 0.99 are also applied. Figure 8 shows the range of values that water longwave radiation heat flux can take depending on the assumed water emissivity coefficient for the selected example of the data set (SET III). $\varepsilon_w = 0.97$ has been used for most calculations performed in the paper.

Evaporation and condensation

Without taking into account the radiation fluxes, the evaporation and condensation heat flux term has the most significant contribution in the total heat budget. However, its value significantly varies among sites and seasons. At the same

Fig. 8 Range of values (a grey area) that water longwave radiation heat flux can take depending on the assumed water emissivity coefficient (values between 0.9 and 0.99) for selected data set (SET III), and longwave back water radiation heat flux (blue line) calculated for water emissivity coefficient equal to 0.97



time, especially within a short time step (like the hourly rate of evaporation), it is difficult to measure evaporation from water in open channels. Commonly accepted methods use the evaporation pan to measure the evaporation rate, but since the hourly evaporation may be of the order of a millimetre or less, very accurate precision is required. Otherwise, the results may become uncertain (Tan et al. 2007). Finally, the evaporation and condensation term in Eq. (1) is usually estimated using one of many indirect methods available in the literature.

Various types of approaches to estimate evaporation are available, for example, radiation-based models, mass or momentum transfer models, temperature-based models and models that used artificial intelligence or mixed-type methods. Comparisons of various approaches may be found, for example, in Ali et al. (2008), Rosenberry et al. (2007), Tan et al. (2007) and Winter et al. (1995). Most of them, however, have been derived or calibrated for long time steps (to estimate daily or monthly evaporation/condensation), and it should be noted that they may not work accurately enough for the hourly time step analysis.

For temperature modelling in open channels and reservoirs, commonly used is the mass transfer approach (see, e.g., Caisie et al. 2007; Sinokrot and Stefan 1993) based on Dalton equation (Dalton 1802). The evaporation is proportional to the difference between the saturation vapour pressure at the water surface— e_s [mb] and the actual vapour pressure of the air above— e_a [mb], and the proportionality factor depends on the value of the wind speed. The heat flux caused by the evaporation and condensation term may be then defined as:

$$q_e = f(u)(e_s - e_a), \quad (16)$$

where $f(u)$ [$\text{W m}^{-2} \text{mb}^{-1}$] is so-called wind speed function (or wind function). The saturation vapour pressure at the water surface is equal to:

$$e_s = r_1 \exp\left(\frac{r_2 T_w}{T_w + r_3}\right), \quad (17)$$

where r_1 [mb], r_2 [–], r_3 [°C] are the same coefficients as in Eq. (10).

The wind speed function is usually a linear or quadratic function of the wind speed and generally may be written as (Ji 2008):

$$f(u) = b_0 + b_1 u + b_2 u^2, \quad (18)$$

where b_0 , b_1 and b_2 are empirical, usually site-specific coefficients, which may take different values. In practical applications when calibration of the wind speed function is not possible, one of many empirical relations available in the literature has to be chosen. However, it is worth to mention that most of the available formulae have been estimated for lakes or reservoirs and later adapted to rivers and streams. The wind speed function coefficients are sometimes estimated based on the meteorological input data. For example, coefficient b_0 (describing the evaporation rate in case of lack of or minimal wind) may be related to the difference between water and air temperature (see, e.g., Arifin et al. 2016; Czernuszenko 1990; Harbeck 1958).

Note that there are different ways to define the heat flux q_e . Depending on the definition, in some cases, the wind function must be divided by latent heat of water and water density to be the same as defined by Eqs. (16) and (18). Above all, it is again essential to notice units for $f(u)$ function. Different units may be used for heat flux and vapour pressure which results in different units for $f(u)$ and finally

different values of the b_0 , b_1 and b_2 coefficients. Similarly to the coefficients in formulae for atmospheric emissivity calculation, also for coefficients in function $f(u)$, we may encounter mistakes derived from simple units calculation errors.

Some researchers, for example (Czernuszenko 1990; Hughes et al. 2011; Trabert 1896), use the relations for wind function that neglect the evaporation in the absence of wind, while others argue that it should also be considered then. Most researchers favour the second opinion. Finally, two forms of the wind formulae are usually met in the literature: linear wind speed function when $b_2=0$ (McJannet et al. 2012; Meyer 1928; Ryan 1973) or quadratic one when $b_1=0$ (Ahsan and Blumberg 1999; Arifin et al. 2016; Brady et al. 1969; Chapra 2008; Ji 2008). Figure 9 shows how the relations $f(u)$ may look like, including the one proposed by Trabert (1896) using square root relationship: $f(u) = 11.25\sqrt{u}$ (dotted line). The values of the coefficients for other wind speed formulae are summarised in Table 6. The general pattern is that wind increases the evaporation rate, but the

value of the proportionality coefficient depends on the chosen formula.

The impact of the choice of the wind function upon the evaporation and condensation heat flux is presented in Fig. 10. As an example, data set (SET III) for the Świder River has been used, with the input data from the local meteorological station (Fig. 10a), and from the station *Meteo Waw* (Fig. 10b). The differences are clearly visible. The results are influenced not only by choice of the wind function formula but also by the wind speed value (grey dots). Since the Świder River is sheltered from the wind by the surrounding trees and bushes, the wind speed values measured locally are much lower than the ones observed in the nearby meteorological stations. Consequently, the evaporation heat fluxes for local input data are generally lower. The local values of the wind speed may often be far away from those measured by the nearby meteorological station. Additionally, it is not meaningless at which height above the surface the wind speed is measured. For rivers, it should be measured typically 2 metres above the surface

Fig. 9 Selected formulae for wind speed function depending on the wind speed value

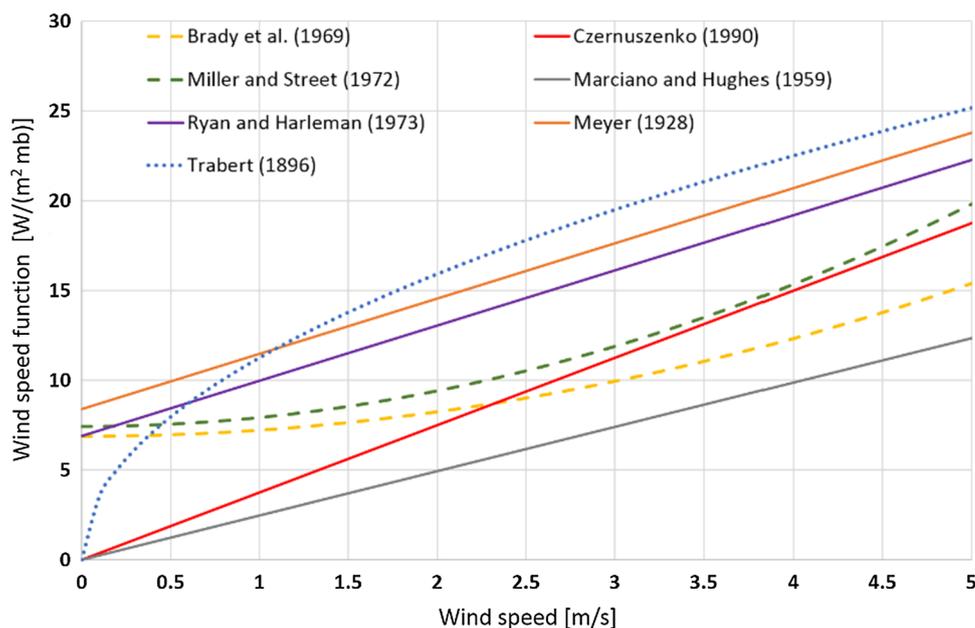
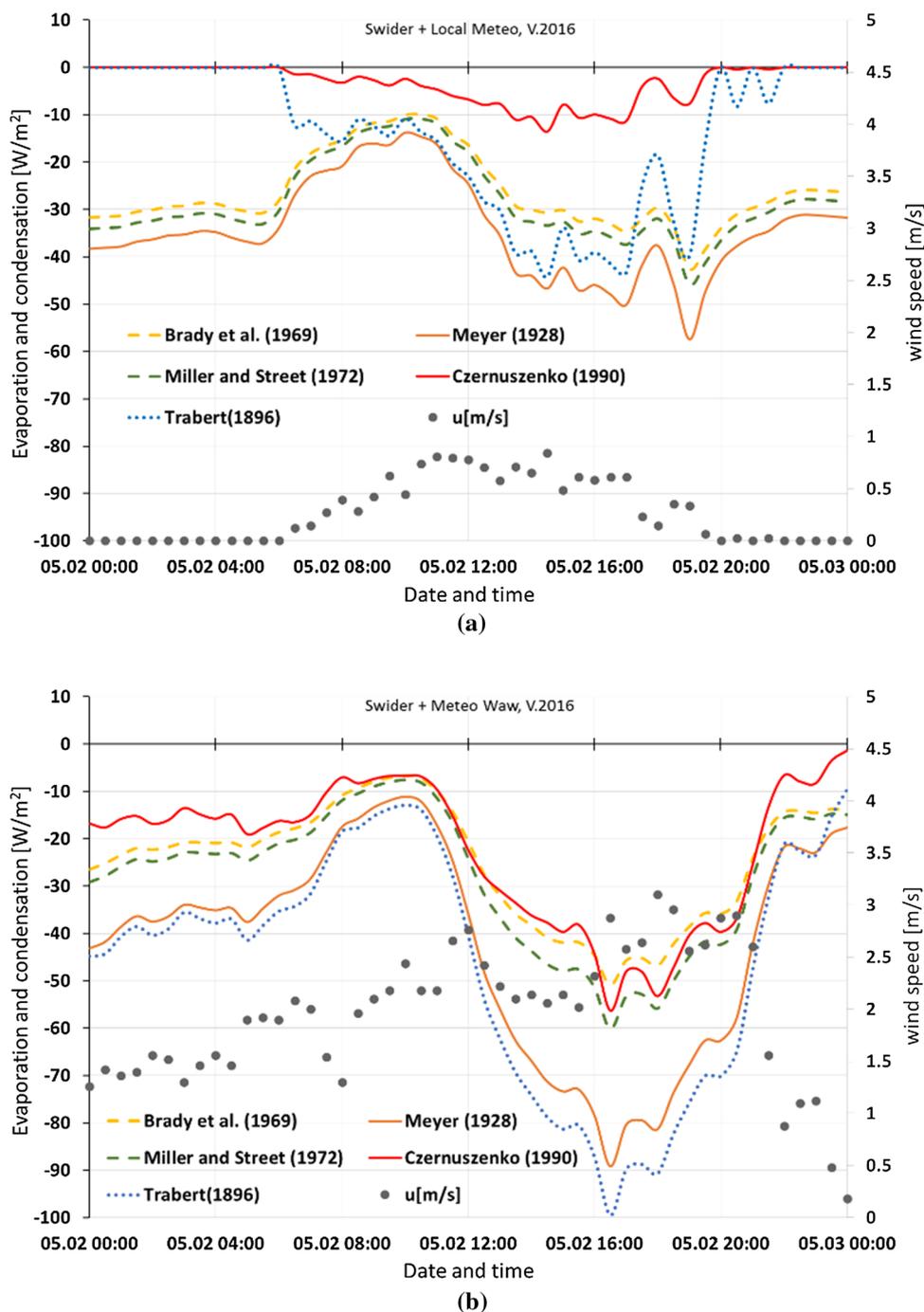


Table 6 Coefficients in the wind speed function empirical formulae used in this paper

The wind speed function formula $f(u)[W m^{-2} mb^{-1}]$	b_0 [$W m^{-2} mb^{-1}$]	b_1 [$W m^{-3} mb^{-1} s$]	b_2 [$W m^{-4} mb^{-1} s^2$]
Brady et al. (1969), Ahsan and Blumberg (1999), Arifin et al. (2016) and Ji (2008)	6.9	0	0.34
Miller and Street (1972)	7.42	0	0.49
Czernuszenko (1990)	0	3.75	0
Marciano and Harbeck (1952)	0	2.47	0
Ryan (1973)	6.9	3.07	0
Meyer (1928)	8.4	3.07	0

Fig. 10 Evaporation and condensation heat flux calculated for the Świder River (SET III) with different formulae for wind speed function, based on input data from two different meteorological stations: **a** local meteorological station—*Meteo Świder Local*, and **b** *Meteo Waw* station



(sometimes 1.5 m), and for lakes and reservoirs, it is usually measured at 7, 8 or 10 metres above the surface. The wind speed generally increases with the height. Although it is possible to recalculate the measured value of the wind speed using log or power wind profile assumption (see, e.g., Allen et al. 1998), this recalculation may be problematic, especially for rivers when measurements are taken far away from the site and for other altitudes. The wind speed values are finally likewise uncertain as wind function coefficients.

According to Winter et al. (1995), errors in the input data may cause serious error in the final evaporation rate value. For example, a few degrees error in measured (or assumed) surface water temperature may cause 40% error in the evaporation rate calculation. In water temperature modelling, usually the depth-averaged values of water temperature (or cross-sectionally averaged values in 1D models) are used and the surface water temperature for sure differs from the depth-averaged values. Although the difference for rivers is usually small, it strongly depends on the local conditions.

To conclude and to obtain accurate results, the best is to calibrate the selected wind speed formula for local conditions using data from local measurements. However, since evaporation measurements are troublesome, and local meteorological data are usually not available, we have to bear in mind that whatever formula we use for calculation of evaporation and condensation heat flux, the results will be very uncertain.

Conduction and convection

Conduction is the process of heat exchange between bodies of different temperatures, which are in direct contact. It consists in transmitting the kinetic energy of the chaotic movement of particles as a result of their collisions.

The process leads to the equalisation of temperature between the bodies. Convection is a process of heat transfer associated with the macroscopic movement of matter. Both processes take place at the border of water and air and can be described using the equation resulting from the Bowen ratio (Bowen 1926; Ji 2008):

$$B = \frac{q_h}{q_e} = C_B \frac{p_a T_w - T_a}{p_0 e_s - e_a}, \quad (19)$$

where B —Bowen ratio [–], C_b —Bowen coefficient (approximately $C_b = 0.62 \text{ mb}^\circ\text{C}^{-1}$), p_0 —reference air pressure (= 1013 mb).

$$q_h = C_b \frac{p_a}{p_0} f(u)(T_w - T_a), \quad (20)$$

Usually, the same wind speed function as for evaporation and condensation heat flux term is used, and the same problems and similar large uncertainty connected with the empirical coefficients and wind speed velocity values apply here. However, the value of the conduction and convection heat flux is usually small compared to the values of other heat fluxes, and its impact on the final results is smaller. Similarly to Fig. 10, Fig. 11 presents the results of computations of the conduction and convection heat flux using different wind function formulae for dataset SET III for the Świder River.

Heat flux with the omission of independent of water temperature terms

The analyses carried out indicate clearly that many factors affect the uncertainty of calculations of heat exchange at the water–air interface, and to obtain the accurate results, many input data are required. In the case of modelling of thermal pollution spreading in rivers, the number of these factors and required input data may be reduced. Since in practical applications we are usually interested in a possible change in natural water temperature caused by artificial heat

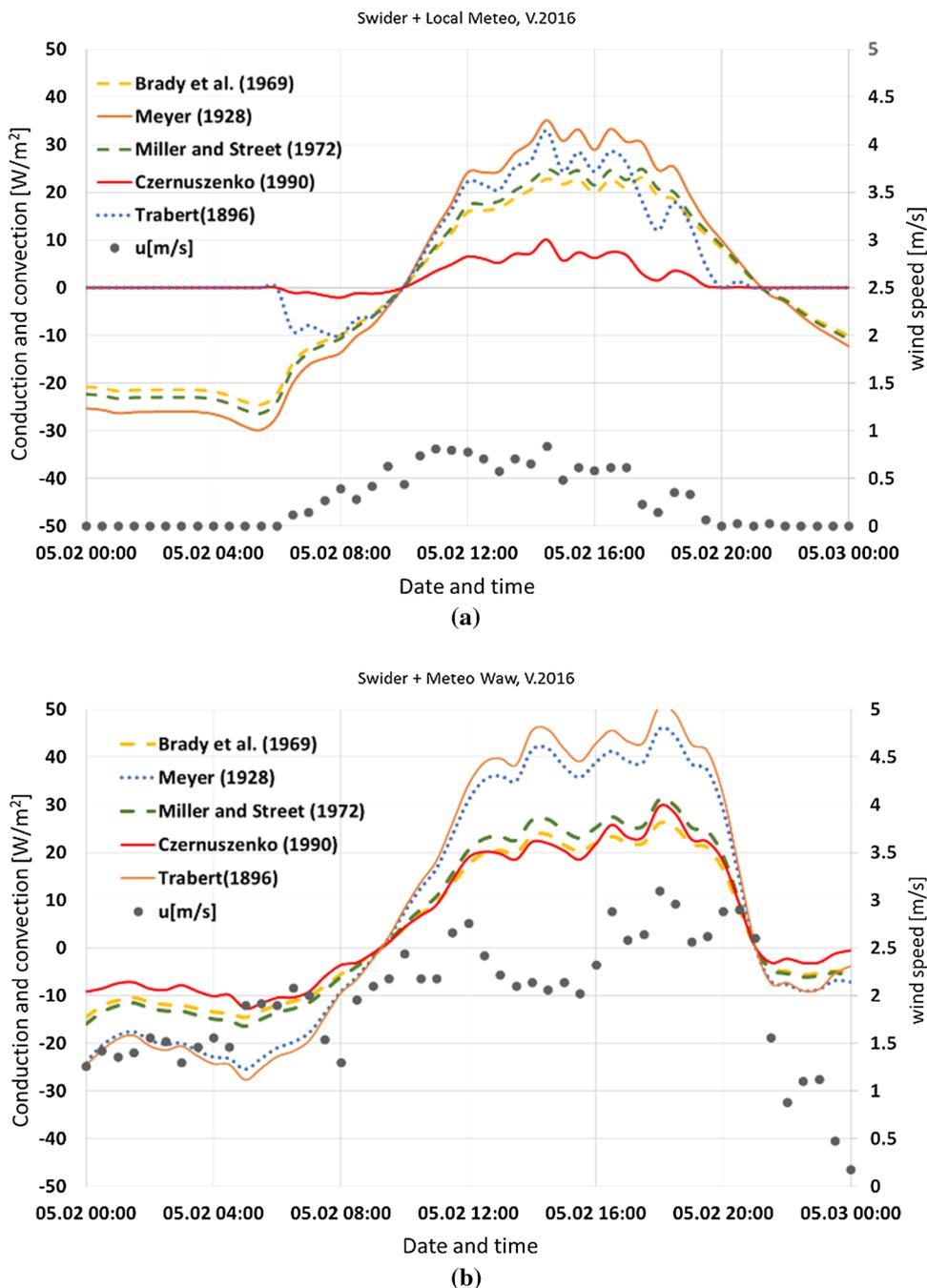
sources, it is useful to compute the temperature change, ΔT_w (difference between actual water temperature and ambient water temperature), instead of the actual water temperature (T_w) itself (see, e.g., Joss and Resele 1987; Kalinowska and Rowinski 2012; Kalinowska et al. 2012). This simplifies the determination of possible heat fluxes resulting from the heat exchange between water and its surrounding environment. All terms that do not depend on water temperature may be omitted during the calculation. They have the same value for different water temperatures and simply counterbalance each other while computing the temperature difference (ΔT_w). In the case of heat exchange with the atmosphere, the shortwave solar radiation and longwave atmospheric radiation terms may be disregarded. Although they significantly affect the natural water temperature T_w (causing large daily changes), they do not affect the temperature increase ΔT_w caused by external factors, such as the discharges of heated water. The net heat flux (Q_A) may be then simplified to:

$$Q_A = -q_b \pm q_e \pm q_h. \quad (21)$$

Finally, while omitting the terms that do not depend on water temperature, acquiring various meteorological input data and problems with the estimation of several factors (e.g., atmospheric emissivity) may be avoided. Unfortunately, the calculation of other factors like the wind speed function will still be problematic. However, in most cases, i.e., while the temperature of the discharged warm water is not very high compared to the temperature of ambient water, it may be assumed that all remaining heat fluxes in Eq. (21) influence both ambient and the warm water similarly. While calculating the water temperature difference, those terms also practically disappear.

For instance, one may calculate the value of the temperature change caused by longwave back water radiation heat flux (which is the most significant water temperature-dependent term at the water–air interface) for natural and heated water temperatures. The temperature of warm water discharged from industrial facilities is usually a few degrees higher than the ambient water temperature. In the mid-field zone, after preliminary vertical mixing, the water temperature is usually not higher than 3 °C. Figure 12 shows the difference (ΔT) between the temperature change for natural water temperature (ΔT_{w0}) and temperature change for heated water temperature—for the assumed 3 °C (ΔT_{w3}) and 7 °C (ΔT_{w7}), respectively—within a half-hour period. Since the value of longwave back water radiation heat flux also depends on the water emissivity coefficient (see, Eq. (15)), additionally, the temperature change for natural water temperature, using different assumed water emissivity coefficients, has been computed. The difference (ΔT) between the temperature change for natural water temperature computed with default emissivity value 0.97 (ΔT_{w0}) and the temperature change for temperature by 3 °C higher than natural

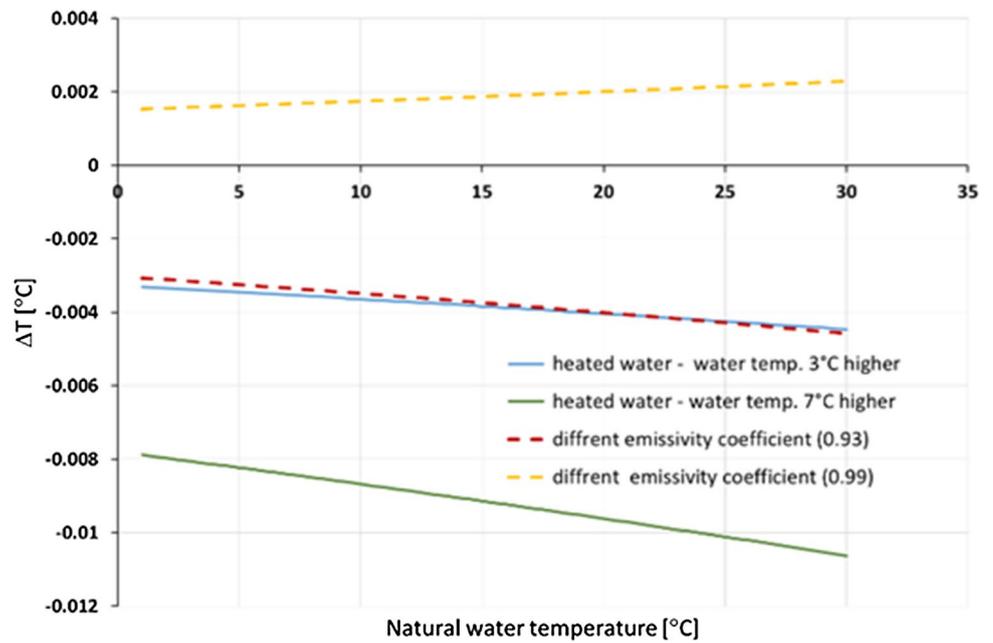
Fig. 11 Conduction and convection heat flux calculated for the Świder River (SET III) with different formulae for wind speed function, based on input data from two different meteorological station: **a** local meteorological station—*Meteo Świder Local*, and **b** *Meteo Waw* station



water temperature is similar to the difference (ΔT) when we calculate the longwave back water radiation heat flux for natural water temperature with emissivity coefficient equal to 0.93. The difference grows with the water temperature; however, it is of the order of 0.001 °C. For the extreme case (natural water temperature equal to 25 °C) when the ambient water is colder than the heated water by 7 °C, the difference reaches only 0.01 °C.

To illustrate the influence of the inclusion of the heat exchange term on the results of thermal pollution modelling in the mid-field zone, the real case study exemplary results have been analysed. The detailed description of the case study may be found in Kalinowska et al. (2012). The River Mixing Model (RivMix), developed by Author, computing the temperature change, has been used (Kalinowska and Rowiński 2008) to predict the water temperature increase caused by heated water discharged

Fig. 12 Difference in temperature change caused by longwave back water radiation when the water temperature is 3 or 7 °C higher than natural water temperature (solid lines) or when different water emissivity coefficients: 0.93 or 0.99 are used (dashed line). Water depth was set to $H=1.8$ m, as averaged water depth in case of the Narew River



from a designed power plant. Two variants of computation have been performed for the examples of chosen (least favoured) conditions. The first variant has been treated with the heat flux caused by water–air heat exchanges included in the calculations (see Fig. 13a), the second one without the heat flux inclusion. After 1 h, the observed difference between the two variants is of the order of 10^{-2} (see Fig. 13b). Such a difference is relatively small

compared to the other sources of errors performed while solving the 2D heat transport equation numerically. For instance, the huge source of uncertainty is the dispersion coefficients estimation. Those coefficients are very difficult to estimate, and therefore, usually different scenarios with various possible values of coefficients are analysed. However, the difference between such scenarios results may be much higher, up to 1 °C (see Kalinowska and Rowinski

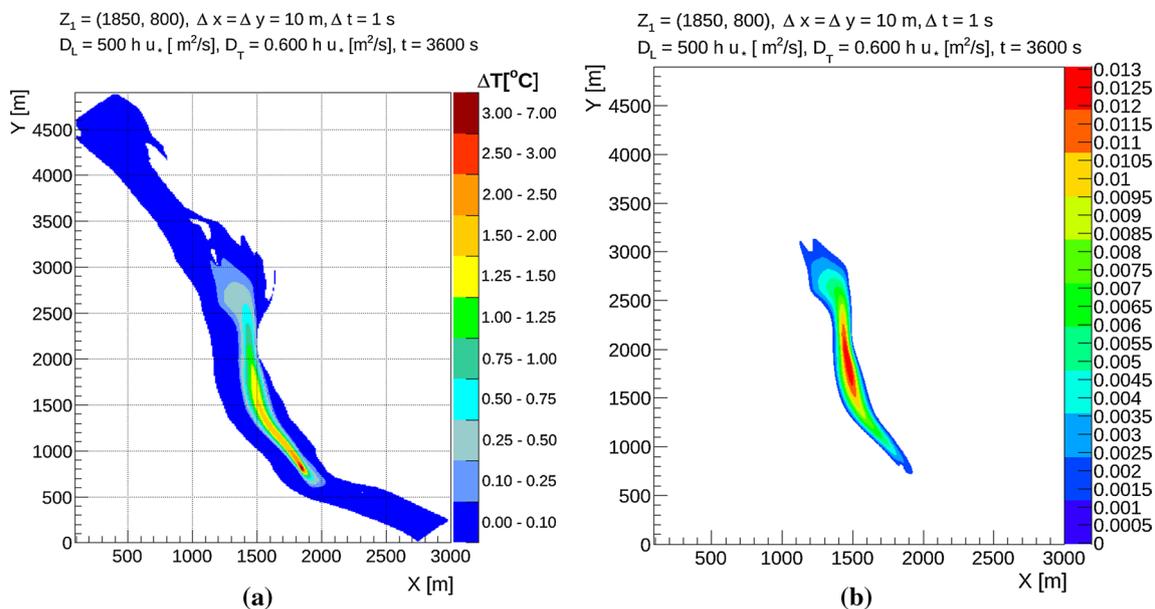


Fig. 13 Examples of results for the real case study (see details in Kalinowska et al. 2012). **a** Predicted temperature increase (ΔT) below the discharge of cooling water from a designed power plant on the Vistula River with the heat flux caused by water–air heat exchange

included in the calculations. **b** The difference in the predicted temperature increase between the results that take into account and results that omit the water–air heat exchange in calculations

2012), compared to the difference obtained with and without the included heat exchange with the atmosphere. Other sources of uncertainty have been analysed in (Kalinowska and Rowinski 2012).

Conclusions

In the paper, for thermal pollution modelling applications, it is recommended to use models that compute the temperature change instead of the actual river temperature whenever it is possible. Such approach reduces the amount of necessary input data and finally the computation error. In most cases in the mid-field zone, omission of all terms related to heat exchange with the environment including the heat exchange with the atmosphere is recommended. Not perfect input data may in some cases introduce much larger error to the final results than just simple omission of the heat fluxes terms. The problem is especially important in practical cases when we deal with limited and not ideal data. We of course fully realise that in some applications it is necessary to include the heat exchange with the atmosphere and/or other heat fluxes. For example, after full mixing between the heated water and the river water, water temperature change depends only on the heat exchange with the environment. Another example is the heat exchange with the riverbed and banks, which is negligible for most practical applications since it is small compared to the value of heat exchange between the water surface and atmosphere and moreover it is also very uncertain due to its variability and complexity. However, when the source of heated water is located near or on the riverbank, it may turn out to have significant value. Also, most of the possible heat sources will not be significant for large, deep rivers, but much more important for very shallow streams. Since the situation very much depends on the considered case and timescale and space scale, to decide whether 1, 2 or 3D approach is appropriate and which additional heat exchange processes should be taken into account, it is very important to make the proper analyses before calculation or applying of any model. First, the expected outcome together with the appropriate timescale and space scale should be defined. Next, all affecting processes should be analysed subject to their significance and the availability of necessary input data, but also taking into account all other errors that may be committed during the calculations. Since the necessary effort to be taken to provide detailed input data (even the perfect one) for computing the heat exchange with the atmosphere (or environment) is not worth the final outcome in many practical cases (while taking into account other unavoidable committed errors).

In the cases where the heat exchange with the atmosphere estimation is necessary, it is important to bear in mind that its estimation is based on empirical formulae that depend on

many uncertain parameters. To be precise, it will be necessary to measure many parameters directly on the site taking into account their space and time dependence to adjust the applied empirical formula to the current conditions. Although such measurements are usually possible for research purpose, in practice, we cannot take advantage of specially planned experiments, whether laboratory or field, allowing to measure a sufficient amount of data. In most cases for various reasons, there are not enough data to perform all the necessary calculations and prediction must be done based on existing historical, often limited and incomplete, sometimes also inaccurate data, as has been shown using two case studies presented in the paper. Therefore, in practical applications, heat exchange with the atmosphere estimation is full of judgements and extremely subjective. The most problematic to estimate is the wind speed function and atmospheric emissivity formulae. The most fragile to local conditions are measured short-wave solar radiation and wind speed value. The analysis and detailed description of particular processes involved in heat exchange with the atmosphere provided in the paper may be useful for practitioners.

Acknowledgements This study has been partly supported by the grant IP2012 028772 from the Polish Ministry of Science and Higher Education and supported within statutory activities No. 3841/E-41/S/2017 of the Ministry of Science and Higher Education of Poland. Author would like to thank Paweł Rowiński for his valuable suggestions and comments; Agnieszka Rajwa-Kuligiewicz for sharing water temperature data; Krzysztof Markowicz for sharing data from IGF UW Meteorological observatory; Małgorzata Krasowska and Piotr Banaszuk for sharing data from Meteo Choroszcz; and the authority of the Narew National Park for sharing data from Narew National Park Weather Station.

Compliance with ethical standards

Conflict of interest The author states that there is no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abramowitz G, Pouyanné L, Ajami H (2012) On the information content of surface meteorology for downward atmospheric long-wave radiation synthesis. *Geophys Res Lett* 39:L04808. <https://doi.org/10.1029/2011GL050726>
- Ahsan AQ, Blumberg AF (1999) Three-dimensional hydrothermal model of Onondaga Lake, New York. *J Hydraul Eng* 125:912–923. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1999\)125:9\(912\)](https://doi.org/10.1061/(ASCE)0733-9429(1999)125:9(912))

- Alados I, Foyo-Moreno I, Alados-Arboledas L (2012) Estimation of downwelling longwave irradiance under all-sky conditions. *Int J Climatol* 32:781–793. <https://doi.org/10.1002/joc.2307>
- Alados-Arboledas L, Jimenez J (1988) Day-night differences in the effective emissivity from clear skies. *Bound-Layer Meteorol* 45:93–101. <https://doi.org/10.1007/BF00120817>
- Alcántara EH, Stech JL, Lorenzetti JA, Bonnet MP, Casamitjana X, Assireu AT, Novo EMLdM (2010) Remote sensing of water surface temperature and heat flux over a tropical hydroelectric reservoir. *Remote Sens Environ* 114:2651–2665. <https://doi.org/10.1016/j.rse.2010.06.002>
- Alduchov OA, Eskridge RE (1996) Improved magnus form approximation of saturation vapor pressure. *J Appl Meteorol* 35:601–609. [https://doi.org/10.1175/1520-0450\(1996\)035%3c0601:IMFAO S%3e2.0.CO;2](https://doi.org/10.1175/1520-0450(1996)035%3c0601:IMFAO S%3e2.0.CO;2)
- Ali S, Ghosh NC, Singh R (2008) Evaluating best evaporation estimate model for water surface evaporation in semi-arid region. *India Hydrol Process* 22:1093–1106. <https://doi.org/10.1002/hyp.6664>
- Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration—guidelines for computing crop water requirements—FAO Irrigation and drainage paper 56, vol 300. Rome
- Anderson ER (1954) Energy-budget studies Waterloss investigations: Lake Hefner studies. Technical Report, US Geological Survey professional paper, vol 269, pp 71–119
- Ångström AK (1915) A study of the radiation of the atmosphere: based upon observations of the nocturnal radiation during expeditions to Algeria and to California, vol 65. Smithsonian Institution, Washington
- Arifin RR, James SC, de Alwis Pitts DA, Hamlet AF, Sharma A, Fernando HJ (2016) Simulating the thermal behavior in Lake Ontario using EFDC. *J Great Lakes Res* 42:511–523. <https://doi.org/10.1016/j.jglr.2016.03.011>
- Benyahya L, Caissie D, El-Jabi N, Satish MG (2010) Comparison of microclimate vs. remote meteorological data and results applied to a water temperature model (Miramichi River, Canada). *J Hydrol* 380:247–259. <https://doi.org/10.1016/j.jhydrol.2009.10.039>
- Berdahl P, Fromberg R (1982) The thermal radiance of clear skies. *Sol Energy* 29:299–314. [https://doi.org/10.1016/0038-092X\(82\)90245-6](https://doi.org/10.1016/0038-092X(82)90245-6)
- Berdahl P, Martin M (1984) Emissivity of clear skies. *Sol Energy* 32:663–664. [https://doi.org/10.1016/0038-092X\(84\)90144-0](https://doi.org/10.1016/0038-092X(84)90144-0)
- Berger X, Buriot D, Garnier F (1984) About the equivalent radiative temperature for clear skies. *Sol Energy* 32:725–733. [https://doi.org/10.1016/0038-092X\(84\)90247-0](https://doi.org/10.1016/0038-092X(84)90247-0)
- Bolton D (1980) The computation of equivalent potential temperature. *Mon Weather Rev* 108:1046–1053. [https://doi.org/10.1175/1520-0493\(1980\)108%3c1046:TCOEPT%3e2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108%3c1046:TCOEPT%3e2.0.CO;2)
- Boltzmann L (1884) Ableitung des Stefan'schen Gesetzes, betreffend die Abhängigkeit der Wärmestrahlung von der Temperatur aus der electromagnetischen Lichttheorie. *Ann Phys* 258:291–294. <https://doi.org/10.1002/andp.18842580616>
- Bowen IS (1926) The ratio of heat losses by conduction and by evaporation from any water surface. *Phys Rev* 27:779. <https://doi.org/10.1103/PhysRev.27.779>
- Brady DK, Graves WL, Geyer JC (1969) Surface heat exchange at power plant cooling lakes. Edison Electric Institute publication, Washington
- Brett JR (1956) Some principles in the thermal requirements of fishes. *Q Rev Biol* 31:75–87. <https://doi.org/10.1086/401257>
- Brewster MQ (1992) Thermal radiative transfer and properties. Wiley, New York
- Brunt D (1932) Notes on radiation in the atmosphere. I. *Q J R Meteorol Soc* 58:389–420. <https://doi.org/10.1002/qj.49705824704>
- Brutsaert W (1975) On a derivable formula for long-wave radiation from clear skies. *Water Resour Res* 11:742–744. <https://doi.org/10.1029/WR011i005p00742>
- Caissie D (2006) The thermal regime of rivers: a review. *Freshw Biol* 51:1389–1406. <https://doi.org/10.1111/j.1365-2427.2006.01597.x>
- Caissie D, Satish MG, El-Jabi N (2007) Predicting water temperatures using a deterministic model: application on Miramichi River catchments (New Brunswick, Canada). *J Hydrol* 336:303–315. <https://doi.org/10.1016/j.jhydrol.2007.01.008>
- Carmona F, Rivas R, Caselles V (2014) Estimation of daytime downward longwave radiation under clear and cloudy skies conditions over a sub-humid region. *Theoret Appl Climatol* 115:281–295. <https://doi.org/10.1007/s00704-013-0891-3>
- Coutant CC (1999) Perspectives on temperature in the pacific northwest's fresh waters. Environmental Protection Agency, United States, Washington. <https://doi.org/10.2172/9042>
- Chapra SC (2008) Surface water-quality modeling. Waveland press, Long Grove
- Choi M, Jacobs JM, Kustas WP (2008) Assessment of clear and cloudy sky parameterizations for daily downwelling longwave radiation over different land surfaces in Florida, USA. *Geophys Res Lett*. <https://doi.org/10.1029/2008GL035731>
- Crawford TM, Duchon CE (1999) An improved parameterization for estimating effective atmospheric emissivity for use in calculating daytime downwelling longwave radiation. *J Appl Meteorol* 38:474–480. [https://doi.org/10.1175/1520-0450\(1999\)038%3c0474:AIPFEE%3e2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038%3c0474:AIPFEE%3e2.0.CO;2)
- Currie RJ, Bennett WA, Beitinger TL (1998) Critical thermal minima and maxima of three freshwater game-fish species acclimated to constant temperatures. *Environ Biol Fishes* 51:187–200. <https://doi.org/10.1023/a:1007447417546>
- Czernuszenko W (1990) Dispersion of pollutants in flowing surface waters. *Environ Fluid Mech* 10:119–168
- Dalton J (1802) On evaporation. Essay III in: Experimental essays on the constitution of mixed gases; on the force of steam or vapour from water or other liquids in different temperatures both in a Torrecellian vacuum and in air; on evaporation; and on the expansion of gases by heat. *Mem Proc Lit Phil Soc Manchester* 5:574–594
- Deacon E (1970) The derivation of Swinbank's long-wave radiation formula. *Q J R Meteorol Soc* 96:313–319. <https://doi.org/10.1002/qj.49709640814>
- Deas ML, Lowney CL (2000) Water temperature modeling review: central valley. California Water Modeling Forum
- Duarte HF, Dias NL, Maggioletto SR (2006) Assessing daytime downward longwave radiation estimates for clear and cloudy skies in Southern Brazil. *Agric For Meteorol* 139:171–181. <https://doi.org/10.1016/j.agrformet.2006.06.008>
- Edinger JE (1974) Heat exchange and transport in the environment. *Rep Electr Power Res Inst* 14:125
- Elsasser WM (1942) Heat transfer by infrared radiation in the atmosphere. *Harvard Meteor Studies* 6:107
- Endrizzi S, Tubino M, Zolezzi G (2002) Lateral mixing in meandering channels: a theoretical approach. In: Bousmar D, Zech Y (eds) Proceedings river flow 2000, International conference on fluvial hydraulics Louvain-La-Neuve Belgium. Swets & Zeitlinger, Lisse
- Evans EC, McGregor GR, Petts GE (1998) River energy budgets with special reference to river bed processes. *Hydrol Process* 12:575–595. [https://doi.org/10.1002/\(SICI\)1099-1085\(19980330\)12:4%3c575:AID-HYP595%3e3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-1085(19980330)12:4%3c575:AID-HYP595%3e3.0.CO;2-Y)
- FAO (1990) Annex V. FAO Penman-Monteith formula. Food and Agriculture Organization of the United Nations, Rome

- Flerchinger G, Xaio W, Marks D, Sauer T, Yu Q (2009) Comparison of algorithms for incoming atmospheric long-wave radiation. *Water Resour Res* 45:1–13. <https://doi.org/10.1029/2008WR007394>
- Garner G, Malcolm IA, Sadler JP, Hannah DM (2014) What causes cooling water temperature gradients in a forested stream reach? *Hydrol Earth Syst Sci* 18:5361–5376. <https://doi.org/10.5194/hess-18-5361-2014>
- Garner G, Malcolm IA, Sadler JP, Hannah DM (2017) The role of riparian vegetation density, channel orientation and water velocity in determining river temperature dynamics. *J Hydrol* 553:471–485. <https://doi.org/10.1016/j.jhydrol.2017.03.024>
- Glose A, Lautz LK, Baker EA (2017) Stream heat budget modeling with HFLUX: model development, evaluation, and applications across contrasting sites and seasons. *Environ Model Softw* 92:213–228. <https://doi.org/10.1016/j.envsoft.2017.02.021>
- Guymier I (1998) Longitudinal dispersion in sinuous channel with changes in shape. *J Hydraul Eng* 124:33–40. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1998\)124:1\(33\)](https://doi.org/10.1061/(ASCE)0733-9429(1998)124:1(33))
- Hannah DM, Malcolm IA, Soulsby C, Youngson AF (2004) Heat exchanges and temperatures within a salmon spawning stream in the Cairngorms, Scotland: seasonal and sub-seasonal dynamics. *River Res Appl* 20:635–652. <https://doi.org/10.1002/rra.771>
- Harbeck GE (1958) Water-loss investigations: Lake Mead studies, vol 298. US Government Printing Office, Washington
- Heitor A, Biga A, Rosa R (1991) Thermal radiation components of the energy balance at the ground. *Agric For Meteorol* 54:29–48. [https://doi.org/10.1016/0168-1923\(91\)90039-S](https://doi.org/10.1016/0168-1923(91)90039-S)
- Hester ET, Doyle MW (2011) Human impacts to river temperature and their effects on biological processes: a quantitative synthesis¹. *JAWRA J Am Water Resour Assoc* 47:571–587. <https://doi.org/10.1111/j.1752-1688.2011.00525.x>
- Hughes D, Kingston D, Todd M (2011) Uncertainty in water resources availability in the Okavango River basin as a result of climate change. *Hydrol Earth Syst Sci* 15:931–941. <https://doi.org/10.5194/hess-15-931-2011>
- Idso SB (1981) A set of equations for full spectrum and 8-to 14- μ m and 10.5-to 12.5- μ m thermal radiation from cloudless skies. *Water Resour Res* 17:295–304. <https://doi.org/10.1029/WR017i002p00295>
- Idso SB, Jackson RD (1969) Thermal radiation from the atmosphere. *J Geophys Res* 74:5397–5403. <https://doi.org/10.1029/JC074i023p05397>
- Iziomon MG, Mayer H, Matzarakis A (2003) Downward atmospheric longwave irradiance under clear and cloudy skies: measurement and parameterization. *J Atmos Solar Terr Phys* 65:1107–1116. <https://doi.org/10.1016/j.jastp.2003.07.007>
- Ji ZG (2008) Hydrodynamics and water quality: modeling rivers, lakes, and estuaries. Wiley, New York
- Jiménez JI, Alados-Arboledas L, Castro-Díez Y, Ballester G (1987) On the estimation of long-wave radiation flux from clear skies. *Theor Appl Climatol* 38:37–42. <https://doi.org/10.1007/bf00866251>
- Jirka GH, Weitbrecht V (2005) Mixing models for water quality management in rivers: continuous and instantaneous pollutant releases. In: Czernuszenko W, Rowiński PM (eds) *Water quality hazards and dispersion of pollutants*. Springer, Boston, pp 1–34. https://doi.org/10.1007/0-387-23322-9_1
- Johnson SL (2004) Factors influencing stream temperatures in small streams: substrate effects and a shading experiment. *Can J Fish Aquat Sci* 61:913–923. <https://doi.org/10.1139/F04-040>
- Joss J, Resele G (1987) Mathematical modelling of the heat exchange between a river and the atmosphere. In: Beniston M, Pielke RA (eds) *Interactions between energy transformations and atmospheric phenomena. A survey of recent research*. Springer, Dordrecht, pp 27–40. https://doi.org/10.1007/978-94-017-1911-7_3
- Kalinowska MB, Rowiński PM (2012) Uncertainty in computations of the spread of warm water in a river—lessons from Environmental Impact Assessment case study. *Hydrol Earth Syst Sci* 16:4177–4190. <https://doi.org/10.5194/hess-16-4177-2012>
- Kalinowska MB, Rowiński PM (2015) Thermal pollution in rivers—modelling of the spread of thermal plumes. In: Rowiński P, Radecki-Pawlik A (eds) *Rivers—physical, fluvial and environmental processes GeoPlanet-Earth and Planetary Sciences*. Springer, Cham, pp 591–613. https://doi.org/10.1007/978-3-319-17719-9_24
- Kalinowska MB, Rowiński PM (2008) Numerical solutions of two-dimensional mass transport equation in flowing surface waters, vol 404(E-8). Institute of Geophysics, Polish Academy of Sciences, Warsaw
- Kalinowska MB, Rowiński PM, Kubrak J, Mirosław-Swiątek D (2012) Scenarios of the spread of a waste heat discharge in a river—Vistula River case study. *Acta Geophys* 60:214–231. <https://doi.org/10.2478/s11600-011-0045-x>
- Kalinowska MB, Mrokowska MM, Rowiński PM (2018) Sensitivity analysis for the water-air heat exchange term. In: Kalinowska MB, Mrokowska MM, Rowiński PM (eds) *Free surface flows and transport processes*. Springer, Cham, pp 219–233
- Katsaros KB, McMurdie LA, Lind RJ, DeVault JE (1985) Albedo of a water surface, spectral variation, effects of atmospheric transmittance, sun angle and wind speed. *J Geophys Res Oceans* 90:7313–7321. <https://doi.org/10.1029/JC090iC04p07313>
- Khatib T, Elmenreich W (2015) A model for hourly solar radiation data generation from daily solar radiation data using a generalized regression artificial neural network. *Int J Photoenergy* 968024:1–13. <https://doi.org/10.1155/2015/968024>
- Kjaersgaard JH, Plauborg F, Hansen S (2007) Comparison of models for calculating daytime long-wave irradiance using long term data set. *Agric For Meteorol* 143:49–63. <https://doi.org/10.1016/j.agrformet.2006.11.007>
- Lawrence MG (2005) The relationship between relative humidity and the dewpoint temperature in moist air: a simple conversion and applications. *Bull Am Meteor Soc* 86:225–234. <https://doi.org/10.1175/bams-86-2-225>
- Leach J, Moore R (2010) Above-stream microclimate and stream surface energy exchanges in a wildfire-disturbed riparian zone. *Hydrol Process* 24:2369–2381. <https://doi.org/10.1002/hyp.7639>
- Lee TY et al (2012) Modeling the effects of riparian planting strategies on stream temperature: increasing suitable habitat for endangered Formosan Landlocked Salmon in Shei-Pa National Park. *Taiwan Hydrol Process* 26:3635–3644. <https://doi.org/10.1002/hyp.8440>
- Lhomme J-P, Vacher J-J, Rocheteau A (2007) Estimating downward long-wave radiation on the Andean Altiplano. *Agric For Meteorol* 145:139–148. <https://doi.org/10.1016/j.agrformet.2007.04.007>
- Li G, Jackson CR, Krasinski KA (2012) Modeled riparian stream shading: agreement with field measurements and sensitivity to riparian conditions. *J Hydrol* 428–429:142–151. <https://doi.org/10.1016/j.jhydrol.2012.01.032>
- Li M, Jiang Y, Coimbra CF (2017) On the determination of atmospheric longwave irradiance under all-sky conditions. *Sol Energy* 144:40–48. <https://doi.org/10.1016/j.solener.2017.01.006>
- Magnus G (1844) Versuche über die Spannkraft des Wasserdampfes. *Ann Phys* 137:225–247. <https://doi.org/10.1002/andp.18441370202>
- Magnusson J, Jonas T, Kirchner JW (2012) Temperature dynamics of a proglacial stream: identifying dominant energy balance components and inferring spatially integrated hydraulic geometry. *Water Resour Res* 48:12. <https://doi.org/10.1029/2011WR011378>

- Marciano Jr J, Harbeck G (1952) Mass transfer studies. Water-loss investigations: volume 1, Lake Hefner Studies, US Navy Electronics Laboratory. Technical Report 327
- Marks D, Dozier J (1979) A clear-sky longwave radiation model for remote alpine areas. *Arch Meteorol Geophys Bioklimatol Ser B* 27:159–187. <https://doi.org/10.1007/BF02243741>
- Marthews TR, Malhi Y, Iwata H (2012) Calculating downward long-wave radiation under clear and cloudy conditions over a tropical lowland forest site: an evaluation of model schemes for hourly data. *Theor Appl Climatol* 107:461–477. <https://doi.org/10.1007/s00704-011-0486-9>
- McJannet DL, Webster IT, Cook FJ (2012) An area-dependent wind function for estimating open water evaporation using land-based meteorological data. *Environ Model Softw* 31:76–83. <https://doi.org/10.1016/j.envsoft.2011.11.017>
- Meyer AF (1928) The elements of hydrology, vol 7. Wiley, New York
- Miller AW, Street RL (1972) Surface heat transfer from a hot spring fed lake. Technical report (Stanford University, Department of Civil Engineering), vol 160. Stanford, California: Department of Civil Engineering, Stanford University
- Monteith J (1961) An empirical method for estimating long-wave radiation exchanges in the British Isles. *Q J R Meteorol Soc* 87:171–179. <https://doi.org/10.1002/qj.49708737206>
- Murray JD (2002) Mathematical biology I. An introduction, vol 17. Interdisciplinary applied mathematics, 3rd edn. Springer, New York. <https://doi.org/10.1007/b98868>
- Piotrowski A, Rowinski P, Napiórkowski J (2006) Assessment of longitudinal dispersion coefficient by means of different neural networks. In: 7th international conference on hydroinformatics, HIC
- Prata A (1996) A new long-wave formula for estimating downward clear-sky radiation at the surface. *Q J R Meteorol Soc* 122:1127–1151. <https://doi.org/10.1002/qj.49712253306>
- Rajwa A, Rowiński PM, Bialik RJ, Karpiński M (2014) Stream diurnal profiles of dissolved oxygen—case studies. In 3rd IAHR Europe Congress, Porto
- Rajwa-Kuligiewicz A, Bialik RJ, Rowiński PM (2015) Dissolved oxygen and water temperature dynamics in lowland rivers over various timescales. *J Hydrol Hydromech* 63:353–363. <https://doi.org/10.1515/johh-2015-0041>
- Raudkivi A (1979) Hydrology: an observed introduction to hydrological processes and modeling. Pergamon Press, New York
- Rosenberry DO, Winter TC, Buso DC, Likens GE (2007) Comparison of 15 evaporation methods applied to a small mountain lake in the northeastern USA. *J Hydrol* 340:149–166. <https://doi.org/10.1016/j.jhydrol.2007.03.018>
- Rowinski PM, Kalinowska MB (2006) Admissible and inadmissible simplifications of pollution transport equations. In: Ferreira RML, Alves CTL, Leal GAB, Cardoso AH (eds) River flow 2006, vols 1 and 2, pp 199–208
- Rowiński PM, Piotrowski A, Napiórkowski JJ (2005) Are artificial neural network techniques relevant for the estimation of longitudinal dispersion coefficient in rivers? *Hydrol Sci J* 50:175–187. <https://doi.org/10.1623/hysj.50.1.175.56339>
- Rutherford J (1994) River mixing. Wiley, New York
- Rutherford JC, Macaskill JB, Williams BL (1993) Natural water temperature variations in the lower Waikato River, New Zealand. *NZ J Mar Freshwat Res* 27:71–85. <https://doi.org/10.1080/00288330.1993.9516547>
- Ryan PJ (1973) An analytical and experimental study of transient cooling pond behavior, Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics, Massachusetts Institute of Technology, Cambridge
- Santos CAC, Silva BB, Rao TVR, Satyamurty P, Manzi AO (2011) Downward longwave radiation estimates for clear-sky conditions over northeast. *Braz Rev Bras Meteorol* 26:443–450. <https://doi.org/10.1590/S0102-77862011000300010>
- Satterlund DR (1979) An improved equation for estimating long-wave radiation from the atmosphere. *Water Resour Res* 15:1649–1650. <https://doi.org/10.1029/WR015i006p01649>
- Sellers WD (1965) Physical climatology. University of Chicago Press, Chicago
- Sinokrot BA, Stefan HG (1993) Stream temperature dynamics: measurements and modeling. *Water Resour Res* 29:2299–2312. <https://doi.org/10.1029/93WR00540>
- Sridhar V, Elliott RL (2002) On the development of a simple downwelling longwave radiation scheme. *Agric For Meteorol* 112:237–243. [https://doi.org/10.1016/S0168-1923\(02\)00129-6](https://doi.org/10.1016/S0168-1923(02)00129-6)
- Swinbank WC (1963) Long-wave radiation from clear skies. *Q J R Meteorol Soc* 89:339–348. <https://doi.org/10.1002/qj.49708938105>
- Szymkiewicz R (2010) Numerical modeling in open channel hydraulics. Springer Science & Business Media, Berlin
- Tan SBK, Shuy EB, Chua LHC (2007) Modelling hourly and daily open-water evaporation rates in areas with an equatorial climate. *Hydrol Process* 21:486–499. <https://doi.org/10.1002/hyp.6251>
- Tang R, Etzion Y, Meir IA (2004) Estimates of clear night sky emissivity in the Negev Highlands. *Israel Energy Convers Manag* 45:1831–1843. <https://doi.org/10.1016/j.enconman.2003.09.033>
- Trabert W (1896) Neue Beobachtungen über Verdampfungsgeschwindigkeiten. [New Observations on Evaporation Rates.]. *Meteorol Z* 13:261–263. <https://doi.org/10.4236/jwarp.2017.912086>
- Vall S, Castell A (2017) Radiative cooling as low-grade energy source: a literature review. *Renew Sustain Energy Rev* 77:803–820. <https://doi.org/10.1016/j.rser.2017.04.010>
- Wallis SG, Manson JR (2004) Methods for predicting dispersion coefficients in rivers. In: Proceedings of the Institution of Civil Engineers-Water Management, vol 3. Thomas Telford Ltd, pp 131–141
- Webb BW, Zhang Y (1999) Water temperatures and heat budgets in Dorset chalk water courses. *Hydrol Process* 13:309–321. [https://doi.org/10.1002/\(SICI\)1099-1085\(19990228\)13:3%3c309:AID-HYP740%3e3.0.CO;2-7](https://doi.org/10.1002/(SICI)1099-1085(19990228)13:3%3c309:AID-HYP740%3e3.0.CO;2-7)
- Webb BW, Hannah DM, Moore RD, Brown LE, Nobilis F (2008) Recent advances in stream and river temperature research. *Hydrol Process* 22:902–918. <https://doi.org/10.1002/hyp.6994>
- Winter TC, Rosenberry DO, Sturrock AM (1995) Evaluation of 11 equations for determining evaporation for a small lake in the north central United States. *Water Resour Res* 31:983–993. <https://doi.org/10.1029/94WR02537>
- Wunderlich WO (1972) Heat and mass transfer between a water surface and the atmosphere. Norris, Tenn: Tennessee Valley Authority, Office of Natural Resources and Economic Development, Division of Air and Water Resources, Water Systems Development Branch
- Xin Z, Kinouchi T (2013) Analysis of stream temperature and heat budget in an urban river under strong anthropogenic influences. *J Hydrol* 489:16–25. <https://doi.org/10.1016/j.jhydrol.2013.02.048>



The water resources of tropical West Africa: problems, progress, and prospects

Christopher E. Ndehedehe¹

Received: 16 September 2018 / Accepted: 30 January 2019 / Published online: 7 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

West Africa plays key roles in global climate and shows one of the strongest variations in hydro-climatic conditions. As it turns out, the region appears to be underrepresented in the existing compendium of Earth science and hydrology-focused journal papers when it comes to significant discussion on terrestrial hydrology and freshwater science. This prominent gap is largely precipitated by increasing number of constraints that include lack of considerable and robust investments in gauge measurements for meteorological and hydrological applications, poor funding of research institutions and other disincentives, among other factors. In this manuscript, the challenges and problems in large-scale terrestrial hydrology-focused investigation in West Africa are reviewed. Using a dossier of some recent contributions in the field of remote sensing hydrology, this review also highlights some of the progress in terrestrial hydrology and the opportunities that exist for hydro-geodetic research in West Africa that leverage on sustained investments in satellite geodetic missions. It is noted that West Africa is still a pristine environment for hydrology-focused research and can benefit from recent advancements in sophisticated space agency programs such as the Gravity Recovery and Climate Experiment, which undoubtedly has revolutionized terrestrial hydrology research around the world for nearly two decades. Given the poor density of gauge stations and limited ground observations, hydrological research in West Africa is expected to benefit more from independent space observations and multi-resolution data. This is because the lack of sufficient in-situ data for the parameterizations and adequate initialization of outputs from hydrological models and reanalysis data for hydrological applications results in poor representation of the West African land surface and hydrological state variables. To further improve our contemporary understanding of West Africa's terrestrial hydrology, the continued evaluation/validation of these observations and space-borne measurements is advocated.

Keywords Terrestrial hydrology · Droughts · Rainfall · River discharge · Climate variability

Introduction

Some new metrics based on the framework of the intersectoral impact model intercomparison project (ISI-MIP) suggest a 40% increase in the number of people living under absolute water scarcity, up from 15% of the global population with a severe decrease in water resources (Schewe et al. 2013). Whereas humanity in general takes water for granted owing to its obvious simplicity (see Falkenmark and Lundqvist 1998), future simulations of freshwater availability show increased water stress in the coming decades owing to population rise

and the impacts of unmitigated climate change. For instance, total water withdrawals as at the year 2000 amounted to 8.4% of global water resources, and this amount is expected to increase up to 12.2% (i.e., between 4600 and 5800 km³/year) by 2025 (Shiklomanov 2000). In a study of global water resources, Vörösmarty et al. (2000) projected estimate of global water use by 2025 puts the value at 4700 km³/year. Regarding water scarcity projections, about 57% and 69% of the world's population are estimated to live in countries facing high water stress by 2025 and 2075, respectively (Alcamo et al. 1997). The argument that considerable portion of the world's population will face water stress by 2025, be it moderate or severe may be somewhat polarized in terms of projected estimates, given the uncertainties and substantial differences in model outputs (e.g., Gleick 2000; Kuylenstierna et al. 1997). Nevertheless, documentary evidence suggests that water shortages (or rather scarcity) will be one of the world's most

✉ Christopher E. Ndehedehe
c.ndehedehe@griffith.edu.au

¹ Australian Rivers Institute and School of Environment and Science, Griffith University, Nathan, Queensland 4111, Australia

pressing challenges in the twenty-first century (e.g., Freitas 2013; Rijsberman 2006; Vörösmarty et al. 2005; Kuylenstierna et al. 1997; Alcamo et al. 1997).

From a contemporary perspective, the rise in drought frequency across the globe in the last few decades (see Ndehedehe et al. 2019; Spinoni et al. 2014; van der Molen et al. 2011) strengthens this argument of anticipated water stress. Considering the notable and often repeated extreme climatic conditions in Africa, the aforementioned projections and indicators of water stress could have significant negative impacts on a considerable proportion of Africa's agricultural biomes, livelihood, and ecosystem services. Arguably, the region's lack of preparedness, limited resources, and plausible policy responses increases its vulnerability to such impacts. It has been reported that about 25% of people living in Africa are already experiencing water stress, while an estimated 13% are direct recipients of drought-related stress once each generation (see Vörösmarty et al. 2005). Because of such concerns and the cumulative impacts of climate variability on water availability, stormwater harvesting was recently (see Fisher-Jeffes et al. 2017) suggested in South Africa as possible alternative water resource that could supplement traditional urban water supplies and improve water security. Water scarcity by itself represents one of the characteristics (e.g., degree of scarcity) that could make water a potential source of strategic rivalry (Gleick 1993). This is true for Texas, Africa, and other regions where water-related conflicts and the impacts of water scarcity on national security have been identified (Freitas 1993, Gleick 2000, 2013). As further mentioned in Freitas (2013), the increasing pressure on water resources in West Africa and other regions of Africa is a potential source of local agitations and interstate tensions that may eventually lead to armed conflicts. But more importantly, despite all of the recurrent climatic, hydrological, and environmental changes that affect the stability of regional income and livelihood, little is known about the terrestrial water systems of West Africa. This limited understanding of the region's hydrology has been attributed to an increasing number of constraints that include lack of sufficient observational networks and/or decline of in situ observations for large-scale hydrological research in West Africa. In this manuscript, critical challenges and the progress made in large-scale terrestrial hydrology-focused research over West Africa are reviewed. In the next section, an overview of the current state of knowledge in West Africa's terrestrial hydrology is indicated while highlighting the specific objectives of the manuscript.

The terrestrial hydrology of West Africa

Water resources

West Africa is enriched with numerous surface water resources (rivers, estuaries, lakes, reservoirs), including major rivers such as the Niger, Senegal, Gambia, and Lake Chad. These rivers take their sources in tropically wet major groundwater basins and regions with considerable amount of annual rainfall (Fig. 1a, b). In addition to the Benue and the Volta river systems (Black and White Volta rivers and Oti), these river systems among others drain the West African subregion. At basin or water management scale, most of West Africa's water resources are shared by participating riparian countries from about 28 transboundary river basins. Some regional and national institutions exist in these river basins that oversee the management and development of water resources. Significant among them is the GLOWA Volta Project, which is saddled with the responsibility of understanding the impact of global change on the use and availability of water resources in the Volta Basin (e.g., de Giesen et al. 2007). In light of the highly variable climatic conditions of West Africa, the West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL, <https://www.wascal.org/about-wascal/welcome-to-wascal/>) is another regional initiative that provides research-based solutions to enhance the resilience of human and environmental systems to the impacts of climate change. This is achieved through capacity building and provision of research infrastructure for climate studies in several West African countries.

As a little caveat to the study area definition, it should be noted that this review is over West Africa, a region defined by the Economic Community of West African States as covering about sixteen countries, and includes Chad and Cameroon. However, given that the latter is a country in Equatorial Africa, this study also takes advantage of the prominent gap in the knowledge of climate variability and its impacts on hydrological changes in Central Equatorial African countries (i.e., the Congo Basin) to extend this synopsis to include the region. These two regions called tropical West Africa which is hereafter referred to as West Africa in the manuscript for simplicity and convenience (Fig. 1a). In terms of water availability, the Guinea Coast countries and those in the Congo Basin exceed other regions of West Africa (Fig. 1b). As with Guinea Coast region, the Congo Basin has numerous networks of surface water resources (Fig. 1a). Annual rainfall amount in the Guinea Coast area is nonetheless the highest (Fig. 1b) and shows the strongest spatiotemporal variability as opposed to the Congo Basin and the Sahel

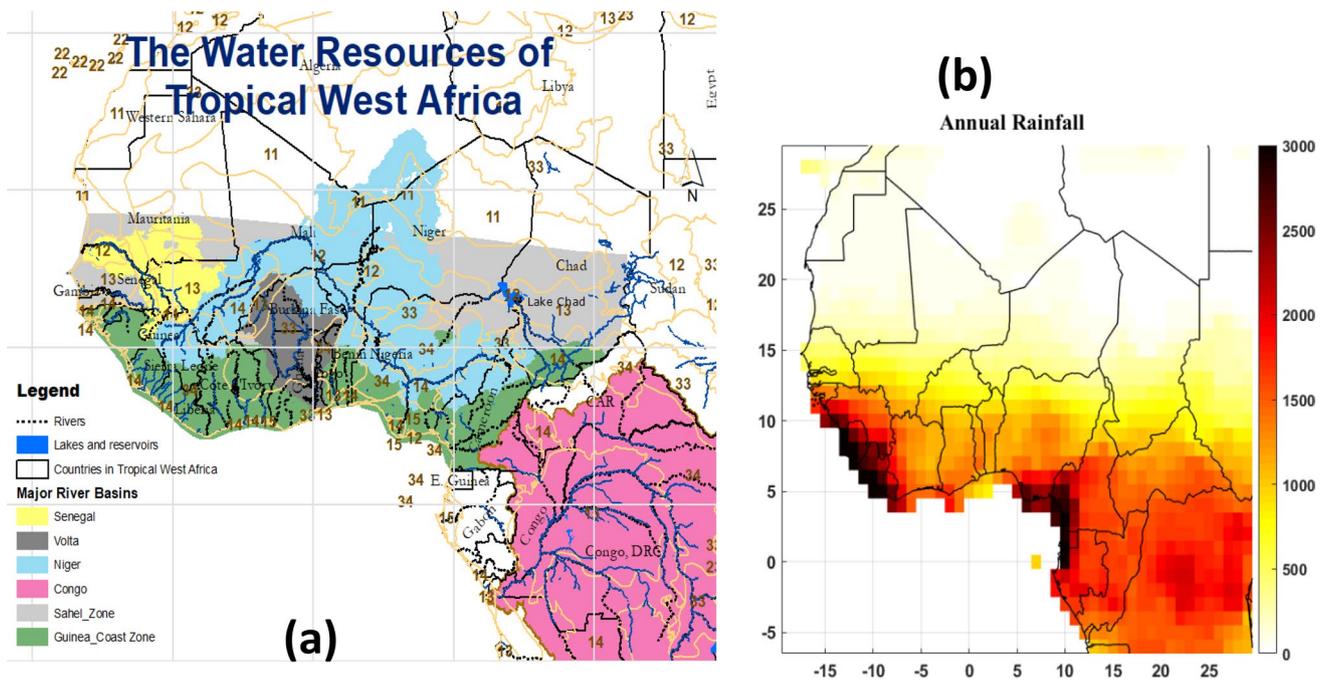


Fig. 1 Map showing some of the water resources in tropical West Africa (i.e., areas that include Central Equatorial Africa or Congo Basin). **a** The hydrological units (rivers, lakes, and groundwater aquifers) and the main river basins, which include Niger, Volta, Congo, and Senegal. The Congo (tulip pink) and Niger (sky blue) river basins are considerably large and apparently the most significant and prominent basins in the region owing to the two major rivers (Niger and Congo) that provide numerous ecosystem services. The aquifers (brown contour lines) are characterized in terms of numbers;

for example, the numbers ranging from 11 to 15 are found in major groundwater basins, while the numbers 33 and 34 on the map are those found in local and shallow aquifers. The aquifer maps and some hydrological units (river distribution networks) are those of Worldwide Hydrogeological Mapping and Assessment Programme. **b** The mean annual rainfall over tropical West Africa based on the Global Precipitation Climatology Center (GPCC) precipitation for the period between 1901 and 2014

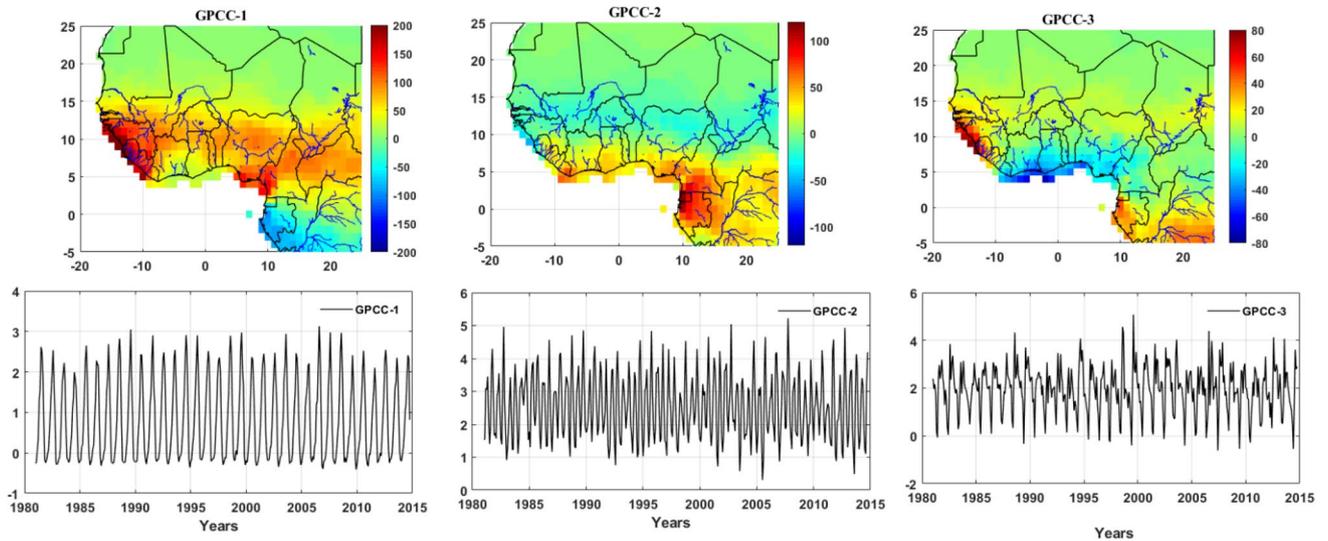


Fig. 2 Statistical decomposition of GPCC-based precipitation (1981–2014) over tropical West Africa using principal component analysis (e.g., Preisendorfer 1988). The EOFs (top) are loadings showing spatial patterns of variations in precipitation over TWA, while the corresponding PCs (bottom) are temporal variations, which are normal-

ized using their standard deviation to be unitless. The total variability accounted for by the three leading orthogonal modes is 60%, 10%, and 5% for GPCC-1, GPCC-2, and GPCC-3, respectively. The blue lines on the spatial patterns (top) are river networks and reservoirs

regions (Fig. 2). The leading orthogonal modes of rainfall over West Africa as shown in Fig. 2 provide an overview of freshwater availability in the region. For example, the regions with stronger spatial loadings (patterns) in rainfall (GPCC-1, Fig. 2) coincide with areas characterized by strong annual amplitudes of rainfall (Fig. 1b), and where strong variability in terrestrial water storage has been observed (Ndehedehe et al. 2016b). The corresponding temporal patterns of leading modes of rainfall over the region are largely characterized by interannual (GPCC-1, which accounts for 60% of total variability), bimodal (GPCC-2), and seasonal variations (GPCC-3) and indicate that Liberia and Guinea receive the most rainfall (Fig. 2). Despite the enormous water resources of these regions (Guinea Coast and the Congo Basin), their increasing vulnerability to several drought intensities in recent times (e.g., Andam-Akorful et al. 2017; Ndehedehe et al. 2016c, 2019) tends to strengthen the argument for the wet getting drier paradigm.

With increasing need of freshwater for several ecosystem functions, water availability is now a concern for some regions in both West and Central Africa that are predominantly semiarid. This is evidenced, for example, in the Lake Chad basin's interest in transferring the surface water resources of the Congo Basin to nourish the nearly desiccated Lake Chad (e.g., Ndehedehe et al. 2016b; Lemoalle et al. 2012). There are ongoing public discourse and debates as to the feasibility of this proposed project. However, the eco-hydrological and cultural implications of this project, in addition to the legal requirements, institutional and other policy solutions, are key issues to be addressed. The enormous shared freshwater resources in the region, which have reduced considerably in the past two decades (e.g., Freitas 2013) as exemplified by the Lake Chad, could exacerbate local tensions due to growing water scarcity and increased competition among riparian countries. To help address some of the problems related to water resources in the region, several regional water initiatives (e.g., Volta Basin Authority), partnerships (e.g., West African Water Partnership), and bilateral cooperations (e.g., agreement between Niger and Benin on the development of surface schemes) (see <https://www.oecd.org/swac/publications/38409569.pdf>) were inaugurated. Ultimately, these institutions and other more recent collaborative efforts with international agencies "Current state of hydrology research" section) are some regional and institutional progress made for sustainability and effective management of water resources.

Hydro-climatic issues

After the notable climatic shifts of the 1970s and 1980s in most West African catchments, some ecological and hydrological impacts were reported. Of note is the loss and

contraction of major freshwater bodies such as the Lake Chad whose surface extent declined by about 90% (Fig. 3). Widespread desertification and degradational transitions in natural vegetation and land use through land clearing and logging (e.g., Knauer et al. 2014; Tucker et al. 1991) were other notable eco-hydrological events that ensued. Few pioneering hydrological papers in addition showed that despite observed precipitation decline during the period, some Sahelian catchments witnessed paradoxical increase in groundwater resources due to changes in soil characteristics and vegetation cover (see, e.g., Mahé and Paturel 2009; Favreau et al. 2009; Séguis et al. 2004; Leduc et al. 1997, 2001). Whereas the efforts to unravel such complex hydrological processes are ongoing (e.g., Gal et al. 2017; Descroix et al. 2009), it has been argued that the tripartite relationship between climate, rural society, and the environment drives land cover changes in the West African Sahel, which in turn drive the distribution of rainfall and water resources in general (Paturel et al. 2017). Moreover, as indicated in global climate projections, the impacts of climate change are expected to have direct and profound negative effects on freshwater availability (see, e.g., Tall et al. 2016; Prudhomme et al. 2014; Schewe et al. 2013). Recently, it has been shown that indices of oceanic variability (e.g., El Niño–Southern Oscillation—ENSO) and sea surface temperature anomalies of the nearby oceans impact on the dynamics and distributions of land water storage in West Africa (e.g., Ndehedehe et al. 2017b). However, a more recent climate projection suggests a climatic shift to drier climate in the 2050 and 2080s in West Africa, largely characterized by increase in drought-prone areas (Gizaw and Gan 2017).

West Africa is one of the world's poorest regions with increasing population and intensity in climate extremes. That population growth and freshwater variability are now significant measures of economic development (e.g., Hall et al. 2014; Brown and Lall 2006), West Africa, under a climate change scenario, perhaps could be the most vulnerable to severe water stress and human-modified drought in the nearest future. As most agricultural goods are produced in regions that are vulnerable to water-related impacts, this will have massive implications not just on the economy of West Africa but other regions of the world that indirectly consume the water resources of West Africa. This arguably emphasizes the strategic importance of West African countries in global food production chain on the one hand. On the other hand, it suggests that the scale of the problem is global.

Current state of hydrology research

Advancements in space agency programs, developments in Earth observation, and its growing applications in hydrology are well documented for other regions of the world.

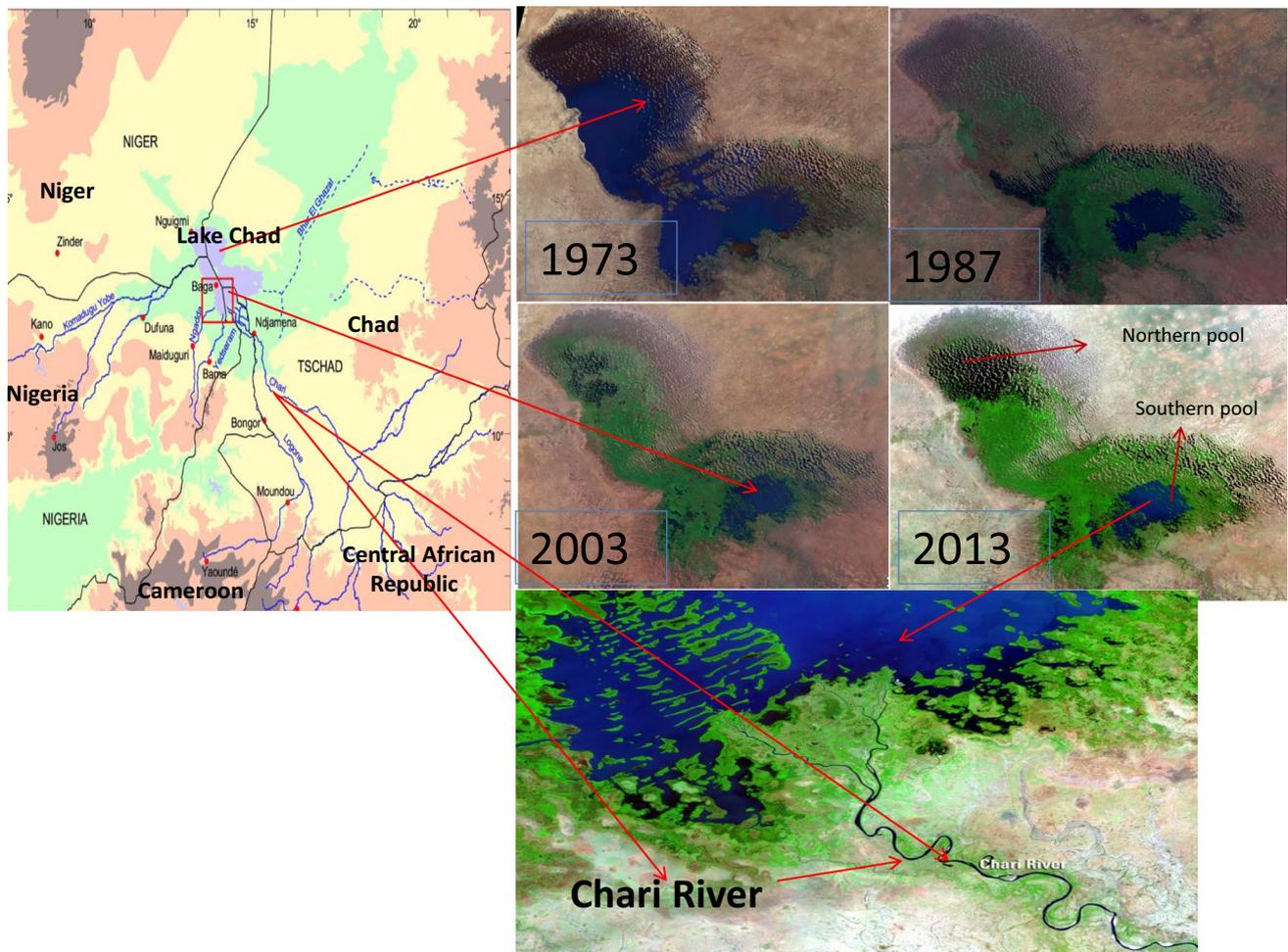


Fig. 3 The spatial and temporal changes in Lake Chad surface area as shown by Landsat imageries for 1973, 1987, 2003, and 2013. The blue lines on the map (left) show the river networks within the basin. The present Lake Chad shows two segmented pools with the northern

pool completely dried up during drought periods (right). The map as adapted from Ndehedehe et al. (2016b) shows the impact of climate change resulting in morphological changes, gradual desiccation, and reduction in Lake Chad's freshwater during the last four decades

However, West Africa's terrestrial hydrology is conspicuously underrepresented in the literature and in fact the least studied. As oppose to other regions of the world with detailed diverse applications of remote sensing data to study hydrological processes (e.g., McCabe et al. 2017; Alsdorf et al. 2016), discipline-specific studies in West Africa that delivers on the terrestrial hydrology and impacts of climate variability on water resources are undeniably limited and somewhat lacking in the volumes of existing studies and catalog of hydrological journals. Extreme climatic conditions (i.e., droughts) of the last 3–4 decades had enormous negative impacts on West Africa's freshwater resources (e.g., Freitas 2013; Wald 1990). Yet the representation of the state of hydrological knowledge in West Africa suggests there are still considerable gaps in knowledge vis-à-vis global perception of Earth science research. One particular area with considerable gap in knowledge in West Africa is

its groundwater and aquifer systems. Although there is some historic evidence regarding groundwater bore measurements and well hydrographs in some locations in the Sahel (e.g., Henry et al. 2011), generally a considerable proportion of West Africa is either lacking hydro-geological data or having sparsely distributed gauged stations and/or rarely monitored well locations. However, some collaborative efforts between the German research program on the global water cycle (GLOWA) and the Integratives Management-Projekt für einen Effizienten und Tragfähigen Umgang mit Süßwasser in West Afrika (IMPETUS) initiative and those of AMMA-CATCH and GHYRAF ("Limited observational networks" section) on water resources development are significant progress made in the region (e.g., Ndehedehe 2017; Andam-Akorful et al. 2017). Be it groundwater or aquifer systems, West Africa seems to be generally dependent on information from external or international agencies like the

World-wide Hydrogeological Mapping and Assessment Programme (Fig. 1a) and the global groundwater information system, among others.

Furthermore, the massive changes in land cover and land use patterns of West Africa in the last 5 decades (e.g., Li et al. 2007) had significant impacts on its terrestrial hydrology causing an imbalance in water budget and complex hydrological processes (see, e.g., Favreau et al. 2009; Séguis et al. 2004; Li et al. 2007; Leduc et al. 1997). However, our contemporary understanding of West Africa's terrestrial hydrology is insufficient and somewhat flagged with inconclusive discourse. Despite the emergence of large-scale hydrological models, the interplay between terrestrial water fluxes and human actions, which has not been fully explored (see Wada et al. 2017), contributes to the limited use of hydrological models for freshwater investigations. Whereas the representation of anthropogenic contributions in global hydrological models is challenging as argued by Wada et al. (2017), the aftermath of water resources assessments from such models for West Africa where various human actions and changes in soil properties drive surface hydrology (e.g., Gal et al. 2017; Favreau et al. 2009) would be somewhat skewed and characterized by inconsistent results. This is because of the failure of hydrological models to faithfully and realistically account for the human dimension that also triggers considerable changes in terrestrial water systems. Therefore, if peer-reviewed science papers represent a state of knowledge and an opportunity to highlight scientific findings based on new or revised data and conducting research in regions with limited knowledge (Alsdorf et al. 2016), this is therefore a clarion call to remotely sensed, large-scale terrestrial hydrology-focused investigation in West Africa. The accentuation of such call is predicated on three vital factors: (i) the argument that most drainage basins in West Africa are located in hydrologically unfavorable environments (see Anyadike 1992), (ii) West Africa plays key roles in global climate and shows one of the strongest variations in climatic conditions, and (iii) the strong marginalization of West Africa in Earth science and hydrology research. Hence, opportunities for hydro-climatic research will bring Africa and its subregions to the fore in leading English-language journals that explore global changes in hydrology.

From a contemporary, large-scale innovative context, hydrological research is still much rather at the embryonic stage in Africa. For prominent river basins like the Congo, despite having a long history of hydrological measurements (Alsdorf et al. 2016), there is a considerable decline in the knowledge of the basin's hydrology as opposed to the last four decades. While this review highlights some of the reasons for such limited understanding in recent times, the focus on West Africa is meant to stimulate research interest in other African subregions and provide relevant insights that will improve our understanding of continental freshwater

dynamics. Hence, a synthesis of some recent contributions in the fields of hydrology and water resources in the region are identified and used to (i) assess the knowledge gaps and significant challenges in hydrological research, (ii) examine contemporary perspectives on hydrological processes and drivers of land water storage, (iii) discuss progress and innovative approaches to large-scale freshwater monitoring, and (iv) articulate new and impressive perspectives on the characterization of space–time evolutions of extreme weather events and key hydrological metrics. Among several other considerations, for example, addressing the 'data problem' and the complementary perspectives on hydrological controls on surface vegetation dynamics, this review further highlights the newly introduced hydro-geodetic approaches to terrestrial hydrology and its potential in the assessment of the region's surface and subsurface hydrology.

Knowledge gaps and challenges to large-scale hydrological research

Hydrological variability

Generally in Africa, hydrological variability and several drought episodes have amazingly large negative impacts on freshwater resources, biodiversity, food security, and health among other factors. These deleterious impacts hamper economic development, e.g., low gross domestic product (e.g., Brown and Lall 2006), and contribute to increased poverty in one of the world's poorest continent. Arguably, reoccurring drought episodes of the 1970s and 1980s, and the incessant impacts of climate variability on the socio-economic systems of West Africa were major triggers that led to the plenitude of climate research at the time. The various aspects of these studies focused mainly on analyzing multiple climate variables such as precipitation, runoff, soil moisture, temperature, and zonal winds, among others (see, e.g., Ndehedehe et al. 2016c; Hua et al. 2016; Tarhule et al. 2015; Nicholson 2013; Mohino et al. 2011; Paeth et al. 2012; Conway et al. 2009; Ali and Lebel 2009; Boone et al. 2009). Moreover, historical gauge data, outputs from global and regional climate models (i.e., GCMs and RCMs), and a host of optical remote sensing data, have been widely used in several hydro-climatic research at all levels (county level, region-specific, and large scales) in the region. These studies as highlighted in subsequent sections heavily explored and diagnosed the mechanisms of meteorological patterns and how it drives other fluxes (runoff and evapotranspiration) and land state variables (soil moisture). There are still knowledge gaps nonetheless particularly in areas regarding large-scale temporal and spatial dynamics of terrestrial water storage (TWS), and the impacts of both climate change and water resources development (e.g., irrigation schemes and

dam constructions for hydropower) on the hydrological systems of West Africa. From a hydrological stand point, water development schemes (e.g., Lake Volta) impact on natural hydrological variabilities, making it difficult to understand the influence of climate on hydrological conditions. A typical example is the Volta Basin where human water management is considered a significant driver in the dynamics of surface water (e.g., Ndehedehe et al. 2017a; Ferreira and Asiah 2015). Further, several scientific reports (e.g., Ndehedehe et al. 2018b; Andam-Akorful et al. 2017; Panthou et al. 2014; Todd et al. 2011; Sheffield and Wood 2008) have outlined exhaustively the influence of global climate and some oceanic hot spots on precipitation changes in West Africa. The influence of climate variability and low-frequency variability that are connected to slow oceanic and climate oscillations from global SST anomaly on the region's TWS has only been recently reported (e.g., Ndehedehe et al. 2018b; Andam-Akorful et al. 2017). Such influence modulates hydrological processes and provides significant control on the spatial and temporal distributions of regional changes in TWS and other water budget quantities (e.g., runoff), resulting in increased acceleration of the water cycle.

Moreover, poor understanding of hydrological variability poses significant challenge to risk management and the prediction of extreme weather events (Hall et al. 2014). The knowledge of hydrological variability in this most financially and climatically challenged region of the world is critical. This is true as variations in hydrology have been identified as one of the key variables causing disparity in the level of economic growth among nations, given that it represents a significant challenge to food security and infrastructure development in the world's poorest regions (see Hall et al. 2014; Brown and Lall 2006). For instance, before the unprecedented droughts of 1980s, the surface water of Lake Chad played significant role in the economy and stability of the region. Apart from other numerous surface water developments for agricultural purposes, Lake Volta and the Kainji Dam are dedicated income-generating hydropower schemes, which sometimes are affected by drought events (e.g., Ndehedehe et al. 2016a, c).

In West Africa, water-related knowledge and innovative technologies are not in the front lines of academic research institutions and government agenda, owing to poor funding and other disincentives. Consequently, the vulnerability of the region to the impacts of climate change (Fig. 3) will continue to grow with ripple effects through its socioeconomic systems. Observed devastating extremes in the climate of West Africa raise some concerns not just for water availability but also for a number of issues that include food security, health, policy and risk management strategies, and socioeconomic challenges (e.g., migration, GDP, etc.). The trajectory of future changes in hydrological conditions and water management practices are complex and will influence

agricultural systems in terms of adaptation measures and mitigation strategies (see, e.g., Paolino et al. 2012; Roudier et al. 2011; Falloon and Betts 2010). Despite the tragedies of frequent extremes in climatic conditions over West Africa, understanding hydrological variability is a wise choice that can unlock sustainable pathways and help mitigate the impacts of hydrological changes on economic growth.

Limited observational networks

Knowledge gaps in large-scale TWS dynamics and the impacts of climate on the hydrology of West Africa exist, primarily because of limited or lack of observational networks to provide the current state of hydrological information. Gauge stations for rainfall and river discharge measurements are in decline globally (e.g., Alsdorf et al. 2007; Alsdorf and Lettenmaier 2003; Lettenmaier 2005). It is even worse in non-industrialized nations such as the African subregion where the gauge networks are not just extremely sparse, discontinuous, and lacking, but their density falls far below the WMO (World Meteorological Organization) guidelines (e.g., Vörösmarty et al. 2001). As further mentioned in Vörösmarty et al. (2005), the routine reporting of African river discharge to relevant climate agencies such as the WMO declined by 90% since 1990. Despite some investments in gauge measurements for meteorological and hydrological applications in some parts of West Africa, many subregions have little or no history of hydrological measurements. This is largely attributed to incessant political instabilities, lack of civilization (i.e., in terms of investing in research), and the costs and logistics implications for the installation of gauges to characterize flow dynamics. The paucity of primary baseline data in West Africa is major challenges that may further create a huge hydro-climatic information gap. This could restrict the availability of information regarding future projections of floods and droughts, in addition to the interannual fluctuations in streamflow in West Africa (Roudier et al. 2014).

The lack of capacity to face the future challenges of climate change impact on water resources, owing to sparse observational record in some gauged river basins, is intensified by political concerns, legal and institutional constraints that have hindered the acquisition of existing data for scientific purposes. Research institutions find it difficult to engage in regional water management and transboundary water sharing conversations because hydro-climatic data from national archives and government repositories are frequently withheld, mostly for political and security reasons. This could only be more damaging for West African countries where water sharing across political boundaries is vastly enormous. Further, the African Monsoon Multidisciplinary Analysis-Couplage Atmosphere Tropicale Cycle Hydrologique (AMMA-CATCH, e.g., Lebel et al. 2009)

and Gravity and Hydrology in Africa (GHYRAF, Hinderer et al. 2009) projects are few among the significant initiatives that are primarily dedicated to studying land surface conditions in West Africa and other African regions through monitoring of hydro-climatological and ecological changes. However, whereas the AMMA-CATCH networks are highly insufficient with presence in only three countries of West Africa, the GHYRAF project is sandwiched between poor funding and transboundary issues (Ndehedehe et al. 2016a). Increased observational networks are critical for the proper initialization of numerical weather prediction models and monitoring climate variability in West Africa (Jenkins et al. 2002). But as indicated in some reports (see, e.g., Ndehedehe et al. 2016a; Todd et al. 2011; Farnsworth et al. 2011; Conway et al. 2009), the paucity of high-quality baseline data remains a major constraint to understanding the climatic influence on changes in the hydrological cycle over the region.

As argued by Alsdorf et al. (2007), gauge observations such as those from river stage are incapable of large-scale monitoring of hydrological conditions. This is because of the physics of water flow across floodplains and wetlands, and since gauge networks essentially provide a one spatial dimension information about the spatial dynamics of surface water extent. For some of these reasons and those presented by Alsdorf et al. (2003) in a related study, large-scale dynamics in surface waters (lakes, wetlands, rivers, reservoirs) are generally unknown. Understanding changes in the hydrological cycle over West Africa would not be possible without large-scale measurements that allow the estimations of the temporal and spatial dynamics of hydrological quantities. Apparently, developments in climate models (e.g., Erfanian et al. 2016; Washington et al. 2013; Otto et al. 2013; Mohino et al. 2011; Cook and Vizy 2006; Li et al. 2005; Koster et al. 2004; Giannini et al. 2003; Lebel et al. 2000), land surface parameterization schemes, and hydrological models (e.g., Sheffield et al. 2014; Thiemiig et al. 2013; Lemoalle et al. 2012; Todd et al. 2011) have shown some prospects as they have been widely employed to forecast and study the West African climatology and large-scale variability in the varied components of the hydrological cycle. Global hydrological models (GHMs) forced by global climate models and the latest greenhouse gases (GHG) concentration scenarios and atmospheric general circulation model (AGCM) have also shown potentials in global water assessment of droughts and water resources (e.g., Prudhomme et al. 2014; Schewe et al. 2013; Corzo Perez et al. 2011). However, the skills of these models (GHMs, Global Climate Models—GCMs, and Regional Climate Models—RCMs) are restricted due to (i) their poor representation of surface water balance (e.g., Alsdorf et al. 2007), (ii) bias and conceptual model uncertainties (e.g., Thiemiig et al. 2013; Todd et al. 2011; Schuol and Abbaspour 2006; Lebel et al. 2000), (iii) dependence

on computational estimates (e.g., Koster et al. 2004), (iv) lack of feedback processes involving anthropogenic impacts (e.g., Piao et al. 2010; Alsdorf and Lettenmaier 2003), and (v) model physics and choice of parameterizations (e.g., Oetli et al. 2011). Considerable progress is therefore required for the improved estimates of land–atmosphere impacts for GCM climate scenarios (Boone et al. 2009). But given the inherent problems of GCM in simulating primary aspects of the West African monsoon (e.g., Cook and Vizy 2006), and uncertainties in satellite estimates of some key hydrological fluxes globally (Lettenmaier 2005), then the implication of the problem is global.

Framework for assessment of hydrological metrics

The lack of a suitable framework to improve the characterization of space–time evolutions of hydrological quantities and extreme weather events limits climate forecasting systems, understanding of TWS dynamics, and the holistic assessment of key hydrological metrics (e.g., magnitude, frequency, duration, predictability, etc.). The impacts of climate variability and other processes of oceanic interannual variability (e.g., ENSO, Pacific Decadal Oscillation—PDO, etc.) in West Africa may result in contrasting outcomes on hydrological regimes. One of such example is the debate on rainfall recovery over West Africa. Nicholson (2005) and Lebel and Ali (2009) have contrasting cases of recovery possibly due to different reference periods. But essentially, they both show that while recovery was marked in some areas, rainfall deficit continued unabated in other Sahelian catchments. The failure of rainfall regime in one region does not translate to failure in other subregions (e.g., Ndehedehe et al. 2016c; Owusu and Waylen 2009), probably owing to different drivers of rainfall, local and land surface conditions. Assuming drought and wet conditions vary in space and time, then the existing methods (e.g., van Huijgevoort 2014; Panthou et al. 2014; Lebel and Ali 2009; Rouault and Richard 2003; Nicholson et al. 2000) used to analyze extreme weather events and hydrological conditions are insufficient, warranting further studies. It was elaborated further in a recent study (Ndehedehe et al. 2016c) that regional average time series of drought indices such as effective drought index (EDI), standardized precipitation index (SPI), groundwater drought index (GWI), and palmer drought severity index (PDSI) among several other indices (see, e.g., Li and Rodell 2015; Kasei et al. 2010; Ali and Lebel 2009; Laux 2009; Vicente-Serrano 2006; Heim 2002; McKee et al. 1993) hide the underlying spatial variability of the index and may lead to generalizations of wet and dry regimes. As the outcome of such drought indices is likely to be skewed, this may not be very helpful for predicting ecological responses, adequate evaluation of drought impacts, crafting of drought policies, and the fostering of campaign on regional adaptation

strategies. Highlights from previous reports overwhelmingly agree that climate oscillations and tropospheric effects are associated with the characteristics of extreme weather events in the region. Hence, a spatiotemporal drought framework based on the localization of hydro-climatic signals could improve our knowledge of these climatic signals. The statistical decomposition of rainfall over West Africa, which indicates the Guinea Coast region is characterized by annual and bimodal rainfall signals (Fig. 2), illustrates the need for a spatiotemporal framework to enhance our understanding of hydrometeorological patterns. Relevant higher-order statistical techniques (e.g., independent component analysis) to support the regionalization of these signals are emerging innovative approaches in hydro-climatic studies (“Higher-order statistical tools and frameworks for spatiotemporal analysis of hydrometeorological data” section).

Climate variability and drivers of land water storage

Impacts of rainfall on stream flow dynamics and surface hydrology

In a study analyzing rainfall and extensive river flow records (1931–1990) in sub-Saharan Africa—SSA, Conway et al. (2009) found that in West Africa, rainfall variability is high and explained 40–60% and 60–80% of the changes in river flows during the periods of 1931–1960 and 1961–1990, respectively. Based on coefficient of variance, Li et al. (2005) showed that rainfall is a principal driver of water fluxes (river discharge and evapotranspiration) in West Africa and that hydrological variability is relatively higher in dry period than wet period. Conway et al. (2009) also confirmed this for SSA, but in addition indicated that river basin physiography and human interventions provide some form of control on interannual and interdecadal variability in river flows. In the Niger river basin, Tarhule et al. (2015) found that streamflow and rainfall fluctuate on cycles that are predominantly 2–4 years long. Further, the hydrological characteristics of three West African countries (Burkina Faso, Ivory Coast and Mali) during the 1951–1997 period showed that rainfall deficit was 15–20%, while river flow deficit was 30–50% and more (Paturel et al. 2003). The regimes of most rivers are expected to be impacted by extreme climatic conditions, topography, soil characteristics, and perhaps vegetation layouts and morphology. Persistent drought conditions, for example, may lead to modifications in ecosystems that also contribute to the nonlinearity of rainfall–runoff transformation in the region (Paturel et al. 2003). The sudden and abrupt change point that occurred in 1969 in the rainfall and streamflow records in all sub-watersheds of the Niger Basin (Tarhule et al. 2015) due to droughts and long periods of

reduced rainfall confirm that rainfall is a significant driver of stream flow in the region. However, several anthropogenic drivers such as the development of water infrastructures, land use change (especially deforestation), and expansion of agricultural areas in West Africa also play key roles in the alteration of river regimes (e.g., Ndehedehe et al. 2017a; Mahe et al. 2013; Paturel et al. 2003; Li et al. 2007). Moreover, several factors that influence runoff in West Africa were studied by Roudier et al. (2014). Apart from rainfall, they showed that potential evapotranspiration was also associated with runoff, in addition to water withdrawal, land use dynamics, and carbon effects, which they argue could also significantly have potential impacts on runoff changes. The surface water resources of West Africa are not only vulnerable to the impacts of climate variability but also to various forms of human activities. This has been demonstrated in some catchments of West Africa where outputs of numerical simulations suggested that deforestation increases annual stream flow by 35–65% (Li et al. 2007). More studies related to the impacts of precipitation, surface water developments, and climate teleconnections as prominent drivers of West Africa’s surface hydrology are summarized in “GRACE hydrological applications in West Africa” section.

Extreme hydro-climatic conditions

Droughts

The highly variable climatic conditions in West Africa pose more threat to numerous water and ecological resources and make it considerably vulnerable to the impact of global climate change. For example, the spatial extent of Lake Chad, a prominent freshwater body in the world’s largest interior drainage basin, declined from 24,000 km² in the 1950s to approximately 1700 km² in recent times (see, e.g., Ndehedehe et al. 2016b; Wald 1990; Birkett 2000; Coe and Foley 2001; Leblanc et al. 2003; Lemoalle et al. 2012). This historic and dramatic decline of Lake Chad surface area (Fig. 3) is the aftermath of the persistent and long drought episodes of the 1960s and 1980s. These severe drought episodes were continental in nature and have been linked by some diagnostic studies (e.g., Bader and Latif 2011; Giannini et al. 2003; Fontaine and Bigot 1993; Janicot 1992) to large-scale climatological shifts and modifications in global SST.

In the Volta Basin, fluctuation in food production was attributed to high variability in rainfall distribution pattern (Kasei et al. 2010), while hydrological drought years were characterized with strong decline in the water levels of Lake Volta (see Ndehedehe et al. 2016c; Bekoe and Logah 2013). The long-term drying observed in the Central Equatorial Africa has been linked to SST variations and circulation changes associated with a weaker West African monsoon (Hua et al. 2016). Combining the long-term drying of

the northern Congolese forest (Zhou et al. 2014) and the reported declines in GRACE-TWS in some parts of the Congo Basin (e.g., Ahmed et al. 2014; Crowley et al. 2006), hydrological conditions in the Congo Basin are therefore not favorable (e.g., Ndehedehe et al. 2018b) and could be a major constraint among other factors, to the proposed and ambitious water transfer from the basin to nourish the Lake Chad basin. It should be noted that despite the notorious impacts of climate variability on water resources, existing dossier on drought evolutions at both regional and basin-specific scales (e.g., Oloruntade et al. 2017; Ndehedehe et al. 2016b, c) is insufficient and does not represent a satisfactory knowledge of hydrometeorological conditions in the region. Given that hydrological drought is said to be largely more driven by temperature than rainfall in the Niger south basin of Nigeria (Oloruntade et al. 2017), drought studies at smaller basin or country scales are important to reveal some latent and intrinsic climatic elements not visible in large-scale studies. This has been demonstrated for the Volta, Lake Chad, and Congo Basins (see Ndehedehe et al. 2016b, c, 2019).

Floods

Whereas droughts, *inter alia*, contribute to food scarcity, famine, hydropower failure, and loss of biodiversity among others (e.g., Ferreira et al. 2018; Ndehedehe et al. 2016c, 2019; Zhou et al. 2014; Shiferaw et al. 2014; Bekoe and Logah 2013), excess water such as the most recent floods across the globe (USA, India, Bangladesh, Nepal, and Nigeria) could even be more devastating. For example, it was mentioned in Basu (2009) that the 2007 severe flood caused by a sequence of above-normal precipitation affected about 800,000 people in West Africa. A climatological diagnostics of this anomalous flood by Paeth et al. (2012) indicated that it was caused by La Niña event in the Pacific and a heightened activity of African easterly waves, among other factors. In 2009, torrential rains and floods affected an estimated 600,000 people in 16 West African countries, killing an estimated 159 people (Basu 2009).

Furthermore, severe floods ravaged West Africa in 2012 and 2013; destroying human lives, farmlands, properties; and causing outbreaks of cholera and other diseases (WMO 2013, 2015). The latest in the strings of multiple extreme wet events in the region is the recent mudslides and torrential flooding in Freetown, the capital city of Sierra Leone. According to several online reports, for example, the American Broadcasting Company (<http://abcnews.go.com/International/wireStory/official-200-dead-sierra-leone-floods-49204542>), more than 400 people died in Sierra Leone, while at least 3000 people were rendered homeless as a result of this torrential flood and mudslides. Intense rainfall, deforestation, and other forms of human activities

have been identified as immediate and preliminary factors for this deadly disaster (<http://news.nationalgeographic.com/2017/08/sierra-leone-mudslide-deadly-video-spd/>). As recently chronicled in West Africa (Ndehedehe et al. 2016b; Nka et al. 2015), wet conditions and flood phenomenon in some Sahelian catchments are becoming more frequent and severe. While the recovery of monsoon rainfall and the continued increase in the occurrence of extreme weather events will unavoidably trigger large changes in the region's TWS, more flood-related investigations are required to understand the vulnerability of the region to extreme wet conditions.

Eco-hydrological processes

Ecosystem dynamics

Studies of ecosystem dynamics in Africa are gradually emerging (e.g., Knauer et al. 2014; Hély et al. 2006) and show an increasing evidence of some level of disturbance. The decline of primary production, widespread desertification, and land degradation are, for example, well-known fallouts of unprecedented droughts and water deficit conditions in the region (e.g., Shiferaw et al. 2014; Bader and Latif 2011; Tucker et al. 1991). A recent study on Sahelian annual vegetation growth and phenology by Pierre et al. (2016) shows that drought shortened the mean vegetation cycle and reduced its amplitude. They argued further that despite the recovery of rainfall in the 1990s, the current conditions for green and dry vegetation are still below pre-drought conditions. It is mentioned in Lovett et al. (2005) that climate change will have major impacts on biodiversity with increased social consequences arising from ecological disturbance. This, as they have further highlighted, will have considerable implications on the continent's readiness to alleviate poverty and meet the Millennium Goals. Large segments of the West African region are heavily reliant on rainfed agriculture, making the impacts of climate variability on freshwater, vegetation dynamics, and weather conditions more deleterious and devastating in the region. As it turns out, economic development is hampered by strong hydrological variability caused by the impacts of climate variability in the region (e.g., Brown and Lall 2006). This may lead to critical water infrastructure which needs to help mitigate the impacts of hydrological variability on food production and livelihood.

Hydrological indicators for surface vegetation dynamics

Numerous studies in the past have used various remotely sensed precipitation and soil moisture products to investigate the water-driven variability of surface vegetation (see, e.g., Guan et al. 2014; Knauer et al. 2014; Chen et al. 2014; Seghieri et al. 2012; Huber et al. 2011; Do et al. 2005; Nicholson

et al. 1990). Because of the complex combination of environmental, social, and multiple strings of anthropogenic factors in Africa, these products are somewhat restricted as drivers of vegetation dynamics. To this end, a new hydro-geodetic approach to understanding the water-driven variability of surface vegetation (the normalized difference vegetation index (NDVI) used as a vegetation proxy) using terrestrial water storage (TWS) inverted from Gravity Recovery and Climate Experiment (more details provided in “Gravity

recovery and climate experiment” section) data is explored in this section. To examine the suitability of other hydrological indicators (e.g., GRACE-derived TWS) for the assessment of climate impacts on eco-hydrological processes, this section evaluates the NDVI–rainfall and NDVI–TWS relationships with focus on the Lake Chad basin. As illustrated in Fig. 4a, b, TWS is modestly associated with NDVI at the annual scale unlike rainfall, which indicated a somewhat weak association with NDVI. A considerably strong linear

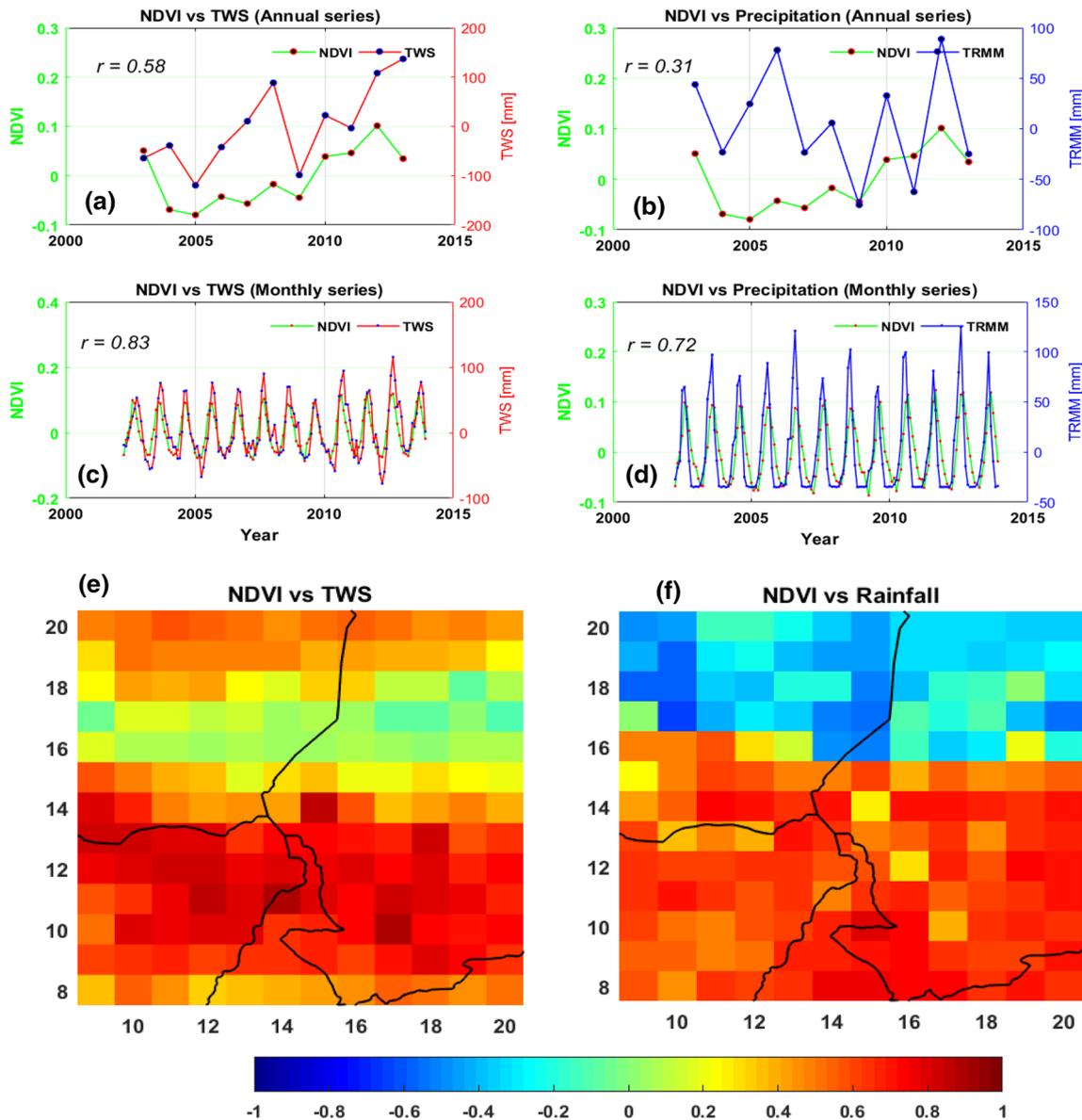


Fig. 4 Association of temporal series of NDVI with TWS and rainfall during the 2003–2013 period in Lake Chad basin—LCB (latitudes 8.25°N–20.25°N and longitudes 9.75°W–20.25°E). **a, b** Correlations of annual departures of aerial averaged NDVI with TWS and tropical rainfall measuring mission (TRMM)-based precipitation. Before computing the annual departures, the annual data (NDVI, precipita-

tion, and TWS) were aggregated from their corresponding monthly time series. **c, d** Correlations of monthly departures of NDVI with TWS and TRMM-based precipitation. **e, f** Spatial correlation coefficients of NDVI with TWS and rainfall, respectively, over the LCB using simple Pearson’s correlation (N/B the scale bar indicates the range of correlation coefficients, i.e., -1 to +1)

correlation of monthly temporal series of NDVI with TWS ($r = 0.83$) and rainfall ($r = 0.72$) is largely indicative of the coupled effects of water availability on surface vegetation changes (Fig. 4c, d). Since changes in TWS components in the Sahelian countries are most likely to emanate from catchment stores (e.g., soil moisture, groundwater, etc.), the strong association of temporal series in NDVI with TWS would be expected. For semiarid ecosystems, the GRACE hydrological signals are changes in the terrestrial water column that includes both the saturated and unsaturated soil zones. Moisture changes in this unsaturated zone as mentioned in Yang et al. (2014) are relatively more sensitive to the impacts of climate variability. Consequently, GRACE-derived TWS changes would mostly be observed moisture changes within the unsaturated zone where the root zone is located. Hence, the result in Fig. 4a–d presents TWS as a rather suitable hydrological indicator on the temporal variations of NDVI in the region. Other land surface processes associated with vegetation dynamics (e.g., evapotranspiration) and the soil characteristics of the region (e.g., Lopez et al. 2016) may contribute to TWS as a better predictor of vegetation dynamics compared to rainfall in the basin. The observed associations of NDVI with rainfall and TWS as indicated in Fig. 4a–d may be dynamic depending on the trajectory of human interference (i.e., in terms of land use change) in the ecosystem and climate change. Such relationship nonetheless provides new perspectives on the ecology of semiarid regions in Africa.

The grid-by-grid relationship of monthly NDVI with TWS and rainfall also indicates the strong potential of TWS and rainfall as hydrological controls in the region (Fig. 4e, f). Overall, TWS shows linear and a more consistent association with NDVI over the entire basin compared to rainfall. For example, TWS shows strong positive correlations with NDVI in the northern and southern catchments of the basin (Fig. 4e), while rainfall shows negative correlations with NDVI (Fig. 4f) in the northern catchment of the basin (rainfall also shows positive correlations with NDVI in the southern flank of the basin). The negative correlations of rainfall with NDVI in this northern catchment (Fig. 4f) suggest that NDVI still oscillates during long and protracted periods of limited rainfall. This perhaps can be caused by the physiological adaptations of Sahelian vegetation where root depth extends to the water table region (Huber et al. 2011; Seghieri et al. 2012). The only indicator of water availability would therefore be soil moisture and other components of catchments stores, which are all represented in TWS. Looking closely, correlations of NDVI with rainfall are predominantly 0.70 or less between latitudes 9°N and 13°N unlike the NDVI–TWS correlations that are mostly between 0.80 and 1.0 (Fig. 4e, f). NDVI correlations with TWS are, moreover, even stronger in the vicinity of the Lake Chad. The strong association of NDVI with TWS in some

parts of southern Chad and northeast Nigeria could be due to changes mostly within the unsaturated soil zone. On the other hand, given that arid and semiarid regions are mostly groundwater-dependent ecosystems, this relationship may suggest complex water use mechanisms, possibly from groundwater (e.g., Guan et al. 2015). The development of such complex eco-hydrological interactions is more likely to emanate from the morphological and physiological adaptations of plants in arid regions that help them maximize water use. It is also critical to note that surface vegetation dynamics of some Sahelian ecosystems are not just driven by climatic factors as anthropogenic influence has been linked to observed changes in vegetation greenness (e.g., Herrmann et al. 2005). It is worthy of note that both at the annual and monthly timescales, TWS is a relatively better indicator of water availability compared to rainfall in LCB, suggesting that interannual variations in TWS can influence crop yield. In such arid regions where rain gauge data and hydrological information are considerably sparse and difficult to acquire, this could further articulate our understanding of the impacts of climate variability on agricultural production and food security.

The role of human interventions on terrestrial hydrology

Anthropogenic factors such as land use change and dam constructions have contributed to freshwater variability. For instance, the impacts of reservoirs and construction of dams along most of the world's major rivers have dramatically change the seasonal flow rates and trends in stream flow rates (e.g., Dai et al. 2009; Yang et al. 2004; Lamers et al. 2001). Human-induced changes in the ecosystem, accompanied by long-term rise in water tables and increase in recharge, were reported for southeast Australia and southwest USA (see Scanlon et al. 2005; Allison et al. 1990). In West Africa, despite the severe drought conditions of the 1970s and 1980s, extensive network of well observations revealed that groundwater resources and water table in Niger increased tremendously (Favreau et al. 2009; Leduc et al. 1997, 2001). This hydrological paradox, popularly known as the “Sahelian paradox,” was attributed to a change in land clearing and changes in land use, which caused an increase in runoff (Descroix et al. 2009; Favreau et al. 2009; Séguis et al. 2004). Apparently, this phenomenon has also been observed in other Sahelian countries of West Africa (see, e.g., Gal et al. 2017; Mahé and Paturol 2009; Mahé and Olivry 1999) and new studies attempting to understand this hydrologically complex scenarios are emerging (e.g., Gal et al. 2017). Although Descroix et al. (2009), in addition, argued from a hydrological stand point that this observed anomaly could possibly emanate from the hydraulic conductivity of the soil and its infiltration capacity, Li et al. (2007)

confirmed the impacts of land use change on the hydrological regimes of Niger and Lake Chad basins. They showed that complete deforestation increased simulated runoff ratio from 0.15 to 0.44 and annual stream flow by 35–65%. The overwhelming litany of evidence from these considerable case studies may lead to the assumption this phenomenon is typical of the Sahel. But this appears not to be the case as a closer look at these Sahelian paradox studies indicate they were conducted at considerably small watershed scales (e.g., 500 km² and 5000 km²). Other hydrological assumptions could lead to the conclusion that this is how the Sahel responds to the composite impacts of water deficit and land use change. A large-scale groundwater assessment from combined GRACE and model observation therefore could provide a broader perspective and perhaps more knowledge on this hydrological phenomenon and the hydrodynamics of the region. Such assessment can benefit from the multivariate methods discussed in “Higher-order statistical tools and frameworks for spatiotemporal analysis of hydrometeorological data” section.

Some recent contributions in the field of remote sensing hydrology to water resources development in West Africa have shown how deliberate human interventions can induce hydrological variability. For instance, Ndehedehe et al. (2017a) recently found that water ponding by the Akosombo Dam in the Lake Volta accounted for about 41% of the increasing trend in GRACE-TWS despite an apparent significant decline in precipitation between 2002 and 2014. The observed increase in TWS over the Volta Basin despite a significant fall in more than a decade precipitation during the same period (e.g., Ndehedehe et al. 2017a; Moore and Williams 2014; Ahmed et al. 2014) re-emphasizes the implications of human water management strategies on hydrology. The installations of hydraulic infrastructures in Senegal triggered rapid evolution of water resources as opposed to the 1980s when hydrological conditions were rather natural (Ngom et al. 2016).

The contraction of the Lake Chad’s surface area was primarily caused by strong declines in regional precipitation patterns, restricting inflow from Chari–Logone river systems, which provide about 95% freshwater of the Lake. However, as indicated in several studies (see, e.g., Ndehedehe et al. 2016b; Birkett 2000; Coe and Foley 2001), water withdrawals from the Lake for irrigation purposes during water deficit periods compounded the effects of extreme droughts, resulting in about 90% decline of the Lake’s surface area (Fig. 3). Apparently, this underscores the role of human activities in not only exacerbating the impacts of climate variability but also reshaping the stability of Lake’s ecosystem. The human-modified droughts of the Lake Chad basin are typical of the Anthropocene, where various forms of human activities impact on catchment storage, soil properties, and hydrological processes,

thereby modifying hydrological drought severity. Van Loon et al. (2016), for example, argued generally that urbanization impacts on recharge and infiltration rates, while soil moisture is influenced by deforestation, afforestation, desertification, and agricultural practices through evapotranspiration. In view of the aforementioned case studies, it can be reconciled that the combined effects of climate variations and human interventions, conjugated with the well-known strong land–atmosphere coupling in the region (e.g., Boone et al. 2009; Koster et al. 2004), will unarguably result in strong and profound influence on water resources systems and hydrological processes.

Earth observations and model simulations in terrestrial hydrology

Advances in remote sensing hydrology evidenced in new space-borne measurements in particular have made it possible to provide reliable estimates of precipitation, soil moisture, streamflow, lakes, soil moisture, glaciers and ice sheets, etc., from orbiting platforms (see, e.g., Petropoulos et al. 2015; Alsdorf et al. 2007; Lettenmaier 2005). Numerous studies (see, e.g., Agutu et al. 2017; Carrao et al. 2016; Funk et al. 2015; Dardel et al. 2014; Chen et al. 2014; Zhou et al. 2014; Du et al. 2013; Wagner et al. 2009; Sheffield et al. 2009; Tucker et al. 1991) have demonstrated the innate potentials of satellite data obtained from several remote sensing platforms such as optical, thermal, and microwave, in drought monitoring, land use change, regional water balance, and hydrological applications.

Multi-resolution data in land surface hydrological studies

Global reanalysis (e.g., Modern-Era Retrospective Analysis for Research and Applications—MERRA) and model data (e.g., WaterGap Global Hydrology Model—WGHM, CPC soil moisture, Global Land Data Assimilation System—GLDAS, etc.) driven by observed meteorological forcing data have also been introduced in understanding global changes in climates and hydrological cycle (see, e.g., Döll et al. 2014; Rienecker et al. 2011; Rodell et al. 2004; Fan and Dool 2004; Dirmeyer et al. 2004). The WGHM model data provide estimates of groundwater recharge from surface water bodies in semiarid and arid regions (Döll et al. 2014), while global reanalysis such as the MERRA data provides atmospheric fields, water fluxes, and global estimates of soil moisture, which are very useful in land surface hydrological studies (Rienecker et al. 2011; Reichle et al. 2011). With the rapid upsurge in wireless and smartphone technologies; unmanned drones, aerial vehicles, and tethered balloons, scientists can now archive estimates of daily

average rainfall, temperatures, floods; map snow depths; and monitor other critical hydrological quantities such as channel depth (McCabe et al. 2017). These sophisticated sensing platforms provide the remote sensing community and hydrologists with a plethora of opportunities to develop new frameworks that advance and facilitate our understanding of global changes in hydro-climatic conditions.

Be it satellite, model, or reanalysis data, their applications have enhanced and revolutionized our knowledge of hydrological cycle, water availability, and global change. For example, using a model soil moisture driven by a hybrid reanalysis–observation forcing dataset, Sheffield and Wood (2008) reported a small wetting trend in global soil moisture (1950–2000) that was attributed to increasing precipitation. Based on new data from the WaterGAP integrated global water resources model, Brauman et al. (2016) recently found periodic water shortage in 71% of world irrigated areas and 47% of large cities suggesting that water security depends largely on reducing society’s vulnerability to water shortages. Multiple satellites and model data have been combined to produce a comprehensive indicator that provides balance for characterization while reflecting aspects of meteorological, agricultural, and hydrological drought information (e.g., Mishra et al. 2015; Niu et al. 2015; Du et al. 2013; Vicente-Serrano et al. 2012; Corzo Perez et al. 2011; Sheffield and Wood 2008). A new multivariate drought framework based on a fourth-order cumulant statistics (Ndehedehe et al. 2016c) also demonstrated the potentials of integrating multiple climate variables (in situ, model, and satellite data) in hydrological drought characterization, contributing to a broad framework of existing methods (“Higher-order statistical tools and frameworks for spatiotemporal analysis of hydrometeorological data” section). Even in a more recent study in East Africa (Agutu et al. 2017), a suite of remote sensing, land surface models, and reanalysis products showed strong potential for agricultural drought characterization, providing alternatives for the in situ data deficient region. Validated satellite precipitation (e.g., Tropical Rainfall Measuring Mission) and soil moisture products have been used to study meteorological processes and map surface soil moisture conditions in West Africa (e.g., Paeth et al. 2012; Pellarin et al. 2009; Nicholson et al. 2003). Net precipitation and other water budget quantities (e.g., rainfall, evapotranspiration, etc) estimated from Global Precipitation Climatology Center, ERA-interim, and GLDAS data over West Africa during the 1979–2010 period were analyzed recently by Andam-Akorful et al. (2017) using wavelet power transforms and coherence analysis. One significant aspect of the findings of the study is that observed decreasing rate of available freshwater (net precipitation) is highly coupled to a low-frequency modulating El Niño activity that induced lower changes in rainfall variance, as well as higher evaporation variance. Put together, these are some obvious

indications of the opportunities that exist for large-scale hydro-climatic research that leverage on satellite programs and other multi-resolution data.

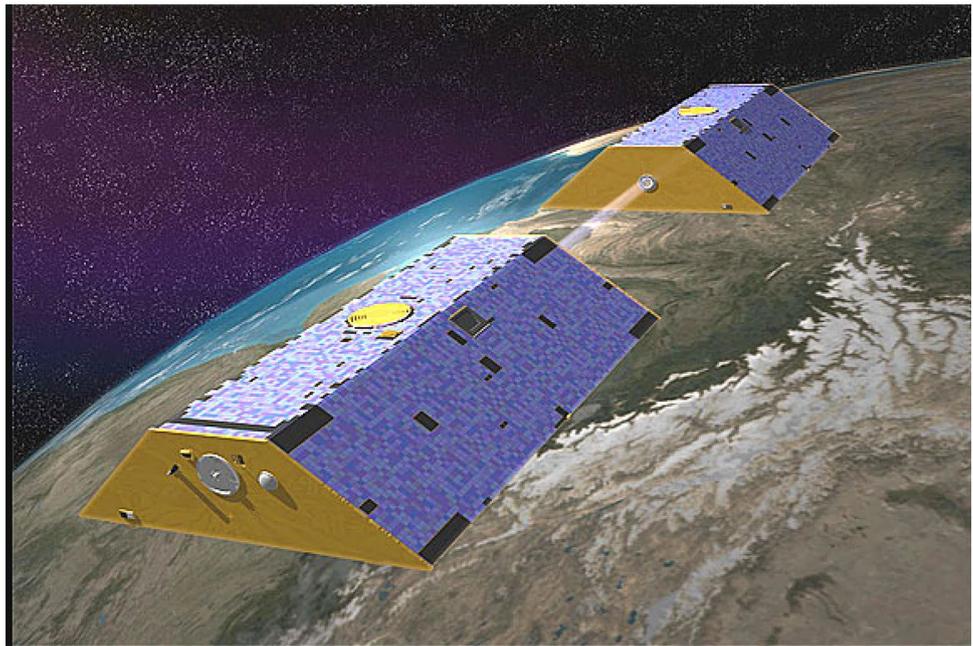
However, some of these datasets are problematic in West Africa as their skills in representing land surface and state variables (e.g., soil moisture) are restricted due to various human interventions, e.g., land use change, surface water developments (Ndehedehe et al. 2018b), and the factors mentioned earlier in “Limited observational networks” section. For this reason, hydrological research in West Africa may tend to benefit more from independent state-of-the-art satellite observations, e.g., the Gravity Recovery and Climate Experiment. This is because the lack of sufficient in situ data for regional configurations and adequate initialization of outputs from hydrological models and reanalysis data for hydrological applications results in limited skills and poor representation of the West African land surface and hydrological state variables. Hence, the continued evaluation and validation of space-borne measurements and outputs from model simulations to assess regional water cycle should be given attention.

Gravity recovery and climate experiment

The Gravity Recovery and Climate Experiment (GRACE, Tapley et al. 2004) is a sophisticated time variable satellite gravity mission, comprising two identical satellites in near-circular orbits with an altitude of ~ 500 km and an inclination angle of 89.5° that are separated by ~ 220 km (Fig. 5). Unlike optical remote sensing platforms, which use visible, near-infrared, and short-wave infrared sensors to collect details of the earth surface, the two GRACE satellites use the satellite-to-satellite tracking system (Fig. 5). Each of the GRACE satellites is further equipped with a K-band microwave ranging system and high-precision accelerometers. Changes in the Earth’s gravity field are caused mostly by dynamic processes that include mass redistribution of water (i.e., continental water storage, ocean, and atmosphere), gravitational tide in the solid Earth, post-glacial rebound, and variations in Antarctic and Greenland ice volumes (e.g., Swenson and Wahr 2002; Wahr et al. 1998). Accurate observations of these gravity fields are obtained from precise measurements of the distance between the two GRACE satellites as they orbit the Earth (see, e.g., Chen et al. 2005; Swenson and Wahr 2002; Wahr et al. 1998).

Since its inception, GRACE has given an unparalleled perspective to global terrestrial hydrology by providing quantitative estimates of monthly changes in TWS (soil moisture; groundwater; surface water—lakes, rivers, wetlands; snow; and canopy) over large spatial scales. Because of its spatial resolution ($200,000 \text{ km}^3$), the dynamics in multilayered land water storage (i.e., all aspects of TWS) can be measured at global or regional scales with an accuracy

Fig. 5 Illustration of the Gravity Recovery And Climate Experiment (GRACE) satellite mission. The two identical twin satellites popularly known as *Tom* and *Jerry* are linked with intersatellite ranging system and equipped with global positioning system receivers and attitude sensors and high-precision accelerometers (Tapley et al. 2004). The ranging system is super-sensitive and can detect separation changes about one tenth the width of a human hair over a distance of 220 km (NASA 2002). The GRACE satellites are designed to map the global gravity field every thirty days. Image adapted from <http://spaceinfo.com.au/2010/07/23/gravity-mission-down-under/>



of 1.5 cm equivalent water height (Famiglietti and Rodell 2013). Apart from its broad applications in droughts, floods, terrestrial water budget, and ecosystem assessments (see, e.g., Zhang et al. 2016; Ndehedehe et al. 2016c; Thomas et al. 2014; Long et al. 2014; Reager et al. 2014; Yang et al. 2014; Long et al. 2013; Houborg et al. 2012; Chen et al. 2010), it is now one of the most vital tools in hydrological research, specifically in monitoring subsurface water storage, aquifer system processes, and evaluating groundwater resources (see, e.g., Castellazzi et al. 2016; Famiglietti et al. 2015; Alley and Konikow 2015; Famiglietti and Rodell 2013; Alsdorf et al. 2010; Henry et al. 2011; Tiwari et al. 2009; Swenson and Wahr 2007). The applications of GRACE data in hydrology research globally are growing and well documented (e.g., Humphrey et al. 2016; Andam-Akorful et al. 2015; Senay et al. 2014; Sneeuw et al. 2014; Ndehedehe et al. 2018c; Wouters et al. 2014, and the references therein) and are only summarized here.

GRACE hydrological applications in West Africa

Early GRACE hydrological applications in Africa were mostly focused on other subregions (East, Central, and North Africa, e.g., Gonçálves et al. 2013; Awange et al. 2013a, b; Lee et al. 2011; Xie et al. 2012; Ramillien et al. 2008; Crowley et al. 2006). However, GRACE-based hydrological studies are gradually emerging in West Africa (see Ndehedehe et al. 2018a, c; Werth et al. 2017; Ndehedehe et al. 2016a, 2017b; Forootan et al. 2014; Lee et al. 2011; Henry et al. 2011). With the validation of GRACE data in the region (Nahmani et al. 2012; Grippa et al. 2011; Hinderer et al.

2009), more studies will be required to deploy GRACE observations in hydro-geological, groundwater, and climatological studies in the region. The utility of GRACE data for identifying long-term regional changes in groundwater storage in southern Mali and Niger basin has been reported (Werth et al. 2017; Henry et al. 2011). The outcomes in the Niger basin and southern Mali indicate opportunities for assessing groundwater dynamics in other major river basins in West Africa. The developments and evolutions of GRACE-derived TWS over West Africa as highlighted in previous reports (“Understanding space–time GRACE hydrological signals” section) suggest complex hydrological processes for the humid parts due to the presence of surface waters, wetlands, and a considerably strong interannual rainfall (see Ndehedehe et al. 2017a, b; Ferreira and Asiah 2015; Moore and Williams 2014; Ahmed et al. 2014). An array of novel findings and interesting insights in the aforementioned areas has been reported recently (see, e.g., Ndehedehe et al. 2016b, 2017a, b; Ferreira and Asiah 2015) and will advance essential aspects of research in remote sensing hydrology in the region.

Some recent perspectives on drivers of land water storage dynamics

The analysis of linear trends in yearly and monthly rainfall totals over West Africa (1980–2010) reveals that the Sahel rainfall has shown recovery marked by statistically significant increase in annual totals (Sanogo et al. 2015). The leading modes of standardized precipitation index over West Africa were also marked by above-normal rainfall

(wet conditions) mostly between 1990 and 2014 over much of the Sahel (Ndehedehe et al. 2017b). These positive trends in rainfall coincide with some significant wetting patterns in observed long-term model derived land water storage over much of the Sahel (Ndehedehe et al. 2018b). Although the 2002–2006 period was somewhat dry in most regions of West Africa due to reduced precipitation, the extreme wet conditions of the 2006–2007 period caused by La Niña triggered anomalous increase in Lake Volta water levels and GRACE-derived TWS over most parts of West Africa (e.g., Ndehedehe et al. 2016a). Regardless of precipitation trends, GRACE-derived TWS has shown strong dominant patterns along the Guinea Coast countries of West Africa and significant increase over the entire region during the last decade (e.g., Ndehedehe et al. 2016a; Ahmed et al. 2014). These results align with a more recent GRACE study in the Niger basin, which reported an estimated linear rates of $93 \pm 61 \text{ km}^3$ in groundwater stocks between 2003 and 2013 (Werth et al. 2017).

As illustrated further in basin-specific studies (Fig. 6a–d), estimated trends in GRACE-derived TWS anomalies using the basin scale approach of Swenson and Wahr (2002) over Lake Chad ($2.32 \pm 1.17 \text{ mm/year}$), Niger ($5.81 \pm 1.19 \text{ mm/year}$), Senegal ($5.22 \pm 1.02 \text{ mm/year}$), and Volta ($13.06 \pm 1.85 \text{ mm/year}$) basins indicate marked increase during the 2002–2014 period. The highest increase in TWS anomalies is observed over the Volta Basin due to strong gravimetric contributions caused by water ponding in the Akosombo Dam as highlighted in previous studies (see Ndehedehe et al. 2016a, 2017a; Moore and Williams 2014; Ahmed et al. 2014). In other basins nonetheless, these linear rates and trends (Fig. 6a–d) are largely perceived as a result of improved rainfall conditions over the same period. In light of the observed pseudo-rise and contributions of Lake Volta to TWS over the Volta Basin, which has been linked to water ponding by the Akosombo Dam (Moore and Williams 2014; Ahmed et al. 2014), understanding hydrological variability of the basin became challenging. For instance, although precipitation, discharge from the Volta river system and net

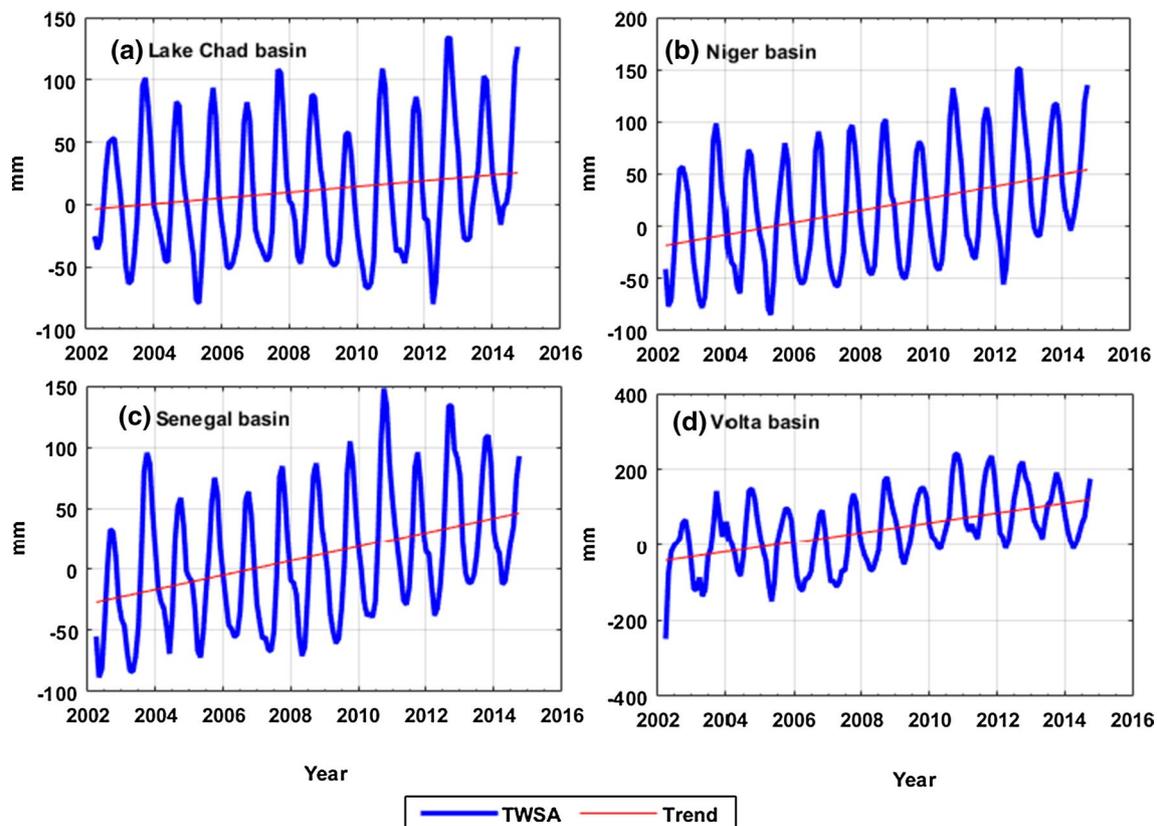


Fig. 6 Trends in GRACE-TWS anomalies over major river basins in West Africa for the period between 2002 and 2014. Trends in TWSA over **a** Lake Chad, **b** Niger, **c** Senegal, and **d** Volta Basins were estimated using the basin scale approach described in Swenson and Wahr (2002). The GRACE (Tapley et al. 2004) satellite mission provides changes in TWS (sum total of changes in groundwater, soil moisture,

surface waters, and canopy) based on observations of the Earth's time variable gravity fields (e.g., Ndehedehe et al. 2016a; Wouters et al. 2014). The GRACE Release-05 (RL05) spherical harmonic coefficients from the center for space research (CSR), truncated at degree 60, covering the period 2002–2014 were used to estimate TWS over these basins

precipitation reduced significantly over a 10-year period that coincided with the GRACE period, TWS over the Volta Basin showed significant and considerable rise during the same period (see Ndehedehe et al. 2017a). By exploring a novel approach based on a two-step procedure that incorporates a weighted least squares formulation of global spherical harmonic analysis and cumulant statistics, a further step was undertaken by Ndehedehe et al. (2017a) recently, to quantify the contribution of Lake Volta to observed trend in GRACE-derived TWS over the Volta Basin (2002–2014). They showed that the observed decline in GRACE-derived TWS (i.e., after removing the strong gravimetric changes of the Lake Volta) was consistent with more than a decade decline in precipitation within the vicinity of the lower Volta Basin where the lake is located. This decline in precipitation over the Volta Basin also coincided with significant reduction in stream flows of the Volta river system and net precipitation. It is here noted that large surface water development schemes such as the Lake Volta will have direct implications in water budget assessment. Hence, the combined impacts of anthropogenic, climatic, and natural drivers (e.g., natural climate variability) are challenging issues that require an integration of science and policy in water resources planning in the basin.

Whereas rainfall at interannual and seasonal timescales drives TWS in West Africa (Ndehedehe et al. 2016a), drivers of hydrological variability are somewhat unclear in the Congo Basin (i.e., much of Equatorial Africa). The drying trends observed in the northern Congolese forest were found to be generally consistent with declines in rainfall, TWS, and aboveground woody and leaf biomass (Zhou et al. 2014). But observed declines in TWS in three sub-basins (Congo, Ubangi, and Sangha) of the Congo river basin during the 2003–2012 period were attributed to deforestation (Ahmed et al. 2014). Pioneer GRACE results over the Congo Basin showed that it lost about 280 km³ of water between April 2002 and May 2006 largely caused by a loss in runoff and evapotranspiration (Crowley et al. 2006). One may therefore argue that the hydrological conditions of the Congo Basin and much of Central Africa are somewhat complex. Apart from non-climatic drivers, the local influence of sea surface temperature (SST), atmospheric circulation features, and mesoscale convective systems, which regulate rainfall through their control on the rain belt, have also been reported (e.g., Farnsworth et al. 2011; Balas et al. 2007). The Congo Basin is an epic biodiversity that houses the world's second largest rainforest. As highlighted by Washington et al. (2013), the basin is among the three prominent convective regions on Earth that dominates global rainfall climatology during transition seasons. Its hydrological characteristics, nonetheless, are less reported compared to the Amazon basin where a relatively large inventory of hydrology-related studies exists (Alsdorf et al. 2016). Hence, more studies will be

required to build a concrete contemporary understanding of hydrological processes, climate feedbacks, and interactions in the region's biophysical systems.

The influence of global climate on terrestrial water storage

Considerable case studies in the past have investigated the impacts of perturbations of the nearby oceans, quasi-periodic phenomena (e.g., ENSO), and other low-frequency climate oscillations on precipitation changes in Africa. While such impacts have been widely reported and less debated, the role of global climate on long-term terrestrial water storage (TWS) has not been studied. Only recently was the influence of three prominent climate teleconnection indices (ENSO, AMO, and IOD) on GRACE-derived TWS and global reanalysis data documented over West Africa by Ndehedehe et al. (2017b). The study revealed that ENSO and other low-frequency climate oscillations of the Atlantic Ocean (AMO) are important drivers of long-term variations in TWS, affecting its distributions and temporal changes. Prior to this, some pioneering studies (see Phillips et al. 2012; Boening et al. 2012) drew the attention of the hydrological community to the impacts of ENSO on global mean sea level and continental water storage. In other reports (e.g., Sheffield and Wood 2008), variations in global soil moisture have been associated with variabilities in ENSO, though AMO supposedly impacts soil moisture variations in other regions including West Africa. Interestingly, ENSO, AMO, and PDO among other climate modes play key roles in the characteristics of extreme climatic conditions in West Africa (e.g., Ndehedehe et al. 2016b, c; Paeth et al. 2012).

Furthermore, new results re-emphasizing and validating the influence of ENSO on variations in TWS are emerging (e.g., Ni et al. 2017). The study by Ndehedehe et al. (2018b) found ENSO-related equatorial Pacific SST fluctuations as a major teleconnection with considerable significant association with long-term TWS reanalysis data over large portions of sub-Saharan Africa. ENSO is arguably one of the leading modes of oceanic variability that explains variability in TWS as the aforementioned studies suggest. However, for West Africa, there are other climate oscillations of the Atlantic SST that are also important drivers of long-term variations in TWS. The observed presence of AMO-driven TWS in much of the Sahel and Guinea Coast regions and the modest association between variabilities in Atlantic Meridional Mode and TWS over West Africa suggest the important role of Atlantic SST variability in the region (Ndehedehe et al. 2017b, 2018b). Away from these large-scale climate teleconnection impacts, Hoff (2009) identified some biophysical, socioeconomic, and institutional teleconnections that exist within global water systems. These teleconnections whose characteristics have been detailed in the aforementioned synopsis, link causes, and effects on water resources

remotely and may lead to serious disturbance and perhaps a breakdown of water-dependent socio-ecological systems. As anthropogenic factors also play key roles in the dynamics of global water systems (“Earth observations and model simulations in terrestrial hydrology” section), water management decisions will have to embrace dynamic policy solutions and meaningful decisions that integrate both climate and human factors.

The future of satellite geodetic missions in hydrology

Apart from the growing applications of GRACE observations in hydrology and Earth science, other sustained investments in satellite geodetic programs have provided large coverage of dynamics in terrestrial water bodies. For example, measurements obtained from interferometric synthetic aperture radar (InSAR), radar altimetry (TOPEX/Poseidon), ERS satellites, Envisat and Jason1 missions, Ice, Cloud and land elevation satellite (ICESat), Shuttle Radar Topography Mission (SRTM), Global Positioning System (GPS), Japan’s Advance Land Observing Satellite (ALOS), among others, have proved to be useful in understanding the impacts of climate variability on the temporal and spatial dynamics of surface water resources (e.g., Tourian et al. 2017; Ndehedehe et al. 2017a; Lee et al. 2014; Alsdorf et al. 2001, 2003, 2007; Frappart et al. 2006; Lettenmaier 2005; Birkett 1995, 2000). All of these satellite programs and global land surface schemes have ignited a plethora of scientific findings that are not only informative but also useful for public policy and management decisions related to water resources (e.g., Ndehedehe 2017). Currently now streaming is the use of InSAR data not just for detecting land subsidence but assessing groundwater abstraction-induced drought (e.g., Castellazzi et al. 2016). Some other novel application areas would be to integrate GRACE observations with InSAR to monitor human-induced changes in groundwater aquifers and predict the likelihood of surface displacements in areas with long history of mining operations and groundwater use. Moreover, such independent observations would be essential in the Sahel of West Africa to address the widely reported hydrological paradox of increasing groundwater stocks during drought periods.

The previous GRACE mission came to an end in December 2017 after more than 15 years of gravity observations across the globe. It was replaced by the launch of a new-generation time variable gravity satellite pair, the GRACE Follow-On (GRACE-FO) mission in May 22, 2018. In non-industrialized regions where hydrological infrastructures and observational networks are in decline, non-operational or sparse (e.g., Alsdorf et al. 2003; Alsdorf and Lettenmaier 2003), the GRACE-FO mission would be the only operational, state-of-the-art,

and most sophisticated gravity mission that would benefit hydrological and climate research. The anticipated finer spatial resolution of GRACE-FO mission would allow not just the quantification of the terrestrial water budget components but the closure of water balance with limited uncertainties. The new spatial resolution would be the aftermath of an optimized system design and the inclusion of an experimental laser ranging system to complement the K-band microwave ranging system in the previous GRACE mission.

Another future space agency mission currently under development and scheduled to be launched by 2020 as published by National Aeronautics and Space Administration (NASA, www.jpl.nasa.gov/missions/surface-water-and-ocean-topography-swot) is the Surface Water and Ocean Topography (SWOT) mission. The SWOT mission is expected to provide detailed measurements of surface water storage variations (i.e., wetlands, lakes, or reservoirs), complementing the GRACE mission. As indicated in the SWOT mission document (https://swot.jpl.nasa.gov/files/swot/SWOT_MSD_1202012.pdf), the scientific rationale for the development of SWOT is twofold. Firstly, to make high-resolution, wide-swath altimetric measurement of the ocean surface topography that will advance the understanding of the oceanic mesoscale and sub-mesoscale processes. Secondly, the quest for a mission that will measure the elevation of water on land that will boost fundamental advances in the knowledge of the spatiotemporal distribution of the storage and discharge of terrestrial water. Nevertheless, some other space agency missions relevant to hydrology have also been flagged for deployment in the coming years. They include Water Cycle Observation Mission (WCOM) and NASA’s ICESat-2 (Ice, Cloud, and land Elevation Satellite). WCOM is expected to be launched by 2020 and is designed to measure snow water equivalent, soil moisture, precipitation, atmospheric water vapor, and other state variables, while ICESat-2 is primarily designed to map ice sheet, and in addition monitor surface water elevations (e.g., McCabe et al. 2017). ICESat-2 is the second generation of the laser altimeter ICESat mission and is optimized to use a micro-pulse multi-beam approach. It is scheduled for launch by 2018 to determine ice sheet mass balance as well as cloud property information, provide topography and vegetation data around the globe, plus the polar-specific coverage over the Greenland and Antarctic ice sheets (see <https://icesat.gsfc.nasa.gov/>).

Higher-order statistical tools and frameworks for spatiotemporal analysis of hydrometeorological data

The plenitude of several available climate, hydrological, or gravimetric data, be it in the form of satellite observations, reanalysis, or model-generated products, has enhanced the continuous monitoring of land surface conditions,

biodiversity, weather, and natural systems. During the last decade, these data have been considerably employed across the globe to study various forms of droughts (meteorological, agricultural, and hydrological), floods, ecological processes, and in the characterization of critical hydrological metrics, among others (see, e.g., Agutu et al. 2017; Ndehedehe et al. 2016c; Van Loon et al. 2016; Velicogna et al. 2015; Sheffield et al. 2014; Zhou et al. 2014; Thomas et al. 2014; Long et al. 2014; Reager et al. 2014; Yang et al. 2014; Vicente-Serrano et al. 2012; Sheffield and Wood 2008; Koster et al. 2004). However, the need to understand and localize the multiple and growing climate signals is now in the forefront of climate science. This interest in regionalising hydro-climatic signals at various time and geographic (e.g., global and regional) scales is increasing and has resulted in several applications of multivariate techniques such as principal component analysis (e.g., Preisendorfer 1988) in the decomposition of climate data, geophysical signal separation, and drought analysis (e.g., Ndehedehe et al. 2016a; Kurnik et al. 2015; Sanogo et al. 2015; Ivits et al. 2014; Nicholson 2014; Santos et al. 2010; Bonaccorso et al. 2003; Semazzi et al. 1988). Other approaches to drought variability and or forecasting however, along with improving general circulation model (GCM) predicted rainfall, have

been studied (though less reported) using wavelets analysis, canonical correlation analysis, among others (e.g., Mehr et al. 2014; Okonkwo 2014; Kumar et al. 2013; Singh et al. 2012; White et al. 2004).

Understanding space–time GRACE hydrological signals

Following some of the poor skills of global hydrological models (e.g., Ndehedehe et al. 2018b), advocating for the improved representation of the African land surface and hydrological state variables (e.g., soil moisture) in these models and reanalysis data has been the basis of some emerging regional studies (e.g., Agutu et al. 2017). Whereas this can be considered a logical step that could promote hydrological research, more validation studies would be essential to further assess the uncertainties in these data, establishing their limits of reliability and use in the region. For example, the MERRA reanalysis and WaterGap model cannot reproduce the temporal variability as captured by GRACE-TWS owing to their limited skills in the Congo Basin unlike the Sahel and some Guinea Coast zones (Figs. 7a and 1a).

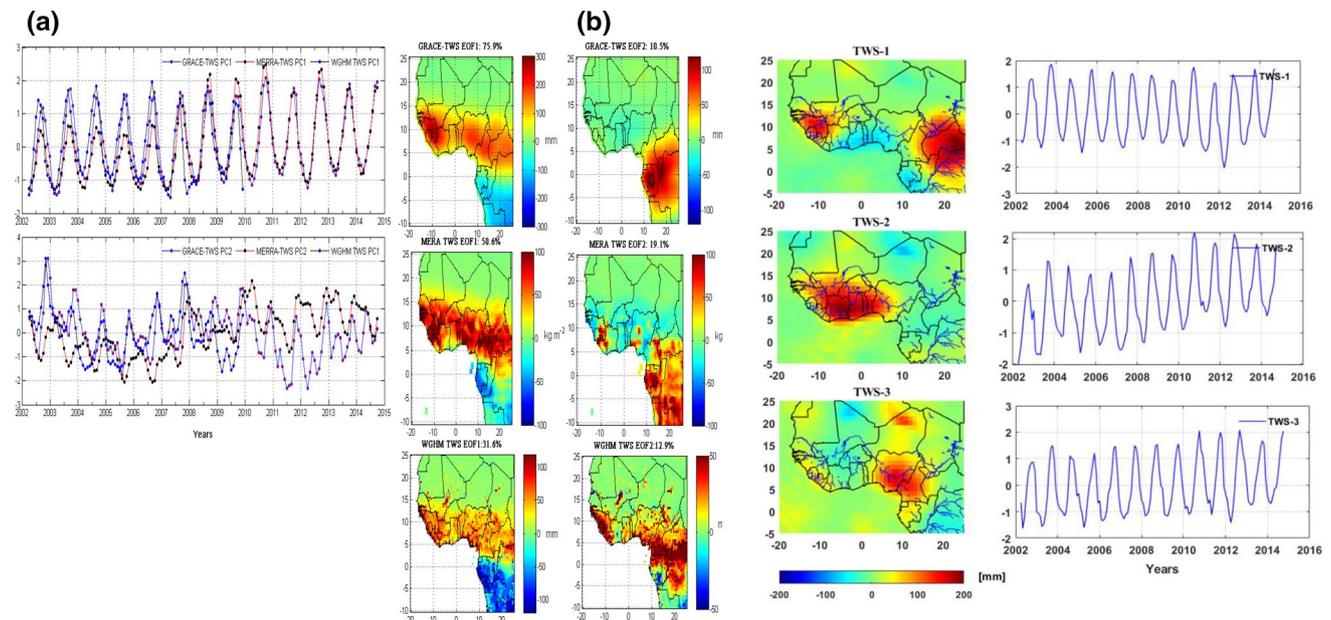


Fig. 7 Illustrating multivariate analyses of land water storage obtained from time-variable gravity observations, reanalyses data, and outputs from hydrological models. **a** Comparing the spatiotemporal patterns of MERRA (2002–2014) and WGHM-based TWS (2002–2009) products over West Africa with GRACE-derived TWS (2002–2014) based on PCA. The EOFs (right) and the corresponding temporal patterns or PCs (left) are constructed similar to Fig. 2. The spatiotemporal patterns of GRACE-TWS here include Central African countries and much of the Congo Basin area (tropical West

Africa—TWA). This map is fully adapted from Ndehedehe et al. (2018b) with permission granted on the November 27, 2018. **b** The space–time patterns of GRACE-derived TWS over TWA using cumulative statistics (e.g., Ndehedehe et al. 2016b; Cardoso and Souloumiac 1993). The independent components (ICs) and spatial patterns are constructed similar to Fig. 8. The blue lines on the spatial patterns corresponding to the ICs are water bodies (rivers, lakes, and reservoirs)

However, in addition to collaborating with local institutions for the provision of existing ground observations for parameterizations and validation purposes, the strong land–atmosphere interaction and the highly variable hydrological conditions should be understood to make necessary adjustments in hydrological models. Understanding this variability and its various drivers has been the basis of some of the author’s pioneering works in the region (e.g., Ndehedehe et al. 2016a, 2018b, c; Andam-Akorful et al. 2017). As the PCA method indicates (Fig. 7a), the two dominant modes of land water storage variability are found at the Guinea Coast/Sahel regions and the Congo Basin area and are consistent with rainfall evolutions (Fig. 2). While the first mode of GRACE-TWS was largely precipitation driven, the strong dynamics of the Congo River accounted for more of the observed variability in GRACE hydrological signals (Ndehedehe et al. 2018b). A further assessment of independent patterns of TWS using cumulant decomposition techniques is illustrated in Fig. 7b. The regionalization of TWS using this technique has been very useful to understanding the evolutionary behavior of TWS in specific locations. As can be seen over the Volta Basin area (TWS-2, Fig. 7b and cf. Fig. 1), the rising trend in the time series associated with the spatial pattern is caused by water ponding of the Akosombo Dam in Ghana, contrary to other regions where temporal patterns are largely influenced by the annual rainfall patterns (TWS-1 and TWS-3, Fig. 7b). Statistical decomposition methods are helpful in identifying climatic drivers such as global teleconnections or other prominent hydrological signals, e.g., variability in groundwater for semiarid regions of the world with massive groundwater abstraction. Space–time evolutions of state variables are essential for improved understanding of variability. The application of these statistical methods on GRACE-derived TWS could become the most efficient approach to quantify and assess the interplay among notable drivers, which include climate change, natural, and human-induced changes that contribute to variations in freshwater stocks. For example, this was demonstrated by Ndehedehe et al. (2018b) who applied PCA to regularized TWS and sea surface temperature anomalies and subsequently used the canonical scheme (canonical correlation analysis) to assess the coupled variability of dynamics in land water storage with SST anomalies. Furthermore, multivariate techniques are also very handy in isolating specific hydrological signals like those emanating from major lakes with strong gravimetric variations (e.g., Ndehedehe et al. 2017a).

Drought characterization and regionalization

To improve the regionalization of drought signals at different aggregation scales, a fourth-order cumulant statistics, the independent component analysis (e.g., Cardoso

and Souloumiac 1993; Cardoso 1991) was recently introduced in region-specific case studies (see Ndehedehe et al. 2016b, c). This cumulant-based approach to the analysis of extreme hydro-climatic conditions has shown great skills in studying space–time developments of droughts compared to previous approaches (e.g., principal component analysis) of regionalizations and can be applied in other regions of the world. Theoretical and mathematical details of cumulant statistics and its applications have been documented, and interested readers can see relevant studies (e.g., Cardoso 1999; Cardoso and Souloumiac 1993; Cardoso 1991; Ndehedehe et al. 2016b, c). As illustrated in Fig. 8, the leading modes of standardized precipitation index (SPI) over West Africa are spatially distinct and in general create a dichotomy of extreme weather events between the semiarid Sahel and the humid coastal areas of West Africa. Drought amplitude and its variability in the region are unique depending on rainfall conditions and perhaps other climatic and human drivers. The notorious droughts of the early 1980s, for example, persisted in the Sahel at the 12-month aggregation scale compared to other regions of West Africa, though the tendency toward wet conditions has increased (with marked fluctuations). Such tendencies are due to recovery of rainfall in the 1990s and improved rainfall conditions during the last decade (Fig. 8). SPI temporal series aggregated at 12-month scale are representative of hydrological droughts (e.g., Ndehedehe et al. 2016b, c; Li and Rodell 2015; Vicente-Serrano 2006; Rouault and Richard 2003; Hayes et al. 1999). Hydrological droughts are usually periods characterized by prolonged deficits in rainfall. This often leads to reduced alimentation, desiccation, and depletion of lakes, reservoirs, stream flows, and, eventually, a decline in groundwater levels, depending on human response to water deficit conditions. This was the case in Lake Chad when protracted periods of hydrological droughts, accompanied with excess water withdrawals from the lake for agricultural purposes, resulted in a humongous contraction of its surface area (Fig. 3). Further, in Fig. 8, this multivariate technique clearly marks out the hydrological drought condition of the Congo Basin between 2003 and 2007 before it experienced the La Niña upsurge later in 2007. It further shows that temporal evolutions of SPI over the Sahel are well associated with AMO unlike the SPI temporal patterns of other regions (Ndehedehe et al. 2017b). This approach supports location-specific assessment of large-scale impacts on extreme climatic conditions. Arguably, this could help reduce the polarizations surrounding the debates on extreme rainfall conditions and the roles of climate modes that induce it variability. This new multivariate technique in spatiotemporal drought analysis is more comprehensive and robust and contributes to a large framework for drought assessment that will complement existing methods.

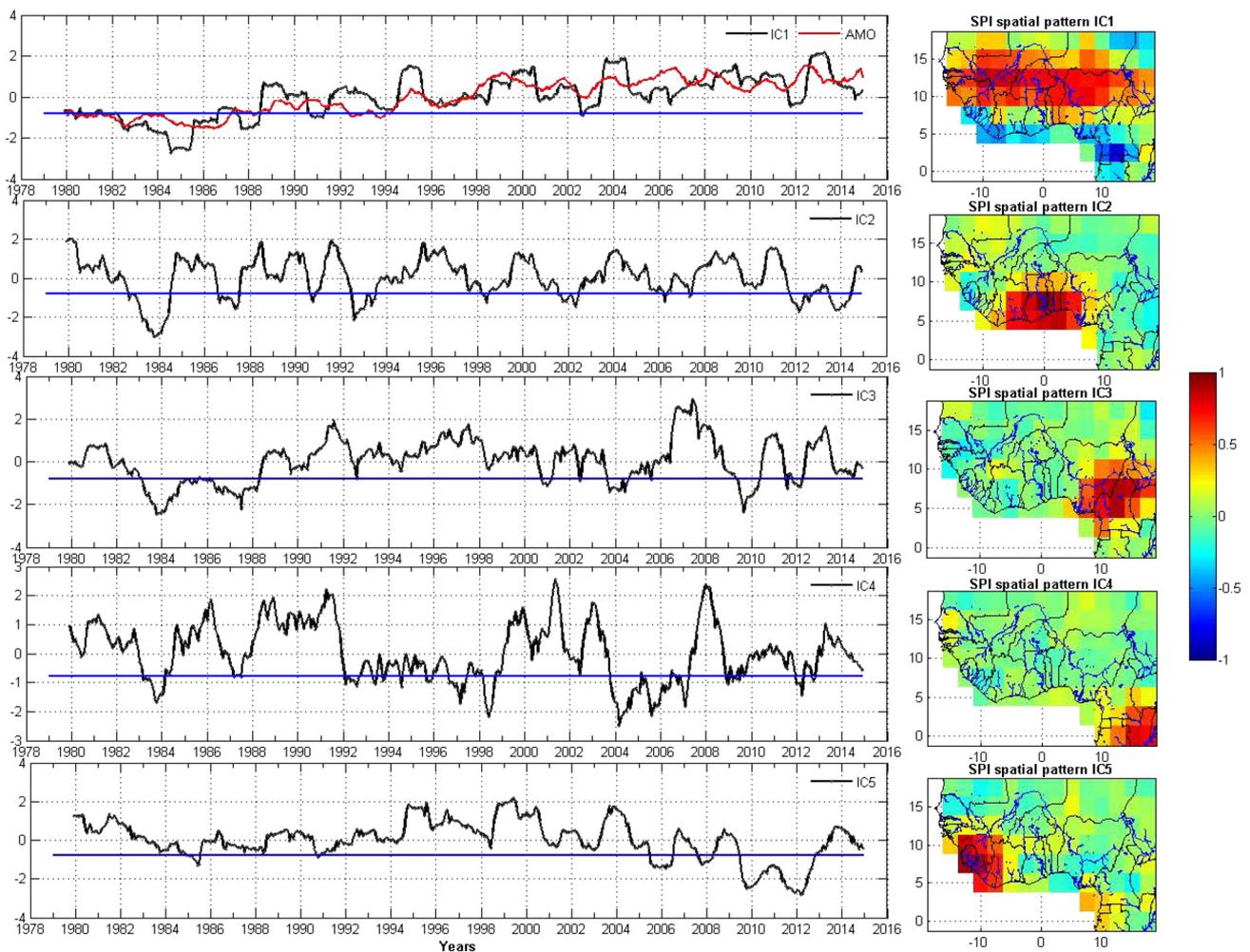


Fig. 8 Illustration of the spatiotemporal SPI (standardized precipitation index) patterns over different hydrological zones of West Africa using 12-months gridded SPI values (adapted from Ndehedehe et al. 2017b). The localized SPI values are based on McKee et al. (1993) description and are computed using GPCP-based precipitation product for the period 1979–2014 (SPI values are in standardized units). The localization of SPI was achieved by rotating significant orthogonal modes of 12-months SPI values using cumulant statistics (e.g., Ndehedehe et al. 2016b; Cardoso and Souloumiac 1993). SPI values for drought classification and characterization are jointly derived and

interpreted together using the temporal evolutions of SPI (left) and their corresponding spatial patterns (right). AMO is associated with the characteristics of SPI over the Sahel region. Blue lines on the SPI spatial patterns are hydrological units (rivers, lakes, and other water bodies), while those on the temporal series are drought thresholds. SPI signals and soil moisture index have also been localized in basin-specific studies over West Africa, where ENSO and several low-frequency climate oscillations such as the PDO, Arctic Oscillation, and North Atlantic Oscillation were associated with drought characteristics (see Ndehedehe et al. 2016c)

Conclusion

The deleterious impacts of unmitigated climate change in West Africa represent significant and considerable negative impacts on agriculture, hydropower, economic development and result in ripple effects on the socioeconomic systems of the region. As most agricultural goods are produced in regions that are vulnerable to water-related impacts, this will have massive implications not just on the economy of West Africa but other regions of the world that indirectly consume the water resources of West Africa. Despite this strategic importance of West African

countries to the global community, there are still considerable and prominent gaps, however, in the knowledge of how global changes in climate impact on the region's water resources systems. Not only is hydrology poorly understood, large-scale temporal and spatial dynamics in TWS, and a framework to characterize key hydrological metrics and extreme weather events are lacking in West Africa. As this review indicates, an increasing number of constraints, e.g., lack of robust investments in gauge measurements for meteorological and hydrological applications among other factors, combine to restrict the availability of in-situ observations for hydro-climatic research

that addresses the aforementioned issues. Advancements in space agency programs and sophisticated sensing platforms undoubtedly have revolutionized terrestrial hydrology research around the world. However, as opposed to other hydrological regions of the world (e.g., the Amazon basin) with considerable reports on hydrological processes, similar studies over West Africa are conspicuously limited in the volumes of existing studies and catalog of hydrological journals.

Understanding the impacts of global climate on regional changes in terrestrial water system is warranted for few reasons: it will (i) support policy and risk management strategies, (ii) promote the need for management and sustainability of water resources, (iii) foster campaign on regional adaptation strategies in the event of extreme hydrological conditions, (iv) advance hydro-climate research in terms of optimizing climate and hydrological models, and (v) provide new perspectives on eco-hydrological processes in West Africa. As this manuscript illustrates, opportunities for hydro-climatic research in West Africa that leverage on sustained investments in satellite geodetic programs exist. The potentials of satellite gravity observations, gauge and high spatial resolution satellite precipitation data in the region, are some opportunities to further assess the representation of the land surface and atmospheric states in global reanalysis models.

The highlights in this manuscript indicate that these geodetic programs (e.g., GRACE and radar altimetry missions), along with other remote sensing satellite missions, and auxiliary data synthesized by forcing global land surface models with historical meteorological data (reanalysis) have only within the last two decades ignited a plethora of scientific findings. These outcomes are not only informative but could be considerably useful for public policy and management decisions related to water resources. Given the poor density of groundwater bores and other observations, hydrological research in West Africa is expected to benefit immensely from independent space geodetic programs such as the GRACE mission. Limited skills in model outputs and poor representation of the West African land surface in reanalysis data are few notable constraints to large-scale terrestrial hydrology-focused investigation in West Africa. Therefore, to improve our contemporary understanding of West Africa's terrestrial hydrology, the continued evaluation/validation of these data (models and reanalysis) and space-borne measurements is advocated.

Acknowledgements The various agencies of government (National Oceanic and Atmospheric Administration, Center for Space Research, and National Aeronautics and Space Administration) are gratefully acknowledged for the data used to support this review paper. This synthesis contains items taken from the literature review component of the author's unpublished PhD thesis, which was completed at Curtin

University, Australia, in 2017. The comments of R W Abrams and two other anonymous reviewers, which helped improved the quality of the manuscript are gratefully acknowledged.

Compliance with ethical standards

Conflict of interest The author declares that there is no conflict of interest.

References

- Agutu N, Awange J, Zerihun A, Ndehedehe C, Kuhn M, Fukuda Y (2017) Assessing multi-satellite remote sensing, reanalysis, and land surface models' products in characterizing agricultural drought in East Africa. *Remote Sens Environ* 194:287–302. <https://doi.org/10.1016/j.rse.2017.03.041>
- Ahmed M, Sultan M, Wahr J, Yan E (2014) The use of GRACE data to monitor natural and anthropogenic induced variations in water availability across Africa. *Earth Sci Rev* 136:289–300. <https://doi.org/10.1016/j.earscirev.2014.05.009>
- Alcamo J, Döll P, Kaspar F, Siebert S (1997) Global change and global scenarios of water use and availability: an application of WaterGAP1.0. Center for Environmental Systems Research (CESR), University of Kassel, Germany. <https://pdfs.semanticscholar.org/a788/80a9f9952f0e04bc88c7c5699ea9c9e0fa30.pdf>. Accessed 18 Aug 2016
- Ali A, Lebel T (2009) The Sahelian standardized rainfall index revisited. *Int J Climatol* 29:1705–1714. <https://doi.org/10.1002/joc.1832>
- Alley WM, Konikow LF (2015) Bringing GRACE down to earth. *Groundwater* 53(6):826–829. <https://doi.org/10.1111/gwat.12379>
- Allison G, Cook P, Barnett S, Walker G, Jolly I, Hughes M (1990) Land clearance and river salinisation in the western Murray Basin, Australia. *J Hydrol* 119(1):1–20. [https://doi.org/10.1016/0022-1694\(90\)90030-2](https://doi.org/10.1016/0022-1694(90)90030-2)
- Alsdorf DE, Lettenmaier DP (2003) Tracking fresh water from space. *Science* 301(5639):1491–1494. <https://doi.org/10.1126/science.1089802>
- Alsdorf D, Birkett C, Dunne T, Melack J, Hess L (2001) Water level changes in a large Amazon lake measured with spaceborne radar interferometry and altimetry. *Geophys Res Lett* 28(14):2671–2674. <https://doi.org/10.1029/2001GL012962>
- Alsdorf D, Lettenmaier D, Vörösmarty G (2003) The need for global, satellite-based observations of terrestrial surface waters. *Eos Trans Am Geophys Union* 84(29):269–276. <https://doi.org/10.1029/2003EO290001>
- Alsdorf DE, Rodríguez E, Lettenmaier DP (2007) Measuring surface water from space. *Rev Geophys* 45(2):RG2002. <https://doi.org/10.1029/2006RG000197>
- Alsdorf D, Han S-C, Bates P, Melack J (2010) Seasonal water storage on the amazon floodplain measured from satellites. *Remote Sens Environ* 114(11):2448–2456. <https://doi.org/10.1016/j.rse.2010.05.020>
- Alsdorf D, Beighley E, Laraque A, Lee H, Tshimanga R, O'Loughlin F, Mahé G, Dinga B, Moukandi G, Spencer RGM (2016) Opportunities for hydrologic research in the congo basin. *Rev Geophys* 54(2):378–409. <https://doi.org/10.1002/2016RG000517>
- Andam-Akorful SA, Ferreira VG, Awange JL, Forootan E, He XF (2015) Multi-model and multi-sensor estimations of evapotranspiration over the Volta Basin, West Africa. *Int J Climatol* 35(10):3132–3145. <https://doi.org/10.1002/joc.4198>
- Andam-Akorful S, Ferreira V, Ndehedehe CE, Quaye-Ballard J (2017) An investigation into the freshwater variability in West

- Africa during 1979 – 2010. *Int J Climatol* 37(S1):333–349. <https://doi.org/10.1002/joc.5006>
- Anyadike RNC (1992) Hydrological regions of West Africa: a preliminary survey based on moisture regimes. *Geogr Ann Ser A Phys Geogr* 74(4):375–382. <https://doi.org/10.2307/521433>
- Awange J, Anyah R, Agola N, Forootan E, Omondi P (2013a) Potential impacts of climate and environmental change on the stored water of Lake Victoria Basin and economic implications. *Water Resour Res* 49(0):8160–8173
- Awange J, Forootan E, Kusche J, Kiema J, Omondi P, Heck B, Fleming K, Ohanya S, Gonçalves R (2013b) Understanding the decline of water storage across the Ramsar–Lake Naivasha using satellite-based methods. *Adv Water Resour* 60:7–23. <https://doi.org/10.1016/j.advwatres.2013.07.002>
- Bader J, Latif M (2011) The 1983 drought in the West Sahel: a case study. *Clim Dyn* 36(3–4):463–472. <https://doi.org/10.1007/s00382-009-0700-y>
- Balas N, Nicholson SE, Klotter D (2007) The relationship of rainfall variability in West Central Africa to sea-surface temperature fluctuations. *Int J Climatol* 27(10):1335–1349
- Basu M (2009) West Africa flooding affects 600,000, U.N. reports. CNN. <http://edition.cnn.com/2009/WORLD/africa/09/08/west.africa.flooding/index.html>. Accessed 30 Aug 2017
- Bekoe EO, Logah FY (2013) The impact of droughts and climate change on electricity generation in Ghana. *Environ Sci* 1(1):13–24
- Birkett CM (1995) The contribution of TOPEX/POSEIDON to the global monitoring of climatically sensitive lakes. *J Geophys Res Oceans* 100(C12):25179–25204. <https://doi.org/10.1029/95JC02125>
- Birkett CM (2000) Synergistic remote sensing of Lake Chad: variability of basin inundation. *Remote Sens Environ* 72:218–236. [https://doi.org/10.1016/S0034-4257\(99\)00105-4](https://doi.org/10.1016/S0034-4257(99)00105-4)
- Boening C, Willis JK, Landerer FW, Nerem RS, Fasullo J (2012) The 2011 La Niña: So strong, the oceans fell. *Geophys Res Lett* 39(19):L19602. <https://doi.org/10.1029/2012GL053055>
- Bonaccorso B, Bordi I, Cancelliere A, Rossi G, Sutera A (2003) Spatial variability of drought: an analysis of the SPI in Sicily. *Water Resour Manag* 17(4):273–296. <https://doi.org/10.1023/A:1024716530289>
- Boone A, Decharme B, Guichard F, Rosnay PD, Balsamo G, Belaars A, Chopin F, Orgeval T, Polcher J, Delire C, Ducharne A, Gascoin S, Grippa M, Jarlan L, Kergoat L, Mougin E, Gusev Y, Nasonova O, Harris P, Taylor C, Norgaard A, Sandholt I, Ottle C, Pocard-Leclercq I, Saux-Picart S, Xue Y (2009) The AMMA land surface model intercomparison project (ALMIP). *Bull Am Meteorol Soc* 90(12):1865–1880. <https://doi.org/10.1175/2009BAMS2786.1>
- Brauman KA, Richter BD, Postel S, Malsy M, Flörke M (2016) Water depletion: an improved metric for incorporating seasonal and dry-year water scarcity into water risk assessments. *Elementa*, 4(83). <https://doi.org/10.12952/journal.elementa.000083>
- Brown C, Lall U (2006) Water and economic development: the role of variability and a framework for resilience. *Natl Resour Forum* 30(4):306–317
- Cardoso J-F (1991) Super-symmetric decomposition of the fourth-order cumulant tensor, blind identification of more sources than sensors. <http://perso.telecom-paristech.fr/~cardoso/Papers.PDF/icassp91.pdf>. Accessed 15 Jan 2016
- Cardoso JF (1999) High-order contrasts for independent component analysis. *Neural Comput* 11:157–192
- Cardoso JF, Souloumiac A (1993) Blind beamforming for non-gaussian signals. *IEE Proc* 140(6):362–370
- Carrao H, Russo S, Sepulcre-Canto G, Barbosa P (2016) An empirical standardized soil moisture index for agricultural drought assessment from remotely sensed data. *Int J Appl Earth Obs Geoinf* 48:74–84. <https://doi.org/10.1016/j.jag.2015.06.011>
- Castellazzi P, Martel R, Galloway DL, Longuevergne L, Rivera A (2016) Assessing groundwater depletion and dynamics using GRACE and InSAR: potential and limitations. *Groundwater* 54(6):768–780. <https://doi.org/10.1111/gwat.12453>
- Chen JL, Rodell M, Wilson CR, Famiglietti JS (2005) Low degree spherical harmonic influences on gravity recovery and climate experiment (GRACE) water storage estimates. *Geophys Res Lett* 32(14):L14405. <https://doi.org/10.1029/2005GL022964>
- Chen JL, Wilson CR, Tapley BD, Longuevergne L, Yang ZL, Scanlon BR (2010) Recent La Plata basin drought conditions observed by satellite gravimetry. *J Geophys Res Atmos* 115(D22):D22108. <https://doi.org/10.1029/2010JD014689>
- Chen T, de Jeu R, Liu Y, van der Werf G, Dolman A (2014) Using satellite based soil moisture to quantify the water driven variability in NDVI: a case study over mainland Australia. *Remote Sens Environ* 140:330–338. <https://doi.org/10.1016/j.rse.2013.08.022>
- Coe MT, Foley JA (2001) Human and natural impacts on the water resources of the Lake Chad basin. *J Geophys Res* 106(D4):3349–3356. <https://doi.org/10.1029/2000JD900587>
- Conway D, Persechino A, Ardoin-Bardin S, Hamandawana H, Dieulin C, Mahé G (2009) Rainfall and water resources variability in Sub-Saharan Africa during the twentieth century. *J Hydrometeorol* 10(1):41–59. <https://doi.org/10.1175/2008JHM1004.1>
- Cook KH, Vizy EK (2006) Coupled model simulations of the West African monsoon system: twentieth- and twenty-first-century simulations. *J Clim* 19(15):3681–3703. <https://doi.org/10.1175/JCLI3814.1>
- Corzo Perez GA, van Huijgevoort MHJ, Voß F, van Lanen HAJ (2011) On the spatio-temporal analysis of hydrological droughts from global hydrological models. *Hydrol Earth Syst Sci* 15(9):2963–2978. <https://doi.org/10.5194/hess-15-2963-2011>
- Crowley JW, Mitrovica JX, Bailey RC, Tamisiea ME, Davis JL (2006) Land water storage within the Congo Basin inferred from GRACE satellite gravity data. *Geophys Res Lett* 33(19):L19402. <https://doi.org/10.1029/2006GL027070>
- Dai A, Qian T, Trenberth KE, Milliman JD (2009) Changes in continental freshwater discharge from 1948 to 2004. *J Clim* 22(10):2773–2792. <https://doi.org/10.1175/2008JCLI2592.1>
- Dardel C, Kergoat L, Hiernaux P, Mougin E, Grippa M, Tucker C (2014) Re-greening Sahel: 30 years of remote sensing data and field observations (Mali, Niger). *Remote Sens Environ* 140:350–364. <https://doi.org/10.1016/j.rse.2013.09.011>
- de Giesen NV, Rodgers C, Vlek P (2007) The GLOWA Volta Project: Interdisciplinary analysis of the impact of global change on a river basin in West Africa. *Int J River Basin Manag* 5(1):3–8. <https://doi.org/10.1080/15715124.2007.9635300>
- Descroix L, Mahé G, Lebel T, Favreau G, Galle S, Gautier E, Olivry J-C, Albergel J, Amogu O, Cappelaere B, Dessouassi R, Diedhiou A, Breton EL, Mamadou I, Sighomnou D (2009) Spatio-temporal variability of hydrological regimes around the boundaries between Sahelian and Sudanian areas of West Africa: a synthesis. *J Hydrol* 375(1–2):90–102. <https://doi.org/10.1016/j.jhydrol.2008.12.012>
- Dirmeyer PA, Guo Z, Gao X (2004) Comparison, validation, and transferability of eight multiyear global soil wetness products. *J Hydrometeorol* 5:1011–1033. <https://doi.org/10.1175/JHM-388.1>
- Do FC, Goudiaby VA, Gimenez O, Diagne AL, Diouf M, Rocheteau A, Akpo LE (2005) Environmental influence on canopy phenology in the dry tropics. *For Ecol Manag* 215(1–3):319–328. <https://doi.org/10.1016/j.foreco.2005.05.022>
- Döll P, Müller Schmied H, Schuh C, Portmann FT, Eicker A (2014) Global-scale assessment of groundwater depletion and related groundwater abstractions: combining hydrological modeling

- with information from well observations and GRACE satellites. *Water Resour Res* 50(7):5698–5720
- Du L, Tian Q, Yu T, Meng Q, Jancso T, Udvardy P, Huang Y (2013) A comprehensive drought monitoring method integrating MODIS and TRMM data. *Int J Appl Earth Obs Geoinf* 23:245–253. <https://doi.org/10.1016/j.jag.2012.09.010>
- Erfanian A, Wang G, Yu M, Anyah R (2016) Multimodel ensemble simulations of present and future climates over West Africa: impacts of vegetation dynamics. *J Adv Model Earth Syst*. <https://doi.org/10.1002/2016MS000660>
- Falkenmark M, Lundqvist J (1998) Towards water security: political determination and human adaptation crucial. *Natl Resour Forum* 22(1):37–51. <https://doi.org/10.1111/j.1477-8947.1998.tb00708.x>
- Falloon P, Betts R (2010) Climate impacts on European agriculture and water management in the context of adaptation and mitigation—the importance of an integrated approach. *Sci Total Environ* 408(23):5667–5687. <https://doi.org/10.1016/j.scitotenv.2009.05.002>
- Famiglietti JS, Rodell M (2013) Water in the balance. *Science* 340(6138):1300–1301. <https://doi.org/10.1126/science.1236460>
- Famiglietti JS, Cazenave A, Eicker A, Reager JT, Rodell M, Velicogna I (2015) Satellites provide the big picture. *Science* 349(6249):684–685. <https://doi.org/10.1126/science.aac9238>
- Fan Y, Dool HV (2004) Climate prediction center global monthly soil moisture data set at 0.5° resolution for 1948 to present. *J Geophys Res* 109:D10102. <https://doi.org/10.1029/2003JD004345>
- Farnsworth A, White E, Williams CJ, Black E, Kniveton DR (2011) Understanding the large scale driving mechanisms of rainfall variability over Central Africa. In: African climate and climate change: physical, social and political perspectives, Springer, Netherlands, pp 101–122. https://doi.org/10.1007/978-90-481-3842-5_5
- Favreau G, Cappelaere B, Massuel S, Leblanc M, Boucher M, Boulain N, Leduc C (2009) Land clearing, climate variability, and water resources increase in semiarid southwest Niger: a review. *Water Resour Res* 45:W00A16. <https://doi.org/10.1029/2007WR006785>
- Ferreira V, Asiah Z (2015) An investigation on the closure of the water budget methods over Volta Basin using multi-satellite data. *Int Assoc Geodesy Symp*. <https://doi.org/10.1007/1345-2015-137>
- Ferreira V, Montecino H, Ndehedehe C, Heck B, Gong Z, Westershaus M, de Freitas S (2018) Space-based observations of crustal deflections for drought characterization in Brazil. *Sci Total Environ* 644:256–273. <https://doi.org/10.1016/j.scitotenv.2018.06.277>
- Fisher-Jeffes L, Carden K, Armitage NP, Winter K (2017) Stormwater harvesting: Improving water security in South Africa's urban areas. *S Afr J Sci* 113(1/2):4. <https://doi.org/10.17159/sajs.2017/20160153>
- Fontaine B, Bigot S (1993) West African rainfall deficits and sea surface temperatures. *Int J Climatol* 13(3):271–285. <https://doi.org/10.1002/joc.3370130304>
- Forootan E, Kusche J, Loth I, Schuh WD, Eicker A, Awange J, Longuevergne L, Diekkruiger B, Shum MSCK (2014) Multivariate prediction of total water storage changes over West Africa from multi-satellite data. *Surv Geophys* 35(4):913–940. <https://doi.org/10.1007/s10712-014-9292-0>
- Frappart F, Calmant S, Cauhopé M, Seyler F, Cazenave A (2006) Preliminary results of ENVISAT RA-2-derived water levels validation over the Amazon basin. *Remote Sens Environ* 100(2):252–264. <https://doi.org/10.1016/j.rse.2005.10.027>
- Freitas A (2013) Water as a stress factor in sub-Saharan Africa. In: European union institute for security studies, pp 1–4. http://www.iss.europa.eu/uploads/media/Brief_12.pdf. Accessed 12 July 2017
- Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L, Hoell A, Michaelsen J (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci Data* 2:150066. <https://doi.org/10.1038/sdata.2015.66>
- Gal L, Grippa M, Hiernaux P, Pons L, Kergoat L (2017) The paradoxical evolution of runoff in the pastoral Sahel: analysis of the hydrological changes over the Agoufou watershed (Mali) using the KINEROS-2 model. *Hydrol Earth Syst Sci* 21(9):4591–4613. <https://doi.org/10.5194/hess-21-4591-2017>
- Giannini A, Saravanan R, Chang P (2003) Oceanic forcing of Sahel rainfall on interannual to decadal time scales. *Science* 302(5647):1027–1030. <https://doi.org/10.1126/science.1089357>
- Gizaw MS, Gan TY (2017) Impact of climate change and El Niño episodes on droughts in sub-Saharan Africa. *Clim Dyn* 49(1):665–682. <https://doi.org/10.1007/s00382-016-3366-2>
- Gleick PH (1993) Water and conflict: fresh water resources and international security. *Int Secur* 18(1):79–112. <https://doi.org/10.2307/2539033>
- Gleick PH (2000) The world's water: Biennial report on freshwater resources 2000–2001. Island Press, Washington, DC
- Goncàlves J, Peterson J, Deschamps P, Hamelin B, Baba-Sy O (2013) Quantifying the modern recharge of the “fossil” Sahara aquifers. *Geophys Res Lett* 40:2673–2678. <https://doi.org/10.1002/grl.50478>
- Grippa M, Kergoat L, Frappart F, Araud Q, Boone A, de Rosnay P, Lemoine JM, Gascoin S, Balsamo G, Otlé C, Decharme B, Saux-Picart S, Ramillien G (2011) Land water storage variability over West Africa estimated by gravity recovery and climate experiment (GRACE) and land surface models. *Water Resour Res* 47(5):W05549. <https://doi.org/10.1029/2009wr008856>
- Guan K, Wood EF, Medvigy D, Kimball J, Pan M, Caylor KK, Sheffield J, Xu X, Jones MO (2014) Terrestrial hydrological controls on land surface phenology of African savannas and woodlands. *J Geophys Res Biogeosci* 119(8):1652–1669. <https://doi.org/10.1002/2013JG002572>
- Guan K, Sultan B, Biasutti M, Baron C, Lobell DB (2015) What aspects of future rainfall changes matter for crop yields in West Africa? *Geophys Res Lett* 42(19):8001–8010. <https://doi.org/10.1002/2015GL063877>
- Hall JW, Grey D, Garrick D, Fung F, Brown C, Dadson SJ, Sadoff CW (2014) Coping with the curse of freshwater variability. *Science* 346(6208):429–430. <https://doi.org/10.1126/science.1257890>
- Hayes MJ, Svoboda MD, Wilhite DA, Vanyarkho OV (1999) Monitoring the 1996 drought using the standardized precipitation index. *Bull Am Meteorol Soc* 80:429–438
- Heim RR (2002) A review of twentieth-century drought indices used in the United States. *Bull Am Meteorol Soc* 83(8):1149–1165
- Hély C, Bremond L, Alleaume S, Smith B, Sykes MT, Guiot J (2006) Sensitivity of African biomes to changes in the precipitation regime. *Glob Ecol Biogeogr* 15(3):258–270. <https://doi.org/10.1111/j.1466-8238.2006.00235.x>
- Henry C, Allen DM, Huang J (2011) Groundwater storage variability and annual recharge using well-hydrograph and GRACE satellite data. *Hydrogeol J* 19:741–755. <https://doi.org/10.1007/s10040-011-0724-3>
- Herrmann SM, Anyamba A, Tucker CJ (2005) Recent trends in vegetation dynamics in the African Sahel and their relationship to climate. *Glob Environ Change* 15(4):394–404. <https://doi.org/10.1016/j.gloenvcha.2005.08.004>
- Hinderer J, de Linage C, Boya J-P, Gegout P, Masson F, Rogister Y, Amalvict M, Pfeffer J, Littel F, Luck B, Bayerb R, Champollion C, Collard P, Moigne NL, Diamante M, Deroussi S, de Viron O, Biancale R, Lemoine J-M, Bonvalot S, Gabalda G, Bock O, Genthon P, Boucher M, Favreau G, Séguis L, Delclaux F, Cappelaere B, Oi M, Descloitresh M, Galleh S, Laurent J-P, Legchenko

- A, Bouinck M-N (2009) The GHYRAF (Gravity and Hydrology in Africa) experiment: description and first results. *J Geodyn* 48:172–181. <https://doi.org/10.1016/j.jog.2009.09.014>
- Hoff H (2009) Global water resources and their management. *Curr Opin Environ Sustain* 1(2):141–147. <https://doi.org/10.1016/j.cosust.2009.10.001>
- Houborg R, Rodell M, Li B, Reichle R, Zaitchik BF (2012) Drought indicators based on model-assimilated gravity recovery and climate experiment (GRACE) terrestrial water storage observations. *Water Resour Res* 48(7):W07525. <https://doi.org/10.1029/2011WR011291>
- Hua W, Zhou L, Chen H, Nicholson SE, Raghavendra A, Jiang Y (2016) Possible causes of the central equatorial African long-term drought. *Environ Res Lett* 11(12):124002. <https://doi.org/10.1088/1748-9326/11/12/124002>
- Huber S, Fensholt R, Rasmussen K (2011) Water availability as the driver of vegetation dynamics in the African Sahel from 1982 to 2007. *Glob Planet Change* 76(3–4):186–195. <https://doi.org/10.1016/j.gloplacha.2011.01.006>
- Humphrey V, Gudmundsson L, Seneviratne SI (2016) Assessing global water storage variability from GRACE: trends, seasonal cycle, subseasonal anomalies and extremes. *Surv Geophys* 37(2):357–395. <https://doi.org/10.1007/s10712-016-9367-1>
- Ivits E, Horion S, Fensholt R, Cherlet M (2014) Drought footprint on European ecosystems between 1999 and 2010 assessed by remotely sensed vegetation phenology and productivity. *Glob Change Biol* 20(2):581–593. <https://doi.org/10.1111/gcb.12393>
- Janicot S (1992) Spatiotemporal variability of West African rainfall. Part i: regionalizations and typings. *J Clim* 5(5):489–497
- Jenkins GS, Adamou G, Fongang S (2002) The challenges of modeling climate variability and change in West Africa. *Clim Change* 52(3):263–286. <https://doi.org/10.1023/A:1013741803144>
- Kasei R, Diekkrüger B, Leemhuis C (2010) Drought frequency in the Volta Basin of West Africa. *Sustain Sci* 5(1):89–97
- Knauer K, Gessner U, Dech S, Kuenzer C (2014) Remote sensing of vegetation dynamics in West Africa. *Int J Remote Sens* 35(17):6357–6396. <https://doi.org/10.1080/0143161.2014.954062>
- Koster RD, Dirmeyer PA, Guo Z, Bonan G, Chan E, Cox P, Gordon CT, Kanae S, Kowalczyk E, Lawrence D, Liu P, Lu C-H, Malyshev S, McAvaney B, Mitchell K, Mocko D, Oki T, Oleson K, Pitman A, Sud YC, Taylor CM, Verseghy D, Vasic R, Xue Y, Yamada T (2004) Regions of strong coupling between soil moisture and precipitation. *Science* 305(5687):1138–1140. <https://doi.org/10.1126/science.1100217>
- Kumar KN, Rajeevan M, Pai D, Srivastava A, Preethi B (2013) On the observed variability of monsoon droughts over India. *Weather Clim Extremes* 1:42–50. <https://doi.org/10.1016/j.wace.2013.07.006>
- Kurnik B, Kajfež-Bogataj L, Horion S (2015) An assessment of actual evapotranspiration and soil water deficit in agricultural regions in Europe. *Int J Climatol* 35(9):2451–2471. <https://doi.org/10.1002/joc.4154>
- Kuylenstierna JL, Björklund G, Najlis P (1997) Sustainable water future with global implications: everyone's responsibility. *Natl Resour Forum* 21(3):181–190. <https://doi.org/10.1111/j.1477-8947.1997.tb00691.x>
- Lammers RB, Shiklomanov AI, Vörösmarty CJ, Fekete BM, Peterson BJ (2001) Assessment of contemporary Arctic river runoff based on observational discharge records. *J Geophys Res Atmos* 106(D4):3321–3334. <https://doi.org/10.1029/2000JD900444>
- Laux P (2009) Statistical modeling of precipitation for agricultural planning in the Volta Basin of West Africa. Doctoral dissertation, Mitteilungen/Institut für Wasserbau, Universität Stuttgart, 179(198). <https://elib.uni-stuttgart.de/opus/volltexte/2009/4016/pdf/Dissertation-Laux.pdf>. Accessed 25 July 2014
- Lebel T, Ali A (2009) Recent trends in the Central and Western Sahel rainfall regime (1990–2007). *J Hydrol* 375(1–2):52–64. <https://doi.org/10.1016/j.jhydrol.2008.11.030>
- Lebel T, Delclaux F, Le Barbé L, Polcher J (2000) From GCM scales to hydrological scales: rainfall variability in West Africa. *Stoch Environ Res Risk Assess* 14(4):275–295. <https://doi.org/10.1007/s004770000050>
- Lebel T, Cappelaere B, Galle S, Hanan N, Kergoat L, Levis S, Vieux B, Descroix L, Gosset M, Mougou E, Peugeot C, Seguis L (2009) AMMA-CATCH studies in the Sahelian region of West-Africa: an overview. *J Hydrol* 375(1–2):3–13
- Leblanc M, Leduc C, Razack M, Lemoalle J, Dagorne D, Mofor L (2003) Applications of remote sensing and GIS for ground-water modelling of large semiarid areas: example of the Lake Chad Basin, Africa. In: Proceedings of the international symposium held at Montpellier, hydrology of the mediterranean and semiarid regions, vol 278, pp 186–191, April 2003
- Leduc C, Bromley J, Schroeter P (1997) Water table fluctuation and recharge in semi-arid climate: some results of the HAPEX-Sahel hydrodynamic survey (Niger). *J Hydrol* 188(188–189):123–138. [https://doi.org/10.1016/S0022-1694\(96\)03156-3](https://doi.org/10.1016/S0022-1694(96)03156-3)
- Leduc C, Favreau G, Schroeter P (2001) Long-term rise in a Sahelian water-table: the continental terminal in south-west Niger. *J Hydrol* 243(1–2):43–54. [https://doi.org/10.1016/S0022-1694\(00\)00403-0](https://doi.org/10.1016/S0022-1694(00)00403-0)
- Lee H, Beighley RE, Alsdorf D, Jung HC, Shum C, Duan J, Guo J, Yamazaki D, Andreadis K (2011) Characterization of terrestrial water dynamics in the Congo Basin using GRACE and satellite radar altimetry. *Remote Sens Environ* 115(12):3530–3538. <https://doi.org/10.1016/j.rse.2011.08.015>
- Lee H, Jung HC, Yuan T, Beighley RE, Duan J (2014) Controls of terrestrial water storage changes over the Central Congo Basin determined by integrating Palsar ScanSar, Envisat Altimetry, and Grace data. *Remote Sens Terr Water Cycle Geophys Monogr* 206:117–129. <https://doi.org/10.1002/9781118872086.ch7/pdf>
- Lemoalle J, Bader J-C, Leblanc M, Sedick A (2012) Recent changes in Lake Chad: observations, simulations and management options (1973–2011). *Glob Planet Change* 80–81:247–254. <https://doi.org/10.1016/j.gloplacha.2011.07.004>
- Lettenmaier DP (2005) Observations of the global water cycle—global monitoring networks. In: Anderson MG (ed) Encyclopedia of hydrological sciences, vol 5. Wiley, New York, pp 2719–2732
- Li B, Rodell M (2015) Evaluation of a model-based groundwater drought indicator in the conterminous U.S. *J Hydrol* 526:78–88. <https://doi.org/10.1016/j.jhydrol.2014.09.027>
- Li KY, Coe MT, Ramankutty N (2005) Investigation of hydrological variability in West Africa using land surface models. *J Clim* 18(16):3173–3188. <https://doi.org/10.1175/JCLI3452.1>
- Li K, Coe M, Ramankutty N, Jong RD (2007) Modeling the hydrological impact of land-use change in West Africa. *J Hydrol* 337(3–4):258–268. <https://doi.org/10.1016/j.jhydrol.2007.01.038>
- Long D, Scanlon BR, Longuevergne L, Sun AY, Fernando DN, Save H (2013) GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas. *Geophys Res Lett* 40(13):3395–3401. <https://doi.org/10.1002/Grl.50655>
- Long D, Shen Y, Sun A, Hong Y, Longuevergne L, Yang Y, Li B, Chen L (2014) Drought and flood monitoring for a large karst plateau in southwest china using extended GRACE data. *Remote Sens Environ* 155:145–160. <https://doi.org/10.1016/j.rse.2014.08.006>
- Lopez T, Antoine R, Kerr Y, Darrozes J, Rabinowicz M, Ramillien G, Cazenave A, Genthon P (2016) Subsurface hydrology of the Lake Chad Basin from convection modelling and observations. *Surv Geophys* 37(2):471–502. <https://doi.org/10.1007/s10712-016-9363-5>

- Lovett JC, Midgley GF, Barnard P (2005) Climate change and ecology in africa. *Afr J Ecol* 43(3):167–169. <https://doi.org/10.1111/j.1365-2028.2005.00598.x>
- Mahé G, Olivry J-C (1999) Assessment of freshwater yields to the ocean along the intertropical Atlantic coast of Africa (1951–1989). *Comptes Rendus de l'Académie des Sciences–Ser IIA–Earth Planet Sci* 328(9):621–626. [https://doi.org/10.1016/S1251-8050\(99\)80159-1](https://doi.org/10.1016/S1251-8050(99)80159-1)
- Mahé G, Paturel J-E (2009) 1896–2006 Sahelian annual rainfall variability and runoff increase of Sahelian Rivers. *Comptes Rendus Geosci* 341(7):538–546. <https://doi.org/10.1016/j.crte.2009.05.002>
- Mahe G, Lienou G, Descroix L, Bamba F, Paturel JE, Laraque A, Meddi M, Habaieb H, Adeaga O, Dieulin C, Chahnez Kotti F, Khomsi K (2013) The rivers of Africa: witness of climate change and human impact on the environment. *Hydrol Process* 27(15):2105–2114
- McCabe MF, Rodell M, Alsdorf DE, Miralles DG, Uijlenhoet R, Wagner W, Lucieer A, Houborg R, Verhoest NEC, Franz TE, Shi J, Gao H, Wood EF (2017) The future of Earth observation in hydrology. *Hydrol Earth Syst Sci* 21(7):3879–3914. <https://doi.org/10.5194/hess-21-3879-2017>
- McKee TB, Doeskin NJ, Kieist J (1993) The relationship of drought frequency and duration to time scales. In: Conference on applied climatology, American meteorological society, Boston, MA, pp 179–184. www.ccc.atmos.colostate.edu/relationshipofdroughtfrequency.pdf. Accessed 27 June 2014
- Mehr AD, Kahya E, Ozger M (2014) A gene-wavelet model for long lead time drought forecasting. *J Hydrol* 517:691–699. <https://doi.org/10.1016/j.jhydrol.2014.06.012>
- Mishra AK, Ines AV, Das NN, Khedun CP, Singh VP, Sivakumar B, Hansen JW (2015) Anatomy of a local-scale drought: Application of assimilated remote sensing products, crop model, and statistical methods to an agricultural drought study. *J Hydrol* 526:15–29. <https://doi.org/10.1016/j.jhydrol.2014.10.038>
- Mohino E, Janicot S, Bader J (2011) Sahel rainfall and decadal to multi-decadal sea surface temperature variability. *Clim Dyn* 37:419–440. <https://doi.org/10.1007/s00382-010-0867-2>
- Moore P, Williams SDP (2014) Integration of altimetry lake levels and GRACE gravimetry over Africa: Inferences for terrestrial water storage change 2003–2011. *Water Resour Res* 50:9696–9720. <https://doi.org/10.1002/2014WR015506>
- Nahmani S, Bock O, Bouin M-N, Santamaría-Gómez A, Boy J-P, Collilieux X, Métivier L, Isabelle Panet PG, de Linage C, Wöppelmann G (2012) Hydrological deformation induced by the West African Monsoon: comparison of GPS, GRACE and loading models. *J Geophys Res Solid Earth*. <https://doi.org/10.1029/2011JB009102>
- NASA (2002). Grace launch. National Aeronautics And Space Administration. <http://www.jpl.nasa.gov/news/press-kits/gracelaunch.pdf> Accessed 9 Feb 2017
- Ndehedehe CE (2017) Remote sensing of West Africa's water resources using multi-satellites and models. PhD thesis, Curtin University, Bentley, Perth, Western Australia. <http://hdl.handle.net/20.500.11937/59637>. Accessed 12 Jan 2018
- Ndehedehe C, Awange J, Agutu N, Kuhn M, Heck B (2016a) Understanding changes in terrestrial water storage over West Africa between 2002 and 2014. *Adv Water Resour* 88:211–230. <https://doi.org/10.1016/j.advwatres.2015.12.009>
- Ndehedehe CE, Agutu NO, Okwuashi OH, Ferreira VG (2016b) Spatio-temporal variability of droughts and terrestrial water storage over Lake Chad Basin using independent component analysis. *J Hydrol* 540:106–128. <https://doi.org/10.1016/j.jhydrol.2016.05.068>
- Ndehedehe CE, Awange JL, Corner R, Kuhn M, Okwuashi O (2016c) On the potentials of multiple climate variables in assessing the spatio-temporal characteristics of hydrological droughts over the Volta Basin. *Sci Total Environ* 557–558:819–837. <https://doi.org/10.1016/j.scitotenv.2016.03.004>
- Ndehedehe CE, Awange J, Kuhn M, Agutu N, Fukuda Y (2017a) Analysis of hydrological variability over the Volta river basin using in-situ data and satellite observations. *J Hydrol Reg Stud* 12:88–110. <https://doi.org/10.1016/j.ejrh.2017.04.005>
- Ndehedehe CE, Awange J, Kuhn M, Agutu N, Fukuda Y (2017b) Climate teleconnections influence on West Africa's terrestrial water storage. *Hydrol Process* 31(18):3206–3224. <https://doi.org/10.1002/hyp.11237>
- Ndehedehe CE, Agutu NO, Okwuashi O (2018a) Is terrestrial water storage a useful indicator in assessing the impacts of climate variability on crop yield in semi-arid ecosystems? *Ecol Indic* 88C:51–62. <https://doi.org/10.1016/j.ecolind.2018.01.026>
- Ndehedehe CE, Awange JL, Agutu NO, Okwuashi O (2018b) Changes in hydro-meteorological conditions over tropical West Africa (1980 – 2015) and links to global climate. *Glob Planet Change* 162:321–341. <https://doi.org/10.1016/j.gloplacha.2018.01.020>
- Ndehedehe CE, Okwuashi O, Ferreira VG, Agutu NO (2018c) Exploring evapotranspiration dynamics over sub-Sahara Africa (2000 – 2014). *Environ Monit Assess* 190(7):400. <https://doi.org/10.1007/s10661-018-6780-6>
- Ndehedehe CE, Anyah RO, Alsdorf D, Agutu NO, Ferreira VG (2019) Modelling the impacts of global multi-scale climatic drivers on hydro-climatic extremes (1901–2014) over the congo basin. *Sci Total Environ* 651:1569–1587. <https://doi.org/10.1016/j.scitotenv.2018.09.203>
- Ngom F, Tweed S, Bader J-C, Saos J-L, Malou R, Leduc C, Leblanc M (2016) Rapid evolution of water resources in the Senegal delta. *Glob Planet Change* 144:34–47. <https://doi.org/10.1016/j.gloplacha.2016.07.002>
- Ni S, Chen J, Wilson CR, Li J, Hu X, Fu R (2017) Global terrestrial water storage changes and connections to ENSO events. *Surv Geophys*. <https://doi.org/10.1007/s10712-017-9421-7>
- Nicholson S (2005) On the question of the “recovery” of the rains in the West African Sahel. *J Arid Environ* 63(3):615–641. <https://doi.org/10.1016/j.jaridenv.2005.03.004>
- Nicholson S (2013) The West African Sahel: a review of recent studies on the rainfall regime and its interannual variability. *ISRN Meteorol* 2013(453521):1–32. <https://doi.org/10.1155/2013/453521>
- Nicholson SE (2014) Spatial teleconnections in African rainfall: a comparison of 19th and 20th century patterns. *Holocene* 24(12):1840–1848. <https://doi.org/10.1177/0959683614551230>
- Nicholson SE, Davenport ML, Malo AR (1990) A comparison of the vegetation response to rainfall in the Sahel and East Africa, using normalized difference vegetation index from NOAA AVHRR. *Clim Change* 17(2):209–241. <https://doi.org/10.1007/BF00138369>
- Nicholson SE, Some B, Kone B (2000) An analysis of recent rainfall conditions in West Africa, including the rainy seasons of the 1997 El Niño and the 1998 La Niña years. *J Clim* 13(14):2628–2640
- Nicholson SE, Some B, Mccollum J, Nelkin E, Klotter D, Berte Y, Diallo BM, Gaye I, Kpabebe G, Ndiaye O, Noukpozoukou JN, Tanu MM, Thiam A, Toure AA, Traore AK (2003) Validation of TRMM and other rainfall estimates with a high-density gauge datasets for West Africa. Part I: validation of GPCP rainfall product and pre-TRMM satellite and blended products. *J Appl Meteorol* 42:1337–1354. [https://doi.org/10.1175/1520-0450\(2003\)042C1337:VOTAOR3E2.0.CO;2](https://doi.org/10.1175/1520-0450(2003)042C1337:VOTAOR3E2.0.CO;2)
- Niu J, Chen J, Sun L (2015) Exploration of drought evolution using numerical simulations over the Xijiang (West River) basin in South China. *J Hydrol* 526:68–77. <https://doi.org/10.1016/j.jhydrol.2014.11.029>
- Nka BN, Oudin L, Karambiri H, Paturel JE, Ribstein P (2015) Trends in floods in West Africa: analysis based on 11 catchments in

- the region. *Hydrol Earth Syst Sci* 19(11):4707–4719. <https://doi.org/10.5194/hess-19-4707-2015>
- Oettli P, Sultan B, Baron C, Vrac M (2011) Are regional climate models relevant for crop yield prediction in West Africa? *Environ Res Lett* 6(1):014008. <https://doi.org/10.1088/1748-9326/6/1/014008>
- Okonkwo C (2014) An advanced review of the relationships between Sahel precipitation and climate indices: a wavelet approach. *Int J Atmos Sci* 2014:11. <https://doi.org/10.1155/2014/759067>
- Oloruntade AJ, Mohammad TA, Ghazali AH, Wayayok A (2017) Analysis of meteorological and hydrological droughts in the Niger-South Basin, Nigeria. *Glob Planet Change* 155(Supplement C):225–233. <https://doi.org/10.1016/j.gloplacha.2017.05.002>
- Otto FEL, Jones RG, Halladay K, Allen MR (2013) Attribution of changes in precipitation patterns in African rainforests. *Philos Trans R Soc Lond B Biol Sci* 368(1625):20120299. <https://doi.org/10.1098/rstb.2012.0299>
- Owusu K, Waylen P (2009) Trends in spatio-temporal variability in annual rainfall in Ghana (1951–2000). *Weather* 64(5):115–119
- Paeth H, Fink A, Pohle S, Keis F, Machel H, Samimi C (2012) Meteorological characteristics and potential causes of the 2007 flood in sub-Saharan Africa. *Int J Climatol* 31:1908–1926. <https://doi.org/10.1002/Joc.2199>
- Panthou G, Vischel T, Lebel T (2014) Recent trends in the regime of extreme rainfall in the central sahel. *Int J Climatol* 34(15):3998–4006. <https://doi.org/10.1002/joc.3984>
- Paolino DA, Kinter JL III, Kirtman BP, Min D, Straus DM (2012) The impact of land surface and atmospheric initialization on seasonal forecasts with CCSM. *J Clim* 25(3):1007–1021. <https://doi.org/10.1175/2011JCLI3934.1>
- Paturol J-E, Ouedraogo M, Servat E, Mahe G, Dezetter A, Boyer J-F (2003) The concept of rainfall and streamflow normals in West and Central Africa in a context of climatic variability. *Hydrol Sci J* 48(1):125–137. <https://doi.org/10.1623/hysj.48.1.125.43479>
- Paturol JE, Mahé G, Diello P, Barbier B, Dezetter A, Dieulin C, Karambiri H, Yacouba H, Maiga A (2017) Using land cover changes and demographic data to improve hydrological modeling in the Sahel. *Hydrol Process* 31(4):811–824. <https://doi.org/10.1002/hyp.11057>
- Pellarin T, Tran T, Cohard J-M, Galle S, Laurent J-P, de Rosnay P, Vischel T (2009) Soil moisture mapping over West Africa with a 30-min temporal resolution using AMSR-E observations and a satellite-based rainfall product. *Hydrol Earth Syst Sci* 13(10):1887–1896. <https://doi.org/10.5194/hess-13-1887-2009>
- Petropoulos GP, Ireland G, Barrett B (2015) Surface soil moisture retrievals from remote sensing: current status, products, and future trends. *Phys Chem Earth, Parts A/B/C* 83–84:36–56. <https://doi.org/10.1016/j.pce.2015.02.009>
- Phillips T, Nerem RS, Fox-Kemper B, Famiglietti JS, Rajagopalan B (2012) The influence of ENSO on global terrestrial water storage using GRACE. *Geophys Res Lett* 39(L16705):2012. <https://doi.org/10.1029/2012GL052495>
- Piao S, Ciais P, Huang Y, Shen Z, Peng S, Li J, Zhou L, Liu H, Ma Y, Friedlingstein P, Liu C, Tan K, Yu Y, Zhang T, Fang J (2010) The impacts of climate change on water resources and agriculture in China. *Nature* 467:43–51. <https://doi.org/10.1038/nature09364>
- Pierre C, Grippa M, Mougou E, Guichard F, Kergoat L (2016) Changes in Sahelian annual vegetation growth and phenology since 1960: a modeling approach. *Glob Planet Change* 143:162–174. <https://doi.org/10.1016/j.gloplacha.2016.06.009>
- Preisendorfer R (1988) Principal component analysis in meteorology and oceanography. Developments in atmospheric science, vol 17. Elsevier, Amsterdam
- Prudhomme C, Giuntoli I, Robinson EL, Clark DB, Arnell NW, Dankers R, Fekete BM, Franssen W, Gerten D, Gosling SN, Hagemann S, Hannah DM, Kim H, Masaki Y, Satoh Y, Stacke T, Wada Y, Wisser D (2014) Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proc Natl Acad Sci* 111(9):3262–3267. <https://doi.org/10.1073/pnas.1222473110>
- Ramillien G, Bouhours S, Lombard A, Cazenave A, Flechtner FR, Schmidt (2008) Land water storage contribution to sea level from GRACE geoid data over 2003–2006. *Glob Planet Change* 60(3–4):381–392. <https://doi.org/10.1016/j.gloplacha.2007.04.002>
- Reager JT, Thomas BF, Famiglietti JS (2014) River basin flood potential inferred using GRACE gravity observations at several months lead time. *Nat Geosci* 7(8):588–592. <https://doi.org/10.1038/ngeo2203>
- Reichle RH, Koster RD, Lannoy GJMD, Forman BA, Liu Q, Mahanama SPP, Touré A (2011) Assessment and enhancement of MERRA land surface hydrology estimates. *J Clim* 24(24):6322–6338. <https://doi.org/10.1175/JCLI-D-10-05033.1>
- Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E, Bosilovich MG, Schubert SD, Takacs L, Kim G-K, Bloom S, Chen J, Collins D, Conaty A, da Silva A, Gu W, Joiner J, Koster RD, Lucchesi R, Molod A, Owens T, Pawson S, Pegion P, Redder CR, Reichle R, Robertson FR, Ruddick AG, Sienkiewicz M, Woollen J (2011) MERRA: NASA’s modern-era retrospective analysis for research and applications. *J Clim* 24(14):3624–3648. <https://doi.org/10.1175/JCLI-D-11-00015.1>
- Rijsberman FR (2006) Water scarcity: fact or fiction? *Agric Water Manag* 80(1–3):5–22. <https://doi.org/10.1016/j.agwat.2005.07.001>
- Rodell M, Houser PR, Jambor U, Gottschalk J, Mitchell K, Meng K, Arsenault CJ, Cosgrove B, Radakovich J, Bosilovich M, Entin JK, Walker JP, Lohmann D, Toll D (2004) The global land data assimilation system. *Bull Am Meteorol Soc* 85(3):381–394. <https://doi.org/10.1175/BAMS-85-3-381.R>
- Rouault M, Richard Y (2003) Intensity and spatial extension of drought in South Africa at different time scales. *Water SA* 29(4):489–500
- Roudier P, Sultan B, Quirion P, Berg A (2011) The impact of future climate change on West African crop yields: what does the recent literature say? *Glob Environ Change* 21:1073–1083. <https://doi.org/10.1016/j.gloenvcha.2011.04.007>
- Roudier P, Ducharne A, Feyen L (2014) Climate change impacts on runoff in west africa: a review. *Hydrol Earth Syst Sci* 18(7):2789–2801. <https://doi.org/10.5194/hess-18-2789-2014>
- Sanogo S, Fink AH, Omosho JA, Ba A, Redl R, Ermert V (2015) Spatio-temporal characteristics of the recent rainfall recovery in West Africa. *Int J Climatol* 35(15):4589–4605. <https://doi.org/10.1002/joc.4309>
- Santos JF, Pulido-Calvo I, Portela MM (2010) Spatial and temporal variability of droughts in Portugal. *Water Resour Res* 46(3):W03503. <https://doi.org/10.1029/2009WR008071>
- Scanlon BR, Reedy RC, Stonestrom DA, Prudic DE, Dennehy KF (2005) Impact of land use and land cover change on groundwater recharge and quality in the southwestern US. *Glob Change Biol* 11(10):1577–1593. <https://doi.org/10.1111/j.1365-2486.2005.01026.x>
- Schewe J, Heinke J, Gerten D, Haddeland I, Arnell NW, Clark DB, Dankers R, Eisner S, Fekete BM, Colón-González FJ, Gosling SN, Kim H, Liu X, Masaki Y, Portmann FT, Satoh Y, Stacke T, Tang Q, Wada Y, Wisser D, Albrecht T, Frieler K, Piontek F, Warszawski L, Kabat P (2013) Multimodel assessment of water scarcity under climate change. *PNAS* 111(9):3245–3250. <https://doi.org/10.1073/pnas.1222460110>
- Schuol J, Abbaspour KC (2006) Calibration and uncertainty issues of a hydrological model (SWAT) applied to West Africa. *Adv Geosci* 9:137–143
- Seghier J, Carreau J, Boulain N, De Rosnay P, Arjounin M, Timouk F (2012) Is water availability really the main environmental factor controlling the phenology of woody vegetation in the central

- Sahel? *Plant Ecol* 213(5):861–870. <https://doi.org/10.1007/s11258-012-0048-y>
- Séguis L, Cappelaere B, Milési G, Peugeot C, Massuel S, Favreau G (2004) Simulated impacts of climate change and land-clearing on runoff from a small Sahelian catchment. *Hydrol Process* 18(17):3401–3413. <https://doi.org/10.1002/hyp.1503>
- Semazzi FHM, Mehta V, Sud Y (1988) An investigation of the relationship between sub-Saharan rainfall and global sea surface temperatures. *Atmos Ocean* 26(1):118–138. <https://doi.org/10.1080/07055900.1988.9649293>
- Senay GB, Velpuri NM, Bohms S, Demissie Y, Gebremichael M (2014) Understanding the hydrologic sources and sinks in the Nile Basin using multisource climate and remote sensing data sets. *Water Resour Res* 50(11):8625–8650. <https://doi.org/10.1002/2013WR015231>
- Sheffield J, Wood EF (2008) Global trends and variability in soil moisture and drought characteristics, 1950–2000, from observation-driven simulations of the terrestrial hydrologic cycle. *J Clim* 21(3):432–458. <https://doi.org/10.1175/2007JCLI1822.1>
- Sheffield J, Ferguson CR, Troy TJ, Wood EF, McCabe MF (2009) Closing the terrestrial water budget from satellite remote sensing. *Geophys Res Lett* 36(7):L07403
- Sheffield J, Wood EF, Chaney N, Guan K, Sadri S, Yuan X, Olang L, Amani A, Ali A, Demuth S, Ogallo L (2014) A drought monitoring and forecasting system for Sub-Sahara African water resources and food security. *Bull Am Meteorol Soc* 95(6):861–882. <https://doi.org/10.1175/BAMS-D-12-00124.1>
- Shiferaw B, Tesfaye K, Kassie M, Abate T, Prasanna B, Menkir A (2014) Managing vulnerability to drought and enhancing livelihood resilience in sub-Saharan Africa: technological, institutional and policy options. *Weather Clim Extremes* 3:67–79. <https://doi.org/10.1016/j.wace.2014.04.004>
- Shiklomanov IA (2000) Appraisal and assessment of world water resources. *Water Int* 25(1):11–32. <https://doi.org/10.1080/02508060008686794>
- Singh A, Kulkarni MA, Mohanty UC, Kar SC, Robertson AW, Mishra G (2012) Prediction of Indian summer monsoon rainfall (ISMR) using canonical correlation analysis of global circulation model products. *Meteorol Appl* 19(2):179–188. <https://doi.org/10.1002/met.1333>
- Sneeuw N, Lorenz C, Devaraju B, Tourian M, Riegger J, Kunstmann H, Bárdossy A (2014) Estimating runoff using hydro-geodetic approaches. *Surv Geophys* 35(6):1333–1359. <https://doi.org/10.1007/s10712-014-9300-4>
- Spinoni J, Naumann G, Carrao H, Barbosa P, Vogt J (2014) World drought frequency, duration, and severity for 1951–2010. *Int J Climatol* 34(8):2792–2804. <https://doi.org/10.1002/joc.3875>
- Swenson S, Wahr J (2002) Methods for inferring regional surface-mass anomalies from gravity recovery and climate experiment (GRACE) measurements of time-variable gravity. *J Geophys Res Solid Earth*. <https://doi.org/10.1029/2001jb000576>
- Swenson S, Wahr J (2007) Multi-sensor analysis of water storage variations of the Caspian Sea. *Geophys Res Lett*. <https://doi.org/10.1029/2007gl030733>
- Tall M, Sylla MB, Diallo I, Pal JS, Faye A, Mbaye ML, Gaye AT (2016) Projected impact of climate change in the hydroclimatology of Senegal with a focus over the Lake of Guiers for the twenty-first century. *Theor Appl Climatol*. <https://doi.org/10.1007/s00704-016-1805-y>
- Tapley B, Bettadpur S, Watkins M, Reigber C (2004) The gravity recovery and climate experiment: mission overview and early results. *Geophys Res Lett* 31:1–4. <https://doi.org/10.1029/2004GL019920>
- Tarhule A, Zume JT, Grijzen J, Talbi-Jordan A, Guero A, Dessouassi RY, Doffou H, Kone S, Coulibaly B, Harshdeep NR (2015) Exploring temporal hydroclimatic variability in the Niger Basin (1901–2006) using observed and gridded data. *Int J Climatol* 35(4):520–539. <https://doi.org/10.1002/joc.3999>
- Thiemig V, Rojas R, Zambrano-Bigiarini M, Roo AD (2013) Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin. *J Hydrol* 499:324–338. <https://doi.org/10.1016/j.jhydrol.2013.07.012>
- Thomas AC, Reager JT, Famiglietti JS, Rodell M (2014) A GRACE-based water storage deficit approach for hydrological drought characterization. *Geophys Res Lett* 41(5):1537–1545. <https://doi.org/10.1002/2014GL059323>
- Tiwari VM, Wahr J, Swenson S (2009) Dwindling groundwater resources in northern India, from satellite gravity observations. *Geophys Res Lett* 36(18):L18401. <https://doi.org/10.1029/2009GL039401>
- Todd M, Andersson L, Ambrosino C, Hughes D, Kniveton DR, Mileham L, Murray-Hudson M, Raghavan S, Taylor R, Wolski P (2011) Climate change impacts on hydrology in Africa: case studies of river basin water resources. In: African climate and climate change: physical, social and political perspectives, Springer, Dordrecht, pp 123–153. https://doi.org/10.1007/978-90-481-3842-5_5
- Tourian M, Schwatke C, Sneeuw N (2017) River discharge estimation at daily resolution from satellite altimetry over an entire river basin. *J Hydrol* 546:230–247. <https://doi.org/10.1016/j.jhydrol.2017.01.009>
- Tucker CJ, Dregne HE, Newcomb WW (1991) Expansion and contraction of the Sahara Desert from 1980 to 1990. *Science* 253(5017):299–300. <https://doi.org/10.1126/science.253.5017.299>
- van der Molen M, Dolman A, Ciais P, Eglin T, Gobron N, Law B, Meir P, Peters W, Phillips O, Reichstein M, Chen T, Dekker S, Doubková M, Friedl M, Jung M, van den Hurk B, de Jeu R, Kruijt B, Ohta T, Rebel K, Plummer S, Seneviratne S, Sitch S, Teuling A, van der Werf G, Wang G (2011) Drought and ecosystem carbon cycling. *Agric For Meteorol* 151(7):765–773. <https://doi.org/10.1016/j.agrformet.2011.01.018>
- van Huijgevoort M (2014) Hydrological drought: characterisation and representation in large-scale models. PhD thesis, Wageningen University, Wageningen, NL. <http://library.wur.nl/WebQuery/wda/2057611>. Accessed 1 Feb 2017
- Van Loon AF, Stahl K, Di Baldassarre G, Clark J, Rangelcroft S, Wanders N, Gleeson T, Van Dijk AIJM, Tallaksen LM, Hannaford J, Uijlenhoet R, Teuling AJ, Hannah DM, Sheffield J, Svoboda M, Verbeiren B, Wagener T, Van Lanen HAJ (2016) Drought in a human-modified world: reframing drought definitions, understanding, and analysis approaches. *Hydrol Earth Syst Sci* 20(9):3631–3650. <https://doi.org/10.5194/hess-20-3631-2016>
- Velicogna I, Kimball JS, Kim Y (2015) Impact of changes in GRACE derived terrestrial water storage on vegetation growth in Eurasia. *Environ Res Lett* 10(12):124024
- Vicente-Serrano SM (2006) Spatial and temporal analysis of droughts in the Iberian Peninsula (1910–2000). *Hydrol Sci J* 51(1):83–97. <https://doi.org/10.1623/hysj.51.1.83>
- Vicente-Serrano SM, Beguería S, Lorenzo-Lacruz J, Camarero JJ, López-Moreno JJ, Azorin-Molina C, Revuelto J, Morán-Tejada E, Sanchez-Lorenzo A (2012) Performance of drought indices for ecological, agricultural, and hydrological applications. *Earth Interact* 16(10):1–27. <https://doi.org/10.1175/2012EI000434.1>
- Vörösmarty CJ, Douglas EM, Green PA, Revenga C (2005) Geospatial indicators of emerging water stress: an application to Africa. *Ambio*
- Vörösmarty CJ, Green P, Salisbury J, Lammers RB (2000) Global water resources: vulnerability from climate change and population growth. *Science* 289(5477):284–288. <https://doi.org/10.1126/science.289.5477.284>

- Vörösmarty C, Askew A, Grabs W, Barry RG, Birkett C, Döll P, Goodison B, Hall A, Jenne R, Kitaev L, Landwehr J, Keeler M, Leavesley G, Schaake J, Strzepek K, Sundarvel SS, Takeuchi K, Webster F (2001) Global water data: a newly endangered species. *EOS Trans Am Geophys Union* 82(5):54–58. <https://doi.org/10.1029/01EO00031>
- Wada Y, Bierkens MFP, de Roo A, Dirmeyer PA, Famiglietti JS, Hanasaki N, Konar M, Liu J, Müller Schmied H, Oki T, Pokhrel Y, Sivapalan M, Troy TJ, van Dijk AIJM, van Emmerik T, Van Huijgevoort MHJ, Van Lanen HAJ, Vörösmarty CJ, Wanders N, Wheeler H (2017) Human-water interface in hydrological modelling: current status and future directions. *Hydrol Earth Syst Sci* 21(8):4169–4193. <https://doi.org/10.5194/hess-21-4169-2017>
- Wagner S, Kunstmann H, Bárdossy A, Conrad C, Colditz R (2009) Water balance estimation of a poorly gauged catchment in West Africa using dynamically downscaled meteorological fields and remote sensing information. *Phys Chem Earth* 34:25–235. <https://doi.org/10.1016/j.pce.2008.04.002>
- Wahr J, Molenaar M, Bryan F (1998) Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *J Geophys Res Solid Earth* 103(B12):30205–30229. <https://doi.org/10.1029/98jb02844>
- Wald L (1990) Monitoring the decrease of Lake Chad from space. *Geocarto Int* 5(3):31–36. <https://doi.org/10.1080/10106049009354266>
- Washington R, James R, Pearce H, Pokam WM, Moufouma-Okia W (2013) Congo basin rainfall climatology: can we believe the climate models? *Philos Trans R Soc Lond B: Biol Sci* 368(1625):20120296. <https://doi.org/10.1098/rstb.2012.0296>
- Werth S, White D, Bliss DW (2017) GRACE detected rise of groundwater in the Sahelian Niger River Basin. *J Geophys Res Solid Earth*. <https://doi.org/10.1002/2017JB014845>
- White WB, Gershunov A, Annis JL, McKeon G, Syktus J (2004) Forecasting Australian drought using Southern Hemisphere modes of sea-surface temperature variability. *Int J Climatol* 24(15):1911–1927. <https://doi.org/10.1002/joc.1091>
- WMO (2013) WMO statement on the status of the global climate in 2012. World Meteorological Organization, WMO No.1108. www.wmo.int/pages/prog/wcp/wcdmp/documents/WMO1108.pdf. Accessed 15 June 2015
- WMO (2015). The climate in Africa: 2013. World Meteorological Organization, WMO No.1147. http://www.wmo.int/pages/prog/wcp/wcdmp/documents/1147_EN.pdf. Accessed 26 Jan 2017
- Wouters B, Bonin JA, Chambers DP, Riva REM, Sasgen I, Wahr J (2014) GRACE, time-varying gravity, earth system dynamics and climate change. *Rep Prog Phys* 77(11):116801. <https://doi.org/10.1088/0034-4885/77/11/116801>
- Xie H, Longuevergne L, Ringler C, Scanlon BR (2012) Calibration and evaluation of a semi-distributed watershed model of Sub-Saharan Africa using GRACE data. *Hydrol Earth Syst Sci* 16(9):3083–3099. <https://doi.org/10.5194/hess-16-3083-2012>
- Yang D, Ye B, Kane DL (2004) Streamflow changes over Siberian Yenisei River Basin. *J Hydrol* 296(1–4):59–80. <https://doi.org/10.1016/j.jhydrol.2004.03.017>
- Yang Y, Long D, Guan H, Scanlon BR, Simmons CT, Jiang L, Xu X (2014) GRACE satellite observed hydrological controls on inter-annual and seasonal variability in surface greenness over mainland Australia. *J Geophys Res Biogeosci* 119(12):2245–2260. <https://doi.org/10.1002/2014JG002670>
- Zhang D, Zhang Q, Werner AD, Liu X (2016) GRACE-based hydrological drought evaluation of the Yangtze River Basin, China. *J Hydrometeorol* 17(3):811–828. <https://doi.org/10.1175/JHM-D-15-0084.1>
- Zhou L, Tian Y, Myneni RB, Ciais P, Saatchi S, Liu YY, Piao S, Chen H, Vermote EF, Song C, Hwang T (2014) Widespread decline of Congo rainforest greenness in the past decade. *Nature* 509(7498):86–90. <https://doi.org/10.1038/nature13265>



Hydro-geochemical analysis of meltwater draining from Bilare Banga glacier, Western Himalaya

Ramesh Kumar¹ · Rajesh Kumar¹ · Atar Singh¹ · Shaktiman Singh^{1,2} · Anshuman Bhardwaj² · Anupma Kumari³ · Ravindra Kumar Sinha^{3,4} · Akhilesh Gupta⁵

Received: 5 September 2018 / Accepted: 4 February 2019 / Published online: 9 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

The changing climate is affecting the melting process of glacier ice and snow in Himalaya and may influence the hydro-geochemistry of the glacial meltwater. This paper represents the ionic composition of discharge from Bilare Banga glacier by carrying out hydro-geochemical analysis of water samples of melting season of 2017. The pH and EC were measured on-site in field, and others parameters were examined in the laboratory. The abundance of the ions observed in meltwater has been arranged in decreasing order for cations as $\text{Ca}^{2+} > \text{Mg}^{2+} > \text{Na}^+ > \text{K}^+$ and for anions as $\text{HCO}_3^- > \text{SO}_4^{2-} > \text{Cl}^- > \text{NO}_3^-$, respectively. Analysis suggests that the meltwater is mostly dominated by Ca^{2+} and HCO_3^- . It has been observed that the ionic concentration HCO_3^- is dominant and Cl^- is the least in the catchment. Piper plot analysis suggests that the chemical composition of the glacier discharge not only has natural origin but also has some anthropogenic input. Hydro-geochemical heterogeneity reflected the carbonate-dominated features ($\text{Ca}^{2+}\text{-HCO}_3^-$) in the catchment. The carbonate weathering was found as the regulatory factor to control the chemistry of the glacial meltwater due to the high enrichment ratio of ($\text{Ca}^{2+} + \text{Mg}^{2+}$) against TZ^+ and ($\text{Na}^+ + \text{K}^+$). In statistical approach, PCA analysis suggests that geogenic weathering dynamics in the catchment is associated with carbonate-dominant lithology.

Keywords Bilare Banga glacier · Hydro-geochemistry · Cations · Anions · Carbonate weathering

Introduction

The Himalayan mountain region is one of the youngest chain of mountains having the world's highest range and huge freshwater reserves as snow and glaciers that helps in sustaining the downstream population and ecosystem (Singh et al. 2016a, b; Shekhar et al. 2017). It holds hydrological importance because of the potential reserve of freshwater

as glacier, lake and river (Schild 2008). Glaciers are considered as indispensable climatic indicator as well as natural resources (Maurer et al. 2016). The component of the glaciers, such as snow and ice melting, considerably contributes to river flow at downstream, which is very crucial for the drinking, irrigation and hydropower generation (Lutz et al. 2014). In general, extreme water discharge in summer season in this region is a result of enhanced melting of snow and glaciers at higher altitude (Singh et al. 2008). It has been reported that the freshwater bodies such as glacial meltwater and glacial lake are getting contaminated under the natural as well as anthropogenic activities (Kanakiya et al. 2014; Kotadiya and Acharya 2014), generating serious problem at the global scale (Iscen et al. 2008). The responsibility of the contamination has been attributed to changing climate and global warming (Szopińska et al. 2018), but there has been no established relationship due to scarcity of database. This has led to a serious concern at regional as well as global level about fate of the glaciers and future freshwater availability. The glacial locations are ideal sections for understanding the process of water–rock interactions and the natural influence

✉ Rajesh Kumar
rajesh.kumar@sharda.ac.in

¹ Department of Environmental Science, SBSR, Sharda University, Greater Noida, India

² Department of Computer Science, Electrical and Space Engineering, Lulea University of Technology, Luleå, Sweden

³ Department of Zoology, Patna University, Patna, India

⁴ Nalanda Open University, Biscoman Bhawan, Gandhi Maidan, Patna, India

⁵ DST, Technology Bhawan, New Mehrauli Road, New Delhi, India

on the hydro-geochemistry as little interventions of human activities are noticed. The varying discharge of glacial meltwater influences the ionic concentration and dissolved load in its river system that influences the hydro-geochemistry and water quality. The glacier dynamics and higher meltwater discharge influence the rate of erosion and hence the process of physical and chemical weathering making it higher than the continental average. Therefore, it is very important for the prospective of changing water quality and needs to conduct research for recognizing the responsible factors for deteriorating water quality (Saleem et al. 2015; Vetrimurugan et al. 2013; Zhu et al. 2013). Hydro-geochemistry of Himalayan glacial meltwater is helpful in understanding the weathering dynamics of glacial environment (Chauhan and Hasnain 1993; Singh and Hasnain 1998a, b; Sharma et al. 2013; Singh et al. 2015; Kumar et al. 2018a, b). The knowledge of the hydro-geochemistry is vital due to the dependency of a large populations on glacial meltwater. The correct assessment of chemical characteristics of ions activity and other physical–chemical parameters is necessary for understanding the weathering reactions, anthropogenic impact on the water resources within the catchment. The characteristics of the glacial meltwater depend on the flow of water under subglacial conduits and its interaction with the rock surface (Haritashya et al. 2010; Kumar et al. 2009). It also depends on the velocity of water and interaction time with bedrocks (Sharma et al. 2013; Singh et al. 2014; Singh and Ramanathan 2015). Numerous glaciological studies have been carried out like glacier surge (Philip and Sah 2004), dynamics of glaciers (Sam et al. 2015), mapping of glacier facies as well as discharge and mass balance reconstruction (Singh et al. 2018), dissolved micronutrient nanoparticles (Kumar et al. 2018a) and sediment dynamics with respect to discharge (Kumar et al. 2018b) in the Shaune Garang basin, but there is not any documented research regarding the geochemical evidence in this basin. The paper attempted to highlight the chemical characteristics of the meltwater draining from Bilare Banga glacier of Baspa basin through the qualitative and quantitative methods. It is affected by the characteristics of the parent rocks as well as the weathering process. This is helpful in addressing the migration and transport mechanism of solute particles to the downstream locations. Therefore, the present study focused on the geochemical characterization of Bilare Banga (Shaune Garang basin) glacial meltwater is able to fill the research gap.

Study area

Geologically, this part of Baspa Valley comes in the group of Tethys Himalaya. A comprehensive study on the geological as well as geographical development of entire basin recommends the control of Raksham group of granite, granodiorite and pegmatite rocks (Dutta et al. 2017). Bilare Banga is the

second largest glacier of Shaune Garang catchment (Philip and Sah 2004). Glacier is located in the Shaune Garang catchment joining the Baspa River in the downstream. The glacier discharge of the catchment meets to the main Shaune Garang River at altitude of 3975 m above the sea level. Location of study region is shown through the map (Fig. 1), while the geomorphic facies of Bilare Banga glacier are presented in Table 1. The altitude of glacier varies from 4250 m to 5160 m asl. The total area of catchment is 8.54 km² having the glacier area of about 3.51 km². The study area falls under the influence of Western disturbance as well as Indian summer monsoon (Kumar et al. 2016, 2018a, b), and the intensity of precipitation events is more in summer compared to winter (Wulf et al. 2010). The chemical characterization of meltwater has not been carried out in this region; hence, this attempt in view of the geological condition of the area becomes important to know the influence in the downstream catchment.

Materials and method

Water sampling protocol and analysis

Glacial meltwater sampling was carried out at different locations of downstream catchment of Bilare Banga stream in the melting season of 2017. Sampling bottles were cleaned by distilled water followed by glacial water, before collection of water at 12 different locations of varying altitudes. Glacial meltwater was filtered on known weight of Millipore membrane (Whatman 0.45 µm) to separate suspended sediment. The meltwater is collected from the five locations in polyethylene bottle (1/4 l) about 10–15 cm deeper to the water surface directly. The in situ measurement of pH, EC and temperature has been performed through handheld multi-parameter (HI-98129, HANNA) instrument. The atomic absorption spectroscopy, (AAS) with accuracy of 0.05 ppm for Ca, Mg, K and Na while 0.01 ppm for silicon, has been used to analyze the cations (Ca²⁺, Mg²⁺, K⁺ and Na⁺). Major anions (Cl⁻, NO₃⁻ and SO₄²⁻) were analyzed through ion chromatography (PERKIN ELMEWR), Dionex ICS 900, USA (accuracy up to 0.1 ppm). The concentration of HCO₃ was analyzed through titration method as defined by the APHA (1998). The utmost precaution has been taken during water sampling, and it was preserved at 4 °C in the laboratory before subjected to analysis. A set of samples was used for evaluation of every calibration curve. Multivariate statistical analysis was applied to assess and understand the variations within the dataset. Correlation matrix analysis has been performed to examine the interrelationship among the measured variables within the dataset. Principal component analysis (PCA) is used for knowing sources of dissolved ionic concentration in glacial meltwater. Bartlett's sphericity

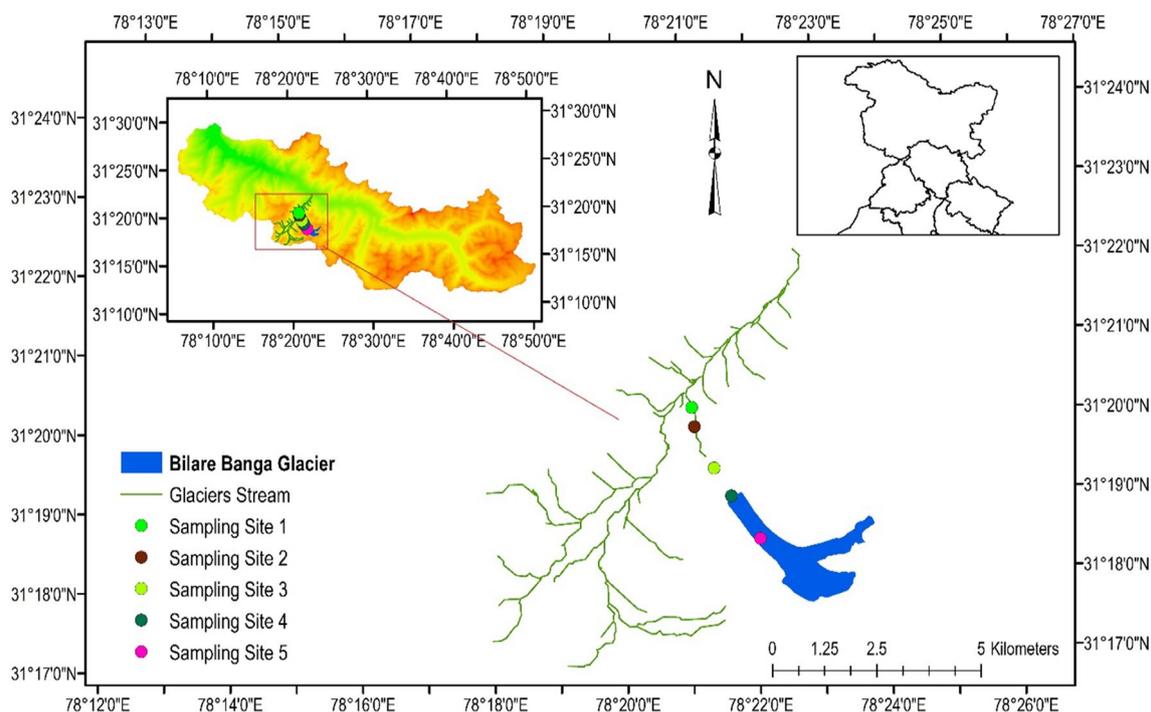


Fig. 1 Map of Bilare Banga glacier presenting the river channels and water sampling site. The location of Bilare Banga glacier (in blue color) in Shaune Garang catchment and its stream channel (in green color) and location of Baspa basin (in red color) shown in inset

Table 1 Geomorphic characteristics of Bilare Banga glaciers. *Source:* Philip and Sah (2004)

Geomorphic characteristics	
Summit height (m)	5555
Upper limit of glacier (m)	5160
Lower limit of glacier (m)	4250
Relative relief of glacier	910
Length of glacier (km)	6.42
Average slope of glacier	08°05'
Glacier area (km ²)	3.51
Glacier area (%)	22.35
Basin area (km ²)	8.54
Basin area (%)	5.47

methods were tested to know the strength and efficiency of PCA which is helpful in reducing the dimensionality in data (Kumar et al. 2014; Pant et al. 2018).

Results and discussion

Hydro-geochemistry

The accuracy of the charge balance error between total cations (TZ^+) and total anions (TZ^-) was calculated as

5.86%, which indicates the precision and quality of the data and considered acceptable. The CBE error was calculated through following equation:

$$CBE = \frac{(TZ^+ - TZ^-)}{(TZ^+ + TZ^-)} \times 100 \quad (1)$$

where TZ^+ is total cation and TZ^- is total anion.

Water draining from the Bilare Banga glacier has been found to be alkaline and may be linked to weathering of rocks (Meybeck 1987; Sharma et al. 2013), and pH varies from 6.98 to 8.23 with an average value of 7.45 ± 0.44 . Electrical conductivity ($\mu\text{S}/\text{cm}$) has shown the range of 54.0–91.0 with an average value of $73.0 \pm 12.9 \mu\text{S}/\text{cm}$. Highly varying electrical conductivity (EC) indicates that the catchment's geochemistry is controlled by meteorological parameters, interaction of water with rock that controls the hydro-chemical facies (Kumar et al. 2014, 2018b). The EC also represents the concentrations of total dissolved ions in meltwater (Shichang et al. 2000; Singh et al. 2015). Total ionic budget of meltwater shows major cation contribution of Ca^{2+} and Mg^{2+} constituting 41.07% and 28.29%, respectively, in the total cationic budget. The average Ca^{2+} and Mg^{2+} concentrations in the meltwater of Bilare Banga glacier was observed to be 288.29 ± 48.95 and $198.56 \pm 32.41 \mu\text{eq}/\text{l}$, respectively.

Similarly, anionic budget of meltwater indicates that bicarbonate (HCO_3^-) is the major anion contributor (47.81%) of the total TZ^- trailed by SO_4^{2-} (42.05%), Cl^- (9.26%) and NO_3^- (0.91%). The concentration of bicarbonate (HCO_3^-) in meltwater ranged from 214 to 350 $\mu\text{eq/l}$ having mean of $298 \pm 45.0 \mu\text{eq/l}$ followed by sulfate (SO_4^{2-}) concentration varied from 200 to 356 $\mu\text{eq/l}$ having mean of $262.4 \pm 57.5 \mu\text{eq/l}$. The concentrations of dissolved silica fluctuated from 14.6 to 25.8 $\mu\text{mol/l}$ with mean value of $20.5 \pm 4.18 \mu\text{mol/l}$. Study shows that Ca^{2+} and HCO_3^- are dominant in the glacial meltwater of Bilare Banga. The cationic abundance in the meltwater follows decreasing order like $\text{Ca}^{2+} > \text{Mg}^{2+} > \text{Na}^+ > \text{K}^+$, and the concentration of anion follows decreasing order from $\text{HCO}_3^- > \text{SO}_4^{2-} > \text{Cl}^- > \text{NO}_3^-$. Excessive concentration of bicarbonate in meltwater is an indicative of higher weathering of silicate-dominant lithology of the area, whereas the higher Cl^- ion may be endorsed to the contribution of rainwater (Singh et al. 2014). The detailed distribution of the dissolved ionic concentration in meltwater of Bilare Banga glacier is presented through Table 2.

Hydro-chemical process in glacial catchment

The evolution of the chemical species in the glacial melt has been accredited to weathering process of parental rock present in the glaciated region (Singh et al. 2012, 2014). The weathering of the chemical species in the rocks such as calcite, dolomite and dissolution or evaporation process of halite gypsum and anhydrite are the main factors in

Table 2 Chemical characteristics of Bilare Banga glacier meltwater during ablation season of year 2017

Parameters	Minimum	Maximum	Average	SD	CV %
pH	6.98	8.23	7.45	0.44	5.86
EC	54	91	73	12.9	17.67
Na+	121	204.45	162.75	25.79	15.84
K ⁺	32	78	52.18	12.87	24.66
Ca ²⁺	178	356	288.29	48.95	16.98
Mg ²⁺	141.01	253.81	198.56	32.41	16.32
Cl ⁻	48	78	57.8	10.07	17.42
SO ₄ ²⁻	200	356	262.45	57.51	21.91
NO ₃ ⁻	2.6	8.98	5.74	2.33	40.6
HCO ₃ ⁻	214	350	298.04	45.07	15.12
H ₄ SiO ₄	14.62	25.82	20.55	4.18	20.33
TZ+	631.21	857.73	701.78	59.01	8.41
TZ-	516.3	703.1	624.02	52.82	8.46
HCO ₃ +SO ₄	464	645.5	560.49	48.91	8.73
Ca+Mg	425.21	609.81	486.85	47.69	9.8
Na+K	175	265.1	214.93	25.97	12.08

Unit: EC in $\mu\text{S/cm}$; TZ⁺, TZ⁻ in $\mu\text{eq/l}$, H₄SiO₄ in $\mu\text{mol/l}$

controlling the hydro-geochemistry of the glacial meltwater (Pant et al. 2018). Richness of cations of calcium, magnesium and carbonates in glacial meltwater may be credited to the presence of carbonate minerals and its weathering. However, the richness of Ca^{2+} , Mg^{2+} , Na^+ and K^+ is also due to the weathering of silicate. Low content of chlorine and sodium ions and high content of bicarbonates and calcium ions in meltwater reflect the stimulus weathering of the rocks having the richness of minerals like carbonates, silicates and sulfide (Tranter et al. 1993) as well as atmospheric CO₂. Chloride (Cl^-) and sulfate (SO_4^{2-}) are primarily generated with the halite and sulfide oxidation in the river catchment. Sulfate minerals such as gypsum are also a main factor for controlling the concentration of Cl^- and SO_4^{2-} in the glacial meltwater (Mortatti and Probst 2003; Thomas et al. 2015). To detect the derivations of chemical species in water deliberately by weathering phenomena, ionic ratios of water are computed and shown through Table 3. Average ratio of $\text{HCO}_3^-/\text{Na}^{2+}$ is 1.87 ± 0.35 , which indicates that glacial meltwater is influenced by carbonate weathering. In addition, the proportion of $(\text{Ca}^{2+} + \text{Mg}^{2+})/\text{TZ}^+$ is 0.69 ± 0.03 and $(\text{Na}^+ + \text{K}^+)/\text{TZ}^+$ is 0.31 ± 0.03 , which indicates the presence of silicate weathering process is lesser than the carbonate weathering (Pant et al. 2018). $(\text{Ca}^{2+} + \text{Mg}^{2+})$ versus $(\text{Na}^+ + \text{K}^+)$ ratio observed to be 2.29 ± 0.29 suggests meltwater is influenced by the minerals of calcite and dolomite. The C-ratio [$(\text{HCO}_3^-)/(\text{HCO}_3^- + \text{SO}_4^{2-})$] explains the importance of proton-producing reactions such as carbonation and sulfide oxidation being an essential component for chemical weathering of carbonate rocks (Huang et al. 2008). C-ratio < 0.50 represents the occurrence of major reaction in the glacial meltwater such as carbonate dissolution (carbonation) as well as sulfide oxidation reaction, although values of the whole C-ratio closer to 1 indicate dominance of carbonic reaction in the region as well as

Table 3 Ionic ratios of various hydro-geochemical species during ablation period of year 2017

Parameters	Minimum	Maximum	Average	SD	CV %
Ca ²⁺ +Mg ²⁺ /TZ ⁺	0.62	0.74	0.69	0.03	4.29
Ca ²⁺ +Mg ²⁺ / Na ⁺ +K ⁺	1.61	2.78	2.29	0.29	12.84
Na ⁺ +K ⁺ /TZ ⁺	0.26	0.38	0.31	0.03	9.72
Ca ²⁺ /Na ⁺	0.89	2.39	1.82	0.39	21.43
Mg ²⁺ /Na ⁺	0.97	1.85	1.24	0.25	20.32
HCO ₃ ⁻ /Na ⁺	1.07	2.39	1.87	0.35	18.88
Na ⁺ /Cl ⁻	1.56	4.17	2.91	0.69	23.73
K ⁺ /Cl ⁻	0.49	1.35	0.91	0.22	23.51
C-ratio	0.41	0.63	0.53	0.08	15.08
SMF	0.37	0.59	0.47	0.08	17.26

atmospheric input of CO₂-derived protons. We observed the C-ratio of meltwater from Bilare Banga as 0.53. The C-ratio of glaciers like Sutri Dhaka, Chhota Shigri, Dudu and Gangotri is represented through Table 4 which reflects the dominance of carbonate weathering in Chhota Shigri, Sutri Dhaka as well as Bilare Banga glacier, while Gangotri and Dudu glaciers are majorly influenced by sulfide weathering. About 85% contribution of bicarbonate in the glacial meltwater is due to higher weathering of carbonate while other 15% is contributed by the weathering of silicate (Raymahasay 1986). The similar result has been reported by Singh and Hasnain (1998a, b) for the Alaknanda basin where the dominance of carbonate weathering is 78% while silicate weathering is of 22%. The Rajkhot watershed of Higher Himalaya also showed the similar trend (Blum et al. 1998). In the same manner, the domination of carbonate weathering designates the partial role of assimilation of atmospheric CO₂ in the

Bilare Banga glacier meltwater. However, the strontium isotopes study would be helpful in confirming the role of atmospheric CO₂ assimilation (Kumar et al. 2009) in the Himalayan region.

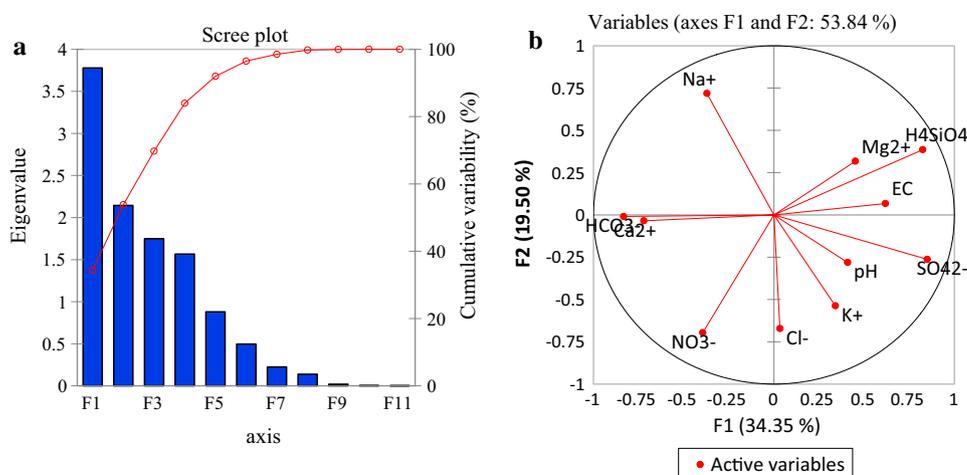
Statistical analysis

Approach of statistical analysis, such as principal component analysis, correlation as well as factor analysis, was executed to understand the interrelationship among the variables. The Bartlett's sphericity test having χ^2 (observed) = 125.76, which is substantially greater than the critical value ($\chi^2 = 38.95$ at 'degree of freedom' 55, p values < 0.0001, significance level 0.05), shows the importance of PCA in understanding substantial decrease in dimensionality of the data (Singh et al. 2011; Kumar et al. 2018b). The PCA helps in standardization of the data collected on different scale of measurements through the diagonalization of correlation

Table 4 Comparison of ionic ratio of glacier meltwater draining from different parts of Himalayan region

	Bilare Banga	Sutri	Gangotri	Chhota Shigri	Dudu
pH	7.45	8.2	7.2	6.5	6.3
EC	73	59.9	81	42.6	31.9
Na ⁺	162.75	5	75	38.8	42.7
K ⁺	52.18	14.7	83	28.9	24.6
Ca ²⁺	288.29	242	206	104	91.1
Mg ²⁺	198.56	43.6	197	98.8	6.5
Cl ⁻	57.8	0.3	11.2	3.3	5
SO ₄ ²⁻	262.45	71.6	401	104	85.4
NO ₃ ⁻	5.74	3.3	1.9	1.5	14.9
HCO ₃ ⁻	298.04	221	266	219	52.3
Ca ²⁺ + Mg ²⁺ /TZ ⁺	0.69	0.94	0.72	0.75	0.59
Ca ²⁺ + Mg ²⁺ /Na ⁺ + K ⁺	2.29	14.5	2.55	3	1.45
Na ⁺ + K ⁺ /TZ ⁺	0.31	0.06	0.28	0.25	0.41
Na ⁺ /Cl ⁻	2.91	16.67	6.7	11.7	8.5
C-ratio	0.53	0.76	0.40	0.68	0.38

Fig. 2 a Scree plot between eigenvalue and factors, b loading score of factor



matrix. The variables in this statistical test automatically give the auto scale to mean = 0 and variance = 1. The scree plot between eigenvalues and factors (Fig. 2) shows that the principle component (PC) values > 1 explain 83.9% of the total variance among four PCs. PC1, PC2, PC3 and PC4 are able to explain 34.34%, 19.49%, 15.90% and 14.23%, respectively, of the total variance. Statistical tools such as correlation matrix are used to find out the interrelationship between two hydro-geochemical species for investigating the grade of dependence of variables among each other (Ramanathan 2007; Vasanthavigar et al. 2013; Pant et al. 2018). Table 5 represents the correlation analysis during the study period. Some of the parameters like EC and Na⁺ are strongly correlated with SO₄²⁻, and similarly, a good correlation is also observed for Ca²⁺–HCO₃⁻, Mg²⁺–H₄SiO₄

and Cl⁻–NO₃⁻. All these water parameters show the strong and positive correlation ($r > 0.6$), indicating that the hydro-geochemistry of Blaie Banga glacier is primarily influenced by them. Factor analysis of the glacial meltwater shows that factor-1 with eigenvalue 3.77 describes 34.3% of the total variance considered as main factor (Table 6). On the basis of results, total variance is very high for Ca²⁺, SO₄²⁻ and HCO₃⁻ but has moderate loadings of electrical conductivity. This result determines that the EC in the meltwater is largely influenced through the presence of sodium and chloride; however, bicarbonate also plays significant role in determining conductivity. Factor 2 describes 19.4% of total variance and demonstrates that the high loading of sodium, nitrate and chloride in the meltwater is due to ion exchange at water interface and dissolution of sodium-bearing minerals and

Table 5 The correlation matrix of chemical parameters of Blaie Banga glacier meltwater

Variables	pH	EC	Na ⁺	K ⁺	Ca ²⁺	Mg ²⁺	Cl ⁻	SO ₄ ²⁻	NO ₃ ⁻	HCO ₃ ⁻	H ₄ SiO ₄
pH	1										
EC	-0.084	1									
Na ⁺	-0.448	-0.349	1								
K ⁺	0.368	-0.035	-0.235	1							
Ca ²⁺	-0.169	-0.062	0.020	-0.331	1						
Mg ²⁺	-0.023	0.313	0.436	0.204	-0.369	1					
Cl ⁻	-0.013	0.105	-0.256	0.336	-0.015	0.254	1				
SO ₄ ²⁻	0.320	0.712	-0.718	0.134	-0.460	0.115	0.127	1			
NO ₃ ⁻	-0.319	-0.277	-0.182	0.204	0.135	-0.136	0.604	-0.134	1		
HCO ₃ ⁻	-0.318	-0.242	0.153	-0.470	0.921	-0.311	0.100	-0.568	0.287	1	
H ₄ SiO ₄	0.144	0.637	-0.077	-0.046	-0.484	0.578	-0.108	0.684	-0.534	-0.512	1

The correlation values greater than 0.5 has been marked bold that signifies the acceptable limit of correlation at the significance level of 0.05

Table 6 Factor analysis of glacial meltwater during ablation period of 2017

	F1	F2	F3	F4	Initial communal-ity	Final commu-nality
pH	0.335	0.197	0.034	-0.374	0.996	0.292
EC	0.595	0.034	0.370	0.489	0.999	0.731
Na ⁺	-0.377	-0.836	-0.371	0.145	0.988	1.000
K ⁺	0.289	0.319	-0.485	-0.178	0.946	0.451
Ca ²⁺	-0.683	0.124	0.433	0.245	0.996	0.729
Mg ²⁺	0.424	-0.398	-0.442	0.469	0.995	0.753
Cl ⁻	0.031	0.559	-0.451	0.491	0.939	0.757
SO ₄ ²⁻	0.870	0.353	0.292	0.111	0.993	0.980
NO ₃ ⁻	-0.366	0.578	-0.408	0.246	0.998	0.695
HCO ₃ ⁻	-0.843	0.092	0.352	0.397	0.993	1.000
H ₄ SiO ₄	0.824	-0.351	0.187	0.274	0.998	0.913
Eigenvalue	3.778	2.145	1.749	1.566		
Variability (%)	34.347	19.496	15.900	14.235		
Cumulative %	34.347	53.843	69.743	83.978		

The factor values greater than 0.5 has been made bold that indicates a significant association of three principle components at the significance level of 0.05

silicate-wearing minerals. Similarly, factors 3 and 4 explain 15.9% and 14.2% of total variance, respectively, during the study period. Factor analysis further suggests a high loading of magnesium that influences the temporary hardness of the water. Thus, all these four PCs are collectively responsible toward 83.96% of total variance. Apart from this, communal-ity of the variables and their proportion define that extracted common factor is larger than almost 0.8. Hence, factor analysis represents all variances of the dataset. The particular control of variables in each factor specifies the partial mixing of different water types. Hierarchical cluster analysis shows that glacial meltwater is divided into two principal

cluster components which are presented through the den-drogram (Fig. 3). Maximum water samples lie in cluster 1, which has Ca-HCO_3^- and mixed Ca-Mg-SO_4^{2-} water-type facies. On the other hand, cluster 2 defines the less dominant water types in the catchment.

Hydro-geochemical facies

To perform the quantification of water type, Piper’s trilinear diagram is broadly used to understand the geochemical evolution in water system (Piper 1944, 1953). The classification of water types in the piper plot demonstrates the basis of ratio of the geochemical characteristics in water (Xiao et al. 2012; Singh et al. 2016a, b). Characterization of glacial meltwater of the study area has been carried out by plotting average data of ablation season of 2017 on Piper’s trilinear diagram (Fig. 4). The outcomes of the study on the basis of piper plot encompass two triangles at base, which designate anions (SO_4^{2-} and $\text{CO}_3^{2-} + \text{HCO}_3^-$) and cations (Ca^{2+} , Mg^{2+} and $\text{Na}^+ + \text{K}^+$), while a diamond shape specifies dominance of anions and cations. Hydro-geochemical results helped in concluding the water type which is mainly influ-enced by Ca-HCO_3^- and mixed influence of Ca-Mg-SO_4^{2-} , demonstrating the dominance of carbonate weathering in the catchment. Ca-HCO_3^- water type is the predominant phase of water in any natural system. Mixed Ca-Mg-SO_4^{2-} facies of water is generally formed with the interaction of pyrite and gypsum minerals present in the catchment. The gradual

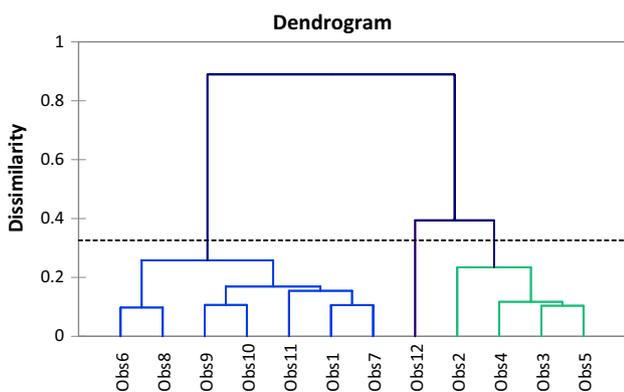
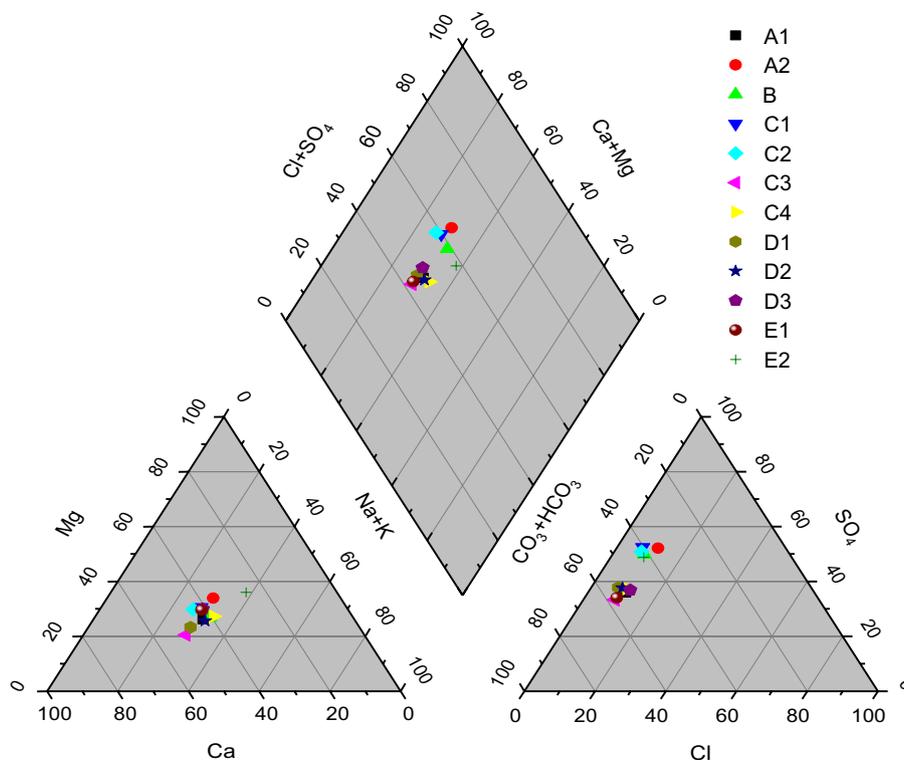


Fig. 3 Hierarchical dendrogram for linkage of 12 water sample parameters of cluster analysis

Fig. 4 Piper plot showing the chemical characterization of glacial meltwater samples from Bilare Banga glacier during ablation season of year 2017



changes in the chemical features may be attributed to the longer interaction between minerals and flowing water. On the other hand, cation plot shows that maximum number of the water samples falls in the middle part of the trilinear plot. Some of samples fall in the bottom left corner, representing the governance of Ca^+ ion in the meltwater (Singh and Ramanathan 2017). Some of the water facies show increased concentration of Na^+ and K^+ ions which is an indicative of stimulus effect of local sources of pollutants that would have accumulated on the glaciers surface after being trapped in the surface ice through deposition from the atmosphere (Karim and Veizer 2000; Ravikumar et al. 2013; Pant et al. 2018). The anion diagram helps in understanding that the maximum water samples fall under the lower left corner to HCO_3^- apex, indicating the dominancy of carbonate. The results show a complete characteristic of hydro-geochemistry, governance of alkali earth elements (69.36%) like calcium and magnesium over the alkaline (30.62%) sodium and potassium concentrations. In anion contribution of weak acid, HCO_3^- (~47.81%) is dominating over the strong acid SO_4^{2-} (~42.05%) in the catchment.

Conclusion

The present study provides fundamental information to chemical characterization of meltwater of Bilare Banga glacier. Chemical characteristics of meltwater show that pH of water moves considerably from acidic to neutral with the presence of Ca^{2+} and HCO_3^- as dominant ions. High ratio of $(\text{Ca}^{2+} + \text{Mg}^{2+})$ versus TZ^+ and $(\text{Na}^+ + \text{K}^+)$ as well as strong correlation between $\text{Ca}^{2+}-\text{HCO}_3^-$, $\text{Ca}^{2+}-\text{Mg}^{2+}$ and $\text{Mg}^{2+}-\text{HCO}_3^-$ indicates that carbonate weathering is the controlling factor of dissolved ions in glacial meltwater. The changing concentration of the ions in the river water is inversely associated with the runoff. It is further observed that the ratio of “ $\text{Na}^+ / (\text{Na}^+ + \text{Ca}^{2+})$ and $\text{Cl}^- / (\text{Cl}^- + \text{HCO}_3^-)$ ” in the catchment is less than 0.5 which is an indicative of the control of weathering on the ionic compositions in the glacial meltwater. Piper diagram shows that $\text{Ca}^{2+}-\text{HCO}_3^-$ is most prevailing water type in this region. The *C*-ratio demonstrates the process of dissolution and dissociation of CO_2 present in the local atmosphere which also controls the production of H^+ (proton). Hierarchical cluster analysis shows that the glacial meltwater is divided into two principal cluster components and most of the water samples fall in the cluster 1, which has $\text{Ca}-\text{HCO}_3^-$ and mixed $\text{Ca}-\text{Mg}-\text{SO}_4^{2-}$ water-type facies. Based on the statistical analysis, it is clear that the hydro-geochemistry in this region is mainly governed through the weathering of carbonate and silicate-dominant lithology. In addition, suspension of sulfate-bearing minerals, pyrite oxidation and precipitation are also contributing for the same.

Acknowledgement The support of the Ministry of Earth Sciences (MoES/PAMC/H&C/61/2015-PCII, dated March 03, 2016) through the project on Shaune Garang glacier sanctioned to Dr. Rajesh Kumar (the PI of the project) is thankfully acknowledged which has been helpful in building the glaciological human resource whose support is vital in the sampling of the meltwater as well as in the analysis. The support of USAID through a project “Contribution to High Asia Runoff from Ice and Snow” (CHARIS) under the collaboration of University of Colorado, Boulder, USA, is also acknowledged. The USAID support through CHARIS project helped not only in the research work but also in producing three PhD theses of students working under me (Dr. Rajesh Kumar). The critical review of the paper and suggestions provided by the anonymous reviewers and editor has been vital in improving the quality of the paper, and authors are grateful for the suggestions.

References

- American Public Health Association (APHA) (1998) Standard methods for the examination of water and waste water, 20th edn. APHA, AWWA, WPCF, Washington
- Blum JD, Gazis CA, Jacobson AD, Chamberlain CP (1998) Carbonate versus silicate weathering in the Raikhot watershed within the High Himalayan Crystalline series. *Geology* 26(5):411–414
- Chauhan DS, Hasnain SI (1993) Chemical characteristics, solute and suspended sediments load in meltwater draining Satopanth and Bhagirathi Kharak glaciers, Ganga basin, India. In: Young GJ (ed) Snow and glacier hydrology, vol 218. IAHS Press, Wallingford, pp 2–10
- Dutta S, Mujtaba SAI, Saini HS, Chunchekar R, Kumar P (2017) Geomorphic evolution of glacier-fed Baspa Valley, NW Himalaya: record of Late Quaternary climate change, monsoon dynamics and glacial fluctuations. In: Pant NC, Ravindra R, Srivastava D, Thompson LG (eds) The Himalayan Cryosphere: past and present, vol 462. Geological Society, London. <https://doi.org/10.1144/SP462.5>
- Haritashya UK, Kumar A, Singh P (2010) Particle size characteristics of suspended sediment transported in meltwater from the Gangotri Glacier, central Himalaya—an indicator of subglacial sediment evacuation. *Geomorphology* 122(1–2):140–152. <https://doi.org/10.1016/j.geomorph.2010.06.006>
- Huang X, Sillanpää M, Duo B, Gjessing ET (2008) Water quality in the Tibetan Plateau: metal contents of four selected rivers. *Environ Pollut* 156(2):270–277
- Iscen CF, Emiroglu O, Ilhan S, Arslan N, Yilmaz V, Ahiska S (2008) Application of multivariate statistical techniques in the assessment of surface water quality in Ulubat Lake, Turkey. *Environ Monit Assess* 144(1–3):269–276
- Kanakiya RS, Singh SK, Sharma JN (2014) Determining the water quality index of an urban water body Dal Lake, Kashmir, India. *IOSR J Environ Sci Toxicol Food Technol* 08(12):64–71. <https://doi.org/10.9790/2402-081236471>
- Karim A, Veizer J (2000) Weathering processes in the Indus River basin: implications from riverine carbon, sulfur, oxygen, and strontium isotopes. *Chem Geol* 170:153–177
- Kotadiya NG, Acharya CA (2014) An assessment of lake water quality index of Manipu Lake of district Ahmedabad, Gujarat. *Int J Sci Res* 03(4):448–450
- Kumar K, Miral MS, Joshi S, Pant N, Joshi V, Joshi LM (2009) Solute dynamics of melt water of Gangotri glacier, Garhwal Himalaya, India. *Environ Geol* 58:1151–1159. <https://doi.org/10.1007/s00254-008-1592-6>
- Kumar A, Verma A, Dobhal DP, Mehta M, Kesarwani K (2014) Climatic control on extreme sediment transfer from Dokriani Glacier

- during monsoon, Garhwal Himalaya (India). *J Earth Syst Sci* 123:109–120
- Kumar R, Singh S, Kumar R, Singh A, Bhardwaj A, Sam L, Randhawa SS, Gupta A (2016) Development of a glaciological model for discharge and mass balance reconstruction. *J Water Resour Manag*. <https://doi.org/10.1007/s11269-016-1364-0>
- Kumar R, Kumar R, Singh A, Sinha RK, Kumari A (2018a) Nanoparticles in glacial melt water. *Mater Today Proc* 5(3P1):9161–9166. <https://doi.org/10.1016/j.matpr.2017.10.037>
- Kumar R, Kumar R, Singh S, Singh A, Bhardwaj A, Kumari A, Randhawa SS, Saha A (2018b) Dynamics of suspended sediment load with respect to summer discharge and temperatures in Shaune Garang glacierized catchment. *Acta Geophys, Western Himalaya*. <https://doi.org/10.1007/s11600-018-0184-4>
- Lutz AF, Immerzeel WW, Shrestha AB, Bierkens MFP (2014) Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. *Nat Clim Change* 4:587–592. <https://doi.org/10.1038/nclimate2237>
- Maurer JM, Rupper SB, Schaefer JM (2016) Quantifying ice loss in the eastern Himalayas since 1974 using declassified spy satellite imagery. *Cryosphere* 10:2203–2215. <https://doi.org/10.5194/tc-10-2203-2016>
- Meybeck M (1987) Global chemical weathering of surficial rocks estimated from river dissolved loads. *Am J Sci* 287:401–428
- Mortatti J, Probst JL (2003) Silicate rock weathering and atmospheric/soil CO₂ uptake in the Amazon basin estimated from river water geochemistry: seasonal and spatial variations. *Chem Geol* 197:177–196
- Pant RR, Zhang F, Rehman FU, Wang G, Ming Y, Zeng C, Tang H (2018) Spatiotemporal variations of hydro-geochemistry and its controlling factors in the Gandaki River Basin, Central Himalaya Nepal. *Sci Total Environ* 622–623(2018):770–782. <https://doi.org/10.1016/j.scitotenv.2017.12.063>
- Philip G, Sah MP (2004) Mapping repeated surges and retreat of glaciers using IRS-1C/1D data: a case study of Shaune Garang glacier, northwestern Himalaya. *Int J Appl Earth Obs Geoinf* 6(2):127–141. <https://doi.org/10.1016/j.jag.2004.09.002>
- Piper AM (1944) A graphic procedure in the geochemical interpretation of water analyses. *Trans Am Geophys Union* 25:914–928. <https://doi.org/10.1029/TR025i006p00914>
- Piper AM (1953) A graphic procedure in the geochemical interpretation of water analysis. *Ground water note 12*. U.S. Dept. of the Interior, Geological Survey, Water Resources Division, Ground Water Branch, Washington, p 63
- Ramanathan AL (2007) Seasonal variation in the major ion chemistry of Pandoh Lake, Mandi district, Himachal Pradesh, India. *Appl Geochem* 22:1736–1747
- Ravikumar P, Somashekar RK, Mehmood MA (2013) Water quality index to determine the surface water quality of Sankey tank and Mallathahalli Lake, Bangalore urban district, Karnataka, India. *Appl Water Sci* 3:247–261
- Raymahasay BC (1986) Geochemistry of bicarbonate in the river water. *J Geol Soc India* 27:114–118
- Saleem M, Jeelani G, Shah RF (2015) Hydro geochemistry of Dal Lake and the potential for present, future management by using facies, ionic ratios, and statistical analysis. *Environ Earth Sci* 74(4):3301–3313. <https://doi.org/10.1007/s12665-015-4361-3>
- Sam L, Bhardwaj A, Singh S, Kumar R (2015) Remote sensing in glacier velocity estimation and a novel approach for debris covered glaciers. *Prog Phys Geogr*. <https://doi.org/10.1177/0309133315593894>
- Schild A (2008) The case of the Hindu Kush–Himalayas: ICIMOD's position on climate change and mountain systems. *Mt Res Dev*. <https://doi.org/10.1659/mrd.mp009>
- Sharma P, Ramanathan AL, Pottakkal JG (2013) Study of solute sources and evolution of hydro-geochemical processes of the Chhota Shigri Glacier meltwaters, Himachal Himalaya, India. *Hydrol Sci J* 58(5):1128–1143. <https://doi.org/10.1080/0262667.2013.802092>
- Shekhar M, Bhardwaj A, Singh S, Ranhotra PS, Bhattacharyya A, Pal AK, Roy I, Torres JM, Zorzano MP (2017) Himalayan glaciers experienced significant mass loss during later phases of little ice age. *Sci Rep* 7:10305. <https://doi.org/10.1038/s41598-017-09212-2>
- Shichang K, Dahe Q, Tandong Y (2000) A study on precipitation chemistry in the late summer in the northern slope of Mt. Xiabangma. *Acta Sci Circumst* 20(5):574–578
- Singh AK, Hasnain SI (1998a) Major ion chemistry and weathering control in a high altitude basin: Alaknanda River, Garhwal Himalaya, India. *Hydrol Sci* 43(6):825–843
- Singh AK, Hasnain SI (1998b) Major ion chemistry and weathering control in a high altitude basin: Alaknanda River, Garhwal Himalaya, India. *Hydrol Sci J* 43(6):825–843. <https://doi.org/10.1080/02626669809492181>
- Singh VB, Ramanathan AL (2015) Assessment of solute and suspended sediment acquisition processes in the Bara Shigri glacier meltwater (Western Himalaya, India). *Environ Earth Sci* 74:2009–2018
- Singh AK, Mondal GC, Kumar S, Singh TB, Tewari BK, Sinha A (2008) Major ion chemistry, weathering processes and water quality assessment in upper catchment of Damodar River basin, India. *Environ Geol* 54(4):745–758. <https://doi.org/10.1007/s00254-007-0860-1>
- Singh CK, Shashtri S, Mukherjee S (2011) Integrating multivariate statistical analysis with GIS for geochemical assessment of groundwater quality in Shiwaliks of Punjab, India. *Environ Earth Sci* 62:1387–1405
- Singh VB, Ramanathan AL, Jose PG, Sharma P, Linda A, Azam MF, Chatterjee C (2012) Chemical characterisation of meltwater draining from Gangotri Glacier, Garhwal Himalaya, India. *J Earth Syst Sci* 121(3):625–636
- Singh VB, Ramanathan AL, Jose PG, Kumar M (2014) Seasonal variation of the solute and suspended sediment load in Gangotri glacier meltwater, central Himalaya, India. *J Asian Earth Sci* 79:224–234
- Singh VB, Ramanathan AL, Sharma P, Pottakkal JG (2015) Dissolved ion chemistry and suspended sediment characteristics of melt water draining from Chhota Shigri glacier, Western Himalaya, India. *Arab J Geosci* 8:281–293
- Singh S, Kumar R, Bhardwaj A, Sam L, Shekhar M, Singh A, Kumar R, Gupta A (2016a) Changing climate and glacio-hydrology in Indian Himalayan Region: a review. *Wiley Interdiscip Rev Clim Change* 7(3):393–410. <https://doi.org/10.1002/wcc.39>
- Singh VB, Ramanathan AL, Mandal A (2016b) Hydrogeochemistry of high-altitude lake: a case study of the Chandra Tal, Western Himalaya, India. *Arab J Geosci* 9(4):308. <https://doi.org/10.1007/s12517-016-2358-1>
- Singh VB, Ramanathan AL (2017) Hydrogeochemistry of the Chhota Shigri glacier meltwater, Chandra basin, Himachal Pradesh, India: solute acquisition processes, dissolved load and chemical weathering rates. *Environ Earth Sci* 76(5):223
- Singh S, Kumar R, Bhardwaj A, Kumar R, Singh A (2018) Changing climate and glacio-hydrology: a case study of Shaune Garang basin, Himachal Pradesh. *Int J Hydrol Sci Technol*. <https://doi.org/10.1504/IJHST.2018.10010353>
- Szopińska M, Szumińska D, Bialik RJ, Chmiel S, Plenzler J, Polkowska Z (2018) Impact of a newly-formed periglacial environment and other factors on fresh water chemistry at the western shore of Admiralty Bay in the summer of 2016 (King George Island, Maritime Antarctica). *Sci Total Environ* 613–614(2018):619–634. <https://doi.org/10.1016/j.scitotenv.2017.09.060>
- Thomas J, Joseph S, Thriwikramji KP (2015) Hydro-chemical variations of a tropical mountain river system in a rain shadow region

- of the southern Western Ghats, Kerala, India. *Appl Geochem* 63:456–471
- Tranter M, Brown GH, Raiswell R, Sharp MJ, Gurnell AM (1993) A conceptual model of solute acquisition by Alpine glacier meltwaters. *J Glaciol* 39(133):573–581
- Vasanthavigar M, Srinivasamoorthy K, Prasanna MV (2013) Identification of groundwater contamination zones and its sources by using multivariate statistical approach in Thirumanimuthar sub basin, Tamil Nadu, India. *Environ Earth Sci* 68:1783–1795
- Vetrimurugan E, Elango L, Rajmohan N (2013) Sources of contaminants and groundwater quality in the coastal part of a river delta. *Int J Environ Sci Technol* 10(3):473–486. <https://doi.org/10.1007/s13762-012-0138-3>
- Wulf H, Bookhagen B, Scherler D (2010) Seasonal precipitation gradients and their impact on fluvial sediment flux in the Northwest Himalaya. *Geomorphology* 118:13–21
- Xiao J, Jin ZD, Zhang F, Wang J (2012) Solute geochemistry and its sources of the ground waters in the Qinghai Lake catchment, NW China. *J Asian Earth Sci* 52:21–30. <https://doi.org/10.1016/j.jseae.2012.02.006>
- Zhu B, Yu J, Qin X, Rioual P, Jiang F, Liu Z, MuY LH, Ren X, Xiong H (2013) Identification of rock weathering and environmental control in arid catchments (northern Xinjiang) of Central Asia. *J Asian Earth Sci* 66:277–294. <https://doi.org/10.1016/j.jseae.2013.02.005>



Analysis of extreme flow uncertainty impact on size of flood hazard zones for the Wronki gauge station in the Warta river

Tomasz Dysarz¹ · Joanna Wicher-Dysarz¹ · Mariusz Sojka² · Joanna Jaskuła²

Received: 6 October 2018 / Accepted: 5 February 2019 / Published online: 13 February 2019
© The Author(s) 2019

Abstract

In this paper, the impact of maximum flow uncertainty on flood hazard zone is analyzed. Two factors are taken into account: (1) the method for determination of maximum flows and (2) the limited length of the data series available for calculations. The importance of this problem is a consequence of the implementation of the EU Flood Directive in all EU member states. The factors mentioned seem to be among the most important elements responsible for potential uncertainty and inaccuracy of the developed flood hazard maps. Two methods are analyzed, namely the quantiles method and the maximum likelihood method. The maximum flows are estimated for the Wronki gauge station located in the reach of the Warta river. This simple river system is located in the central part of Poland. The length of the available data is 44 years. Hence, the series of the lengths 40, 30 and 20 years are tested and compared with reference calculations for 44 years. The hydrodynamic model HEC-RAS is used to calculate water surface profiles in steady state flow. The Python scripting language is applied for automation of HEC-RAS calculations and processing of final results in the form of inundation maps. The number of trials for each factor is not huge to keep the presented methodology useful in practice. The chosen measure of uncertainty is the range of variability for maximum flow values as well as inundation areas. The estimated values stressed the great importance of the factors analyzed for the uncertainty of the maximum flows as well as inundation areas. The impact of the data series length on the maximum flows is straightforward; a shorter data series gives a wider range of variability. However, the dependencies between other factors are more complex. Hence, the application of methodology based on the simulation and GIS data processing for assessment of this problem seems to be quite a good approach.

Keywords Flood hazard maps · Hydrological uncertainty · River flow simulation · Geo-processing · Python scripting

Introduction

The main focus in the present paper is uncertainty of flood hazard maps. The motivation for such research is implementation of the EU Flood Directive (European Parliament 2007) in EU member states. Although the concepts introduced in the EU Flood Directive could significantly reduce the risk related to flood damage, there are two general problems. These are (1) non-uniformity of the approaches applied in different EU member states and (2) uncertainty of

the developed maps. The first is reported by many researchers and consists in the differences at many steps of flood hazard map preparation. The most important are (a) different quality of topographic data (Van Alpen and Passchier 2007), (b) different methods applied for calculation of maximum flows with a specific return period (Van Alpen and Passchier 2007), (c) different return periods for representation in flood hazard maps (Van Alpen and Passchier 2007; Nones 2015) and (d) different hydraulic models applied (Van Alpen and Passchier 2007). Additionally, the lack of necessary data for flood damage analysis and related risk assessment is also reported (e.g., Albano et al. 2017) Although some researchers express the need for standardization over the EU member states (Nones 2017), legal regulations in this area have not been established yet.

The second problem seems to be more important, because it is related to our lack of knowledge. The uncertainty of the flood hazard maps should be investigated taking into

✉ Tomasz Dysarz
tomasz.dysarz@up.poznan.pl

¹ Department of Hydraulic and Sanitary Engineering, Poznan University of Life Sciences, Poznan, Poland

² Institute of Land Improvement, Environmental Development and Geodesy, Poznan University of Life Sciences, Poznan, Poland

account the multistage procedure of the maps' development including many different elements (e.g., Bates et al. 2014; Teng et al. 2017). Generally, the uncertainty in flood inundation modeling can be categorized into seven major types: (1) topographic data, (2) hydrologic data, (3) data for preliminary estimation of roughness, (4) method applied for final roughness calibration, (5) method applied for estimation of maximum flows, (6) structure of the applied model and (7) transition of the model results to the maps (e.g., Refsgaard and Storm 1990; Cook and Merwade 2009; Calenda et al. 2009; Liu and Merwade 2018). Some of the above-listed elements become less uncertain due to the development of the technology, e.g., increasing accuracy and resolution of LIDAR data for DEM elaboration (e.g., Gilles et al. 2012; Sampson et al. 2012; Walczak et al. 2013; Laks et al. 2017), involvement of satellite and remote sensing data (e.g., Jung et al. 2014; Arseni et al. 2017), elaboration of more detailed databases on flood events (e.g., Kundzewicz et al. 2017) and broader availability of more accurate hydraulic models (e.g., Szydłowski et al. 2013; Gašiorowski 2013; Brunner 2016a; Kolarski 2018). A specific problem is the uncertainty related to the channel and floodplain roughness (e.g., Dimitriadis et al. 2016; Pappenberger et al. 2008; Engeland et al. 2016; Liu and Merwade 2018). The lack of proper data for the initial assessment of these factors may be corrected by robust application of the calibration method. On the other hand, the time-consuming procedure of model parameters' identification may be reduced by relatively good quality data describing land cover with cover of the channel bed.

In the area of uncertainty assessment, one of the still unsolved problems is the estimation of maximum flows, also called design floods. The magnitudes of such design floods depend on the assumed return period uniquely linked to probability of exceedance. In fact, the maximum flows are the basis of the whole procedure for elaboration of flood hazard maps after the model calibration. The estimation of their magnitudes consists of a number of steps (e.g., Calenda et al. 2009). Each of these steps may be prone to potential uncertainty sources (e.g., Merz and Thielen 2005; Griffiths and Stedinger 2007; Laio et al. 2011). As it is reported by many researchers, the most important sources are (1) measurement errors and rating curve evaluation (Di Baldassarre and Montanari 2009; Di Baldassarre et al. 2010, 2012), (2) plotting position formula; (Hirsh and Stedinger 1987; Ewemoje and Ewemoje 2011), (3) assuming the randomness, stationarity, homogeneity, independence of the data (Serinaldi and Kilsby 2015; Serago and Vogel 2018), (4) choice of the observation period and data sampling (Calenda et al. 2009; Schendel and Thongwichian 2015, 2017), (5) probability distribution function (Calenda et al. 2009; Schendel and Thongwichian 2015, 2017; Yuan et al. 2017; Sun et al. 2017) and (6) method for estimation of parameters (Romanowicz and Beven 2003; Calenda et al. 2009; Beven

and Hall 2014; Schendel and Thongwichian 2015, 2017; Parkes and Demeritt 2016; Sun et al. 2017). In fact, some of the listed problems could be reduced with advances in the methodology applied. But two of them seem to be present independently of the methods developed. These are factors (4) and (6) in the above list. The first, choice of the observation period and data sampling, is related to the limited historical data available. This problem will not be overcome with new methods. The second, the method for estimation of parameters, is still a matter of subjective choice. Hence, the error made with such a decision is very important for the potential modeler.

In this study, we focus on the uncertainty related to the method applied for determination of design floods and the length of the data series. Of course, there are also other elements potentially important for the accuracy and reliability of flood hazard maps, e.g., digital elevation models (DEM), type of hydrodynamic simulation software. Taking into account achieved nowadays, quality of available data and applicability of the available methodologies, these two elements analyzed in our research seems to be the most crucial factors causing uncertainty in flood hazard assessment and management. The analysis was performed on the basis of the data collected in one selected gauge station. This is the Wronki gauge station on the Warta river. The station is located in the town of Wronki in the western part of Poland. The main purpose of the paper is to estimate the uncertainty of the maximum flow evaluation and its impact on the uncertainty of the flood hazard zone determination. The uncertainty of the maximum flow calculation is assessed with respect to factors mentioned earlier: the method for evaluation of the maximum flows and length of the data series. The uncertainty of the flood hazard zone is analyzed taking into account a very basic measure: inundation area.

Study site description

The Warta river is the third longest river in Poland. The length of the river is 808.2 km, and the total watershed area is 54,480 km². It is the greatest tributary of the Oder river (Fig. 1a). The total annual precipitation in the Warta watershed ranges from less than 500 mm in the central part to approximately 650 mm in upper part.

The mean discharge estimated for the whole period is 124.7 m³/s. The selected discharges are presented in Fig. 2. The presented light brown diamonds, dark blue line and red dots denote the annual minimum, mean and maximum flows, respectively, in Fig. 2a. In this part, the greatest floods are also marked. As one can see, the most dramatic flood occurred in 1979, although other flood events are shown in Fig. 2a in the years 1975, 1981, 2010 and 2011, with dangerous flooding generally occurring every 2–5 years. The

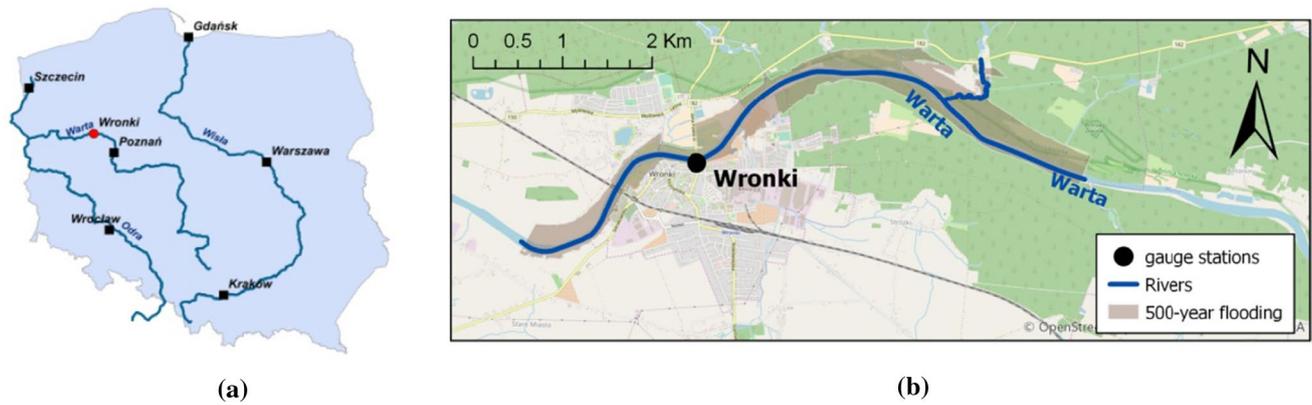


Fig. 1 Chosen case study: **a** location in Poland, **b** modeled river reach

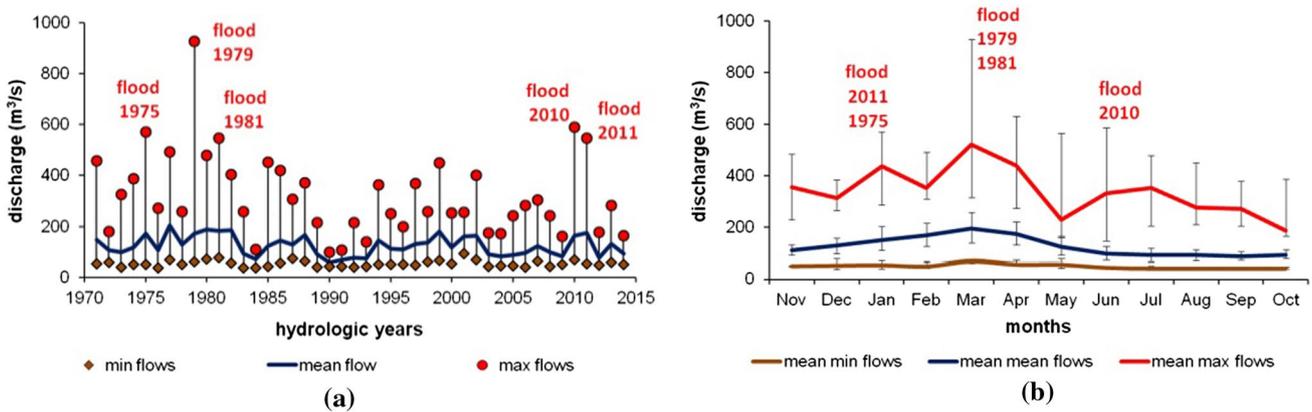


Fig. 2 Variability of discharge in the Wronki gauge station for the period 1971–2014: **a** annual min, mean and max flows, **b** monthly min, mean and max flows

second part, Fig. 2b, presents characteristics of the monthly flows. The brown, blue and red lines denote average minimum, mean and maximum flows observed in each month of the hydrologic year. The error bars show the variability of the minimum, mean and maximum monthly flows. The same floods are marked. As we can see, the greatest flood events occur in March, during the season change from winter to spring. However, the floods were also observed in other months.

There is a high risk of flooding in the Warta river valley. The flood hazard maps were determined for the whole Warta river (Fig. 1a) during implementation of the EU Flood Directive (European Parliament 2007). It was done in the frame of the ISOK project, mentioned earlier. One of the most interesting reaches along the Warta is the channel near the town of Wronki (Fig. 1b). The reach seen in Fig. 1b is about 7.6 km. The average width of the channel is 70 m there. Along this short reach, the width changes slightly. The depth is about 1.5–1.6 m. The selected reach is located in the lower part, along the last 68 km of the river course,

where the slope is only 0.13‰ (www.mdwe70.pl). The flood threat along this reach is serious. In Fig. 1b, the range of the flood hazard zone determined for a 500-year flood was showed. The existing infrastructure, the town of Wronki and two bridges, is prone to flooding. Additionally, new road investments are planned there. The third bridge is going to be built upstream of the town. In the town of Wronki, the gauge station is located (Fig. 1b).

Materials and methods

Data collection

In this study, four data sets were used: (1) hydrologic data for the Wronki gauge station, (2) a digital elevation model (DEM) for the area surrounding modeled reach, (3) the measurements of the channel cross sections and structures, (4) additional GIS layers supporting preparation of the model and visualization of results.

The hydrologic data for the Wronki gauge station were obtained from the Institute of Meteorology and Water Management (IMGW—Polish: *Instytut Meteorologii i Gospodarki Wodnej*). The location of the gauge station is presented in Fig. 3a, b as a black dot. The data are daily observed water stages and estimated discharges for the period 1971–2014. The homogeneity of the hydrologic data was tested with the Mann–Kendall test recommended by Banasik et al. (2017). No significant trends were detected. The DEM applied in this research was obtained from the Head Office of Geodesy and Cartography (GUGiK—Polish *Główny Urząd Geodezji i Kartografii*). Its spatial resolution is $1\text{ m} \times 1\text{ m}$. The vertical accuracy equals 15 cm. The modeled reach of the Warta river is presented in Fig. 2b. The DEM represents the surrounding area and it is seen in Fig. 3b. It covers about 122 km^2 . These data are stored as an ESRI GRID file. The memory size is about 467 MB. The cross-section measurements were obtained from the National Board for Water Management (KZGW—Polish: *Krajowy Zarząd Gospodarki Wodnej*). They are shown in Fig. 3a as green dots. There are 11 cross sections measured along the modeled reach. The average distance between measurements is 756 m. The obtained set of measurements also includes hydro-structures. In the modeled reach, two bridges are present. They are denoted as red squares in Fig. 3a. The measurements of cross sections near bridges are spaced closer. Their average distance apart is 52 m. On the other hand, the maximum distance between cross sections is 1460 m. The measurement of cross sections was done with a GPS device. The data have the form of a vector layer including points. The average number of measurement points in a single cross section is 136. The table of attributes includes position (X, Y), elevation (Z), code of point type and code of the land cover. The processing of the data enables interpolation of the bed, e.g., (Dysarz 2018a) and preparation

of the model (Dysarz et al. 2015). The computational cross sections are presented in Fig. 3b.

The additional GIS layers include vector layers as river centerline, river banks and lines of the computational cross sections as well as rasters such as orthophotomap, topographic map and OpenStreetMap (OSM). The listed vector layers were created during the process of model preparation. The river centerline is based on the Map of Hydrological Division of Poland (MPHP—Polish: *Mapa Podziału Hydrograficznego Polski*) available from KZGW. The two mentioned raster layers are applied as WMTS (Web Map Tile Service) servers linked to Geoportal 2—a web application prepared for management of spatial data sets. The OSM is a layer shared by OpenStreetMap Foundation (OSMF) on the Internet. All these layers are used here for better visualization of the results.

Estimation of maximum flows

For the period 1971–2014, annual maximum flow was determined. The annual maximum flows were the basis for calculating design floods with return periods of 10, 100 and 500 years. Two methods are applied for estimation of the maximum flows. These are: (1) the method of quantiles and (2) the maximum likelihood method. The methods are chosen due to their wide applicability in Polish conditions. Both of them are recommended in Banasik et al. (2017) as elements of flood hazard elaboration in Poland. In both cases, the probability distribution is modeled with the Pearson III function. In the first case, the quantiles of the distribution for cumulative probabilities 10%, 50% and 90% are estimated on the basis of the historical data. Then, the Pearson III curve was fitted in such a way that quantiles from historical data equal those calculated from the theoretical formulae. In the second method, the parameters of the distribution are set in such a way that the likelihood is maximized. The likelihood

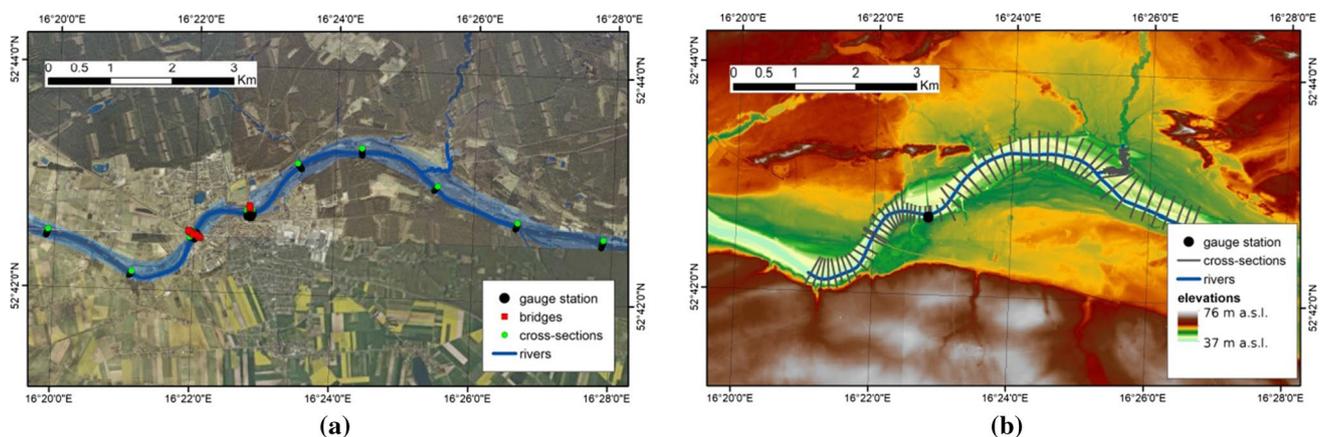


Fig. 3 Map of the chosen study case: **a** orthophotomap with inundation area for 100-year flood, **b** DEM and measured cross sections

is derived from the theoretical formula. Both methods are described in several hydrological books, e.g., Chow et al. (1988).

Modeling of river flow

The HEC-RAS is a well-known hydrodynamic model for rivers and water reservoirs. This program was designed at the Hydrologic Engineering Center (HEC). The second term in the name defines its application: River Analysis System (RAS). The concepts applied in the package are well described by Brunner (2016a). The HEC-RAS is applied for simulation of flow and transport processes in river networks, including floodplains and reservoirs. The modeled flow conditions include steady and unsteady longitudinal flow. The first is based on the simple Bernoulli equation. For description of the second model, the numerical solution of the St. Venant system of equations is implemented. There is also a module enabling the so-called quasi-unsteady flow simulation—simplified flow simulation in unsteady conditions on the basis of the Bernoulli equation. In the last versions of the package, the 2D flow module is also available. The HEC-RAS modules for simulation of transport processes include different transport of solutes, heat and sediments with deposition and erosion. The HEC-RAS package also includes several useful tools for data preparation and results processing. These tools include the module for GIS data processing called RAS Mapper (Brunner 2016b).

The model applied here is prepared for the reach of the Warta river near the town of Wronki. The length of the reach is about 7.6 km. In the model, there are 26 cross sections and 2 bridges defined. The average distance between computational cross sections is about 130 m (Fig. 3b). However, this distance is much more shorter in the case of cross sections located upstream and downstream of the bridges. It is a little bit more than 60 m in such a case. There is also a tributary in the model. It is a short reach of the Smolnica river (Fig. 3b). The length of this reach is about 1050 m. The role of the Smolnica river is not huge, but in the valley and floodplains of this river, a backwater may occur during flooding in the Warta river.

The maximum flows estimated for the Wronki gauge station are the basis for the calibration process. The values of the discharges are $560 \text{ m}^3 \text{ s}^{-1}$, $928 \text{ m}^3 \text{ s}^{-1}$ and $1211 \text{ m}^3 \text{ s}^{-1}$ for 10-, 100- and 500-year floods, respectively. The flood hazard maps are used to read the expected water surface elevations determined for each flood along the reach. The water surface elevations in the reach outlet are used as downstream boundary conditions. Although the changes of the discharge would also induce the changes in the water stage in the downstream cross section, there is a lack of data precisely describing this relationship. It is worth to remember that the water level in particular

cross section may be influenced by hydraulic conditions occurring downstream of this location, e.g., local weirs, flow contractions, etc. Because we focused on the selected reach, this problem is beyond the scope of our research. Hence, approach assuming consistency with other ISOK results is implemented. It is expected that the impact of the downstream boundary condition is not huge, though the problem was not carefully analyzed.

The rest of the water levels are applied as reference values in the process of model calibration. The implemented method of calibration is based on the standard trial-and-error procedure. However, the marching nature of the algorithm for water surface calculation is used. It let to search the optimal roughness values from downstream to upstream, what simplifies the whole process. This procedure is described by Dysarz et al. (2015). One of the relatively poorly known elements of HEC-RAS is HEC-RAS Controller (Goodell 2014). It enables access to the HEC-RAS elements, because it is compiled as the Component Object Module (COM). Hence, the HEC-RAS Controller is an application programming interface (API). Any program able to read the COM library may be used to control HEC-RAS computations. In this research, the HEC-RAS Controller is used to set the discharges for a simulation, to run a simulation, to read the results and store them in DBF tables.

Tools of spatial analysis

The ArcGIS 10.5.1. software developed by Esri company is applied in this research (e.g., Docan 2016; Law and Collins 2018). It enables quite easy processing of GIS data such as vector and raster layers. An integral part of ArcGIS is the ArcToolbox. It is a module including the main external tools and methods. Some of them are concurrent to methods available in the basic ArcGIS interface. Others extend the capabilities of the standard interface. Extension of the ArcGIS is possible also with specific plug-ins installed as ArcGIS toolbars. One such plug-in is HEC-GeoRAS (Cameron and Ackerman 2012). It is a toolbar with methods designed to support preparation of the river flow model.

The most important are tools applied for the generation of flood hazard maps including range of inundation, maps of depths and maps of water surface elevations. These are mainly spatial interpolation and map algebra methods. The natural neighbor method (e.g., Sibson 1981; Van der Graaf 2016) is applied for modeling of water surface spatial distribution. The conditional method for rasters and reclassification is applied to properly process the interpolation results. The data for interpolation are read from DBF tables, where the simulation results were stored before.

Automation with Python scripts

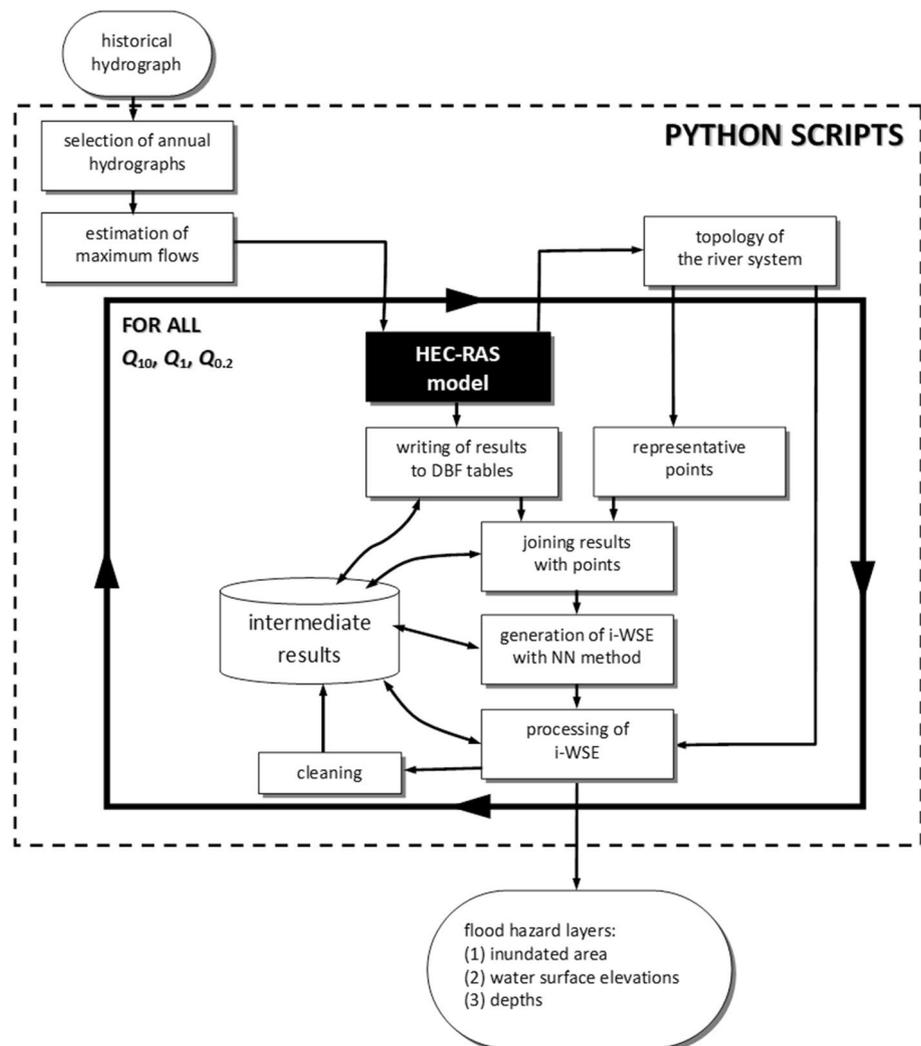
Python is one of the most popular programming languages today and it is still under development. In this paper, version 2.7.12 is used. Its usefulness in many areas has been reported. This scripting language is relatively simple for the beginner, but it is also very powerful if applied by an experienced coder. A description of this language may be found in many books or on internet websites, e.g., Downey et al. (2002), Python Software Foundation (2017a). The Python capabilities may be extended with specific modules dedicated to particular problems. The import of a module or single function is very simple. For the purposes of this research, two modules are used: (1) Pywin32 and (2) ArcPy. The methods available in the first module are applied to access COM (Component Object Module) objects and COM servers in Python code (Hammond and Robinson 2000; PythonCOM Documentation Index 2017). Brief descriptions of this package may be found in PythonCOM Documentation Index

(2017). The application of Python language with Pywin32 library is described in Dysarz (2018b). The second module, ArcPy, is fully integrated with ArcGIS (Zandbergen 2013). It enables access to all ArcToolbox methods. There are also mechanisms enabling access to objects of the vector layers such as points, lines and polygons.

Generation of flood hazard zones

The computations are configured in several steps shown in Fig. 4 and described below. The first is random choice of the annual discharge hydrographs for the estimation of maximum flows. The hydrographs are chosen from available data, namely years 1971–2014. The number of selected hydrographs depends on the length of the hydrological series assumed. There are four series assumed, namely 44 years, 40 years, 30 years and 20 years. Although the series shorter than 30 years are not recommended by Banasik et al. (2017), the shortest data scenarios are presented here to indicate what

Fig. 4 Scheme of the data and results processing



kind of uncertainty should be expected if this recommendation is ignored. Because the total number of years in this data set is 44, there is only one scenario tested for 44 years. The maximum flows estimated for this scenario are used as reference for other tests. For the rest of the tested series, the numbers of scenarios are given in Table 1. The numbers of tested scenarios were chosen arbitrarily taking into account the practical applicability of the presented methodology. In real cases, the number of simulations has to be controlled due to the time limitations of project elaboration. As can be seen, the number of scenarios is inversely proportional to the length of the series. When the scenarios are selected, the maximum flows are estimated for each scenario. The quantiles and maximum likelihood methods described above are used for this purpose. The maximum flows used further are Q_{10} , Q_1 and $Q_{0.2}$ which correspond to 10-, 100- and 500-year floods.

The preliminary step of the simulations is reading of the river network topology from the model. Python scripting with the access to HEC-RAS Controller is applied for this purpose. The river centerline, the cross-section cut lines and bank lines are read and transformed into shapefiles. The last step is done with GIS functions available in the ArcPy module. From these layers, the points of results' location are generated. These are starting and ending vertexes of the cross sections as well as cutting points of cross sections with river centerline and banks. Finally, there are 5 representative points for each cross section.

After simulation, the results saved in DBF tables are joined with representative points generated earlier. A new point layer is generated for each simulation result. The points with water surface elevation values recorded are used for interpolation. The Natural Neighbor method from the Spatial Analyst extension of ArcGIS is used to generate the surface of spatially distributed water elevation. Then, the generated surfaces are processed for each simulation result. In the first step, the raster of differences between interpolated water surface elevations (i-WSE) and digital elevation model (DEM) is calculated. This is the raster of pseudo-depth including negative and positive values. The negative values are removed by simple reclassification, and the positive values are replaced with a single value, e.g., unity. Such a raster is transformed into polygons. In the next step, the polygonal objects are selected. The polygons intersecting with the river centerline are considered as inundated area. The inundated area is used as mask for the final processing of the water surface map and depth map generated after extraction of the previously computed difference between i-WSE and DEM. During this process, a number of

intermediate vector and raster layers are created. These are storage consuming processed data. At the end of the processing, the memory is cleaned. The final results for any simulation are three layers: (1) polygons of inundated area, (2) the raster of depths and (3) the raster of water surface elevations.

Applied methods for assessment of uncertainty

Two elements are of concern in the present research. These are the method applied for estimation of the maximum discharges and the length of the data series. The impact of these two elements on the obtained values of the discharges and the range of the inundation area are analyzed. The analysis is based on direct comparison of obtained results presented in the form of graphs or probabilistic maps.

The first results are analyzed as mean values obtained for particular series of data and deviations from these values. The values are maximum discharges and inundation areas for three flood events, namely 10-year, 100-year and 500-year floods. Such comparisons show the potential stability of the mean values. They also present the potential range of variability if the data series of a particular length is taken as the basis of the analysis.

The final results of uncertainty assessment are based on calculation of deviations from reference values. In any case, the reference values are those measures which are calculated for the total length of data series equaling 44 years. The deviations from reference are calculated for each of the analyzed flood events mentioned above. The final results are presented as the total range of variability, which seems to be the most suitable variable in the case of relatively small number of trials. If the number of trials is increased, the range is never smaller. It might be only greater. Hence, the range of variability in the case of a relatively small number of trials should be treated as the minimum, which may be only greater.

Results and discussion

Selected flood hazard maps

There are 25 data scenarios tested (Table 1). Taking into account the method and the scenario, the maximum flows are determined 39 times. The quantiles method was tested with all lengths of the data series. The maximum likelihood method was not tested with the 20-year scenario. The results of each maximum flow estimation are three discharges, namely the 10-year flood, 100-year flood and 500-year flood. Each discharge is used for hydraulic simulation, which gives 117 calculations of water surface profiles. For each profile, three layers are processed: the inundation zone, the water surface elevations and spatial distribution

Table 1 Numbers of scenarios tested for each series and method

Length of series (years)	44	40	30	20
Quantiles method	1	5	8	11
Maximum likelihood method	1	5	8	–

of depth. The first is a vector layer of polygon features. The last two are raster layers. The total number of processed layers is 351. An example is shown in Fig. 5. The presented maps cover only the area around the town of Wronki. These results were obtained with the quintile method and randomly chosen 20-year data series. Part (a) of Fig. 5 presents the inundation zone for the 10-year flood. The background is an orthophotomap of this area. The second map (Fig. 5b) shows the spatial distribution of the depth for the same flood. It is presented on the background of DEM. The gauge station is marked in both maps.

The obtained maps are then compared and analyzed. The results of these analyses are presented below.

Impact of observation period for each method

In Figs. 6 and 7, the impact of the data series length on the obtained discharges and inundation areas is presented. These analyses included three maximum flows: 10-year

(blue color), 100-year (red) and 500-year (green). In both figures, the continuous lines represent mean values obtained for particular series of tests characterized by length of the data series and method chosen for the estimation of maximum discharges. The dots denote particular results. The dashed lines mark the range of variability. In the graphs, the range of uncertainty is written as numbers expressed as a percentage of the mean value determined for a particular length of the data series.

In Fig. 6, the results for the quantiles method are presented. It may be seen that the uncertainty of discharge evaluation increases as the length of the data series is shortened (Fig. 6a). In the case of the shortest series, the uncertainty dramatically increases with the magnitude of the flow. Although the uncertainty also seems to increase with the magnitude of the flow in the cases of longer series, 30 and 40 years long, this effect is not so significant as it is in the case of 20-year-long series. In Fig. 6b, the uncertainty of the inundation area is presented. The dependence of the

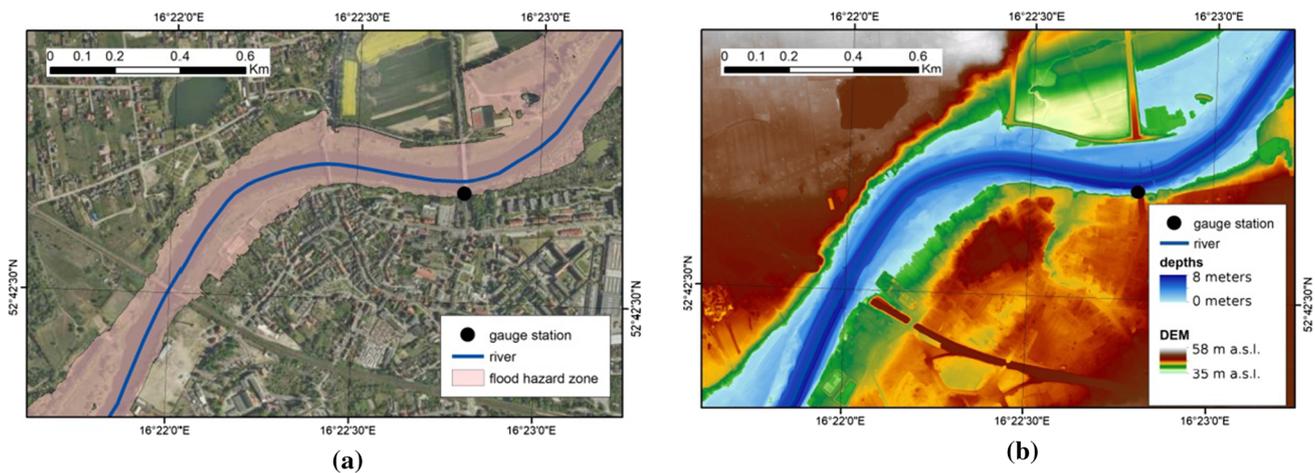


Fig. 5 Examples of flood hazard maps—quantiles method, 10-year flood: **a** inundation zone, **b** map of depths

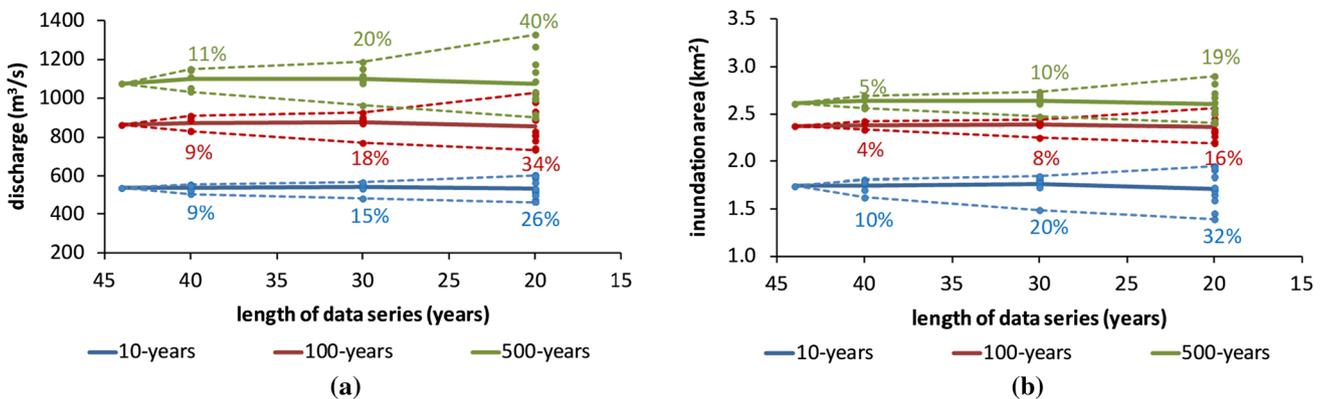


Fig. 6 Impact of the observation period for quantiles method: **a** variability of extreme flows, **b** variability of inundation area

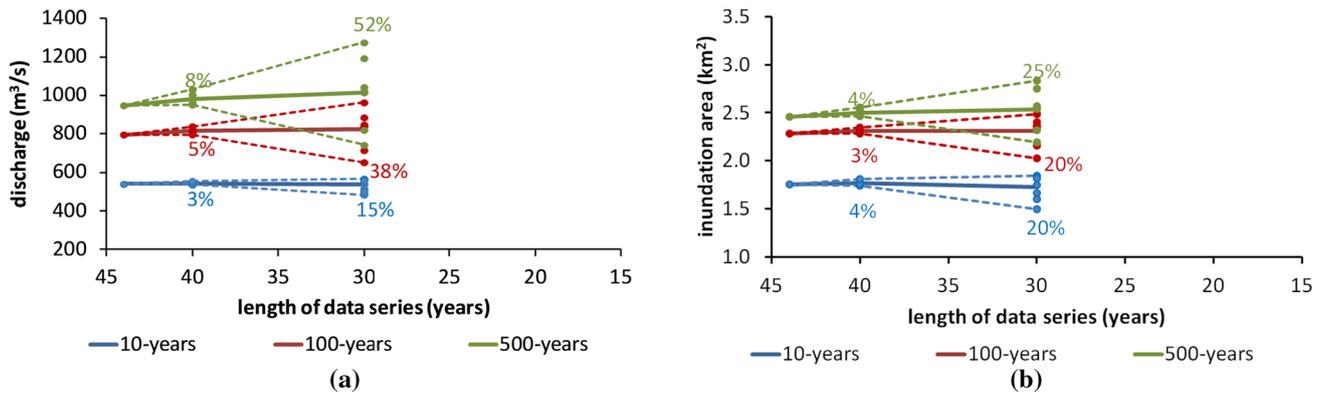


Fig. 7 Impact of the observation period for maximum likelihood method: **a** variability of extreme flows, **b** variability of inundation area

uncertainty on length of the data series is still visible. However, the relationship between the magnitude of the flow and the length of the series is changed. The greatest uncertainties are observed in the case of the lowest flows, 10-year floods. The impact of the terrain topography on the inundation area is the most important in the case of lower discharges.

In Fig. 7, the results obtained with the maximum likelihood method are presented. In this case, only the 40 and 30 year-series are tested, because the maximum flows were not determined for shorter series. It may be seen that this method gives smaller uncertainty of discharges for 40-year-long series (Fig. 7a). However, the uncertainty “jumps” to greater values than was observed previously as the length of the series is shortened to 30 years. The uncertainty of the inundation area is also smaller for longer series. For shorter series, the results are not so unique. It gives a greater range of variability for higher flows, 100-year flood and 500-year flood. At the same time the uncertainty for the lowers one, the 10-year flood, is smaller. A decrease in uncertainty with the increase in the flow magnitude is not seen in these results.

In Fig. 8, the comparison of the methods for the 100-year flood is presented. In part (a) we may see the changes of maximum flows, while part (b) shows variability of the inundation area. Both methods are presented as seen in Figs. 6 and 7, but colored polygons representing ranges of variability are added to indicate the convergence and discrepancy areas. The horizontal shift of the polygons is caused by different values of maximum flows obtained from each method. The comparison presents well the previous findings. The uncertainty of the maximum likelihood methods is smaller for longer series, but it dramatically increases as the series is shortened.

Propagation of uncertainty

The next graphs presented in Fig. 9 show the dependence of inundation area uncertainty on the discharge uncertainty. The axes represent values of difference between the results of the particular computations and results obtained for the reference scenario of 44 years long. These differences are expressed as the percentage of change with respect to the

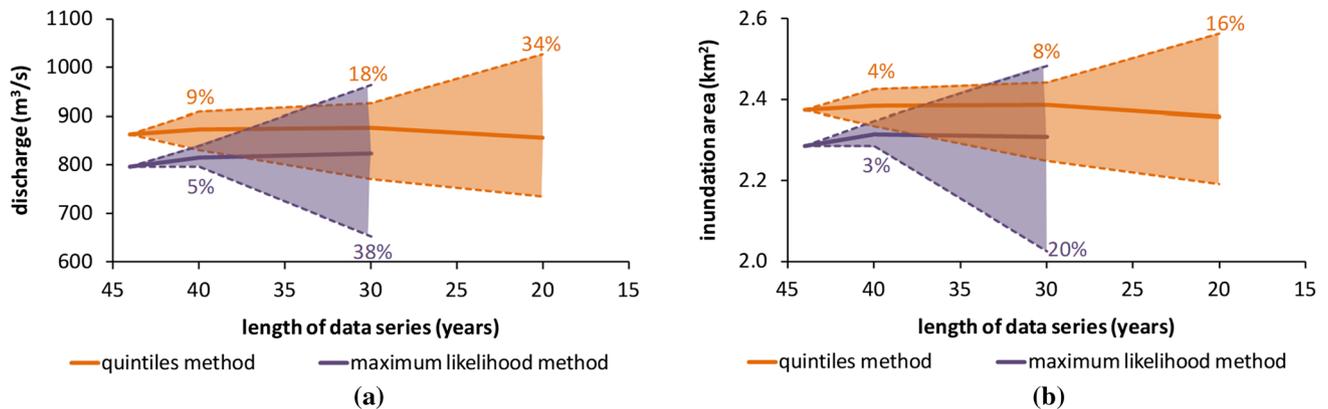


Fig. 8 Comparison of methods for 100-year flood: **a** variability of extreme flows, **b** variability of inundation area

reference results. The changes are positive and negative, which means that the obtained discharges and inundation areas could be greater or smaller than reference values. The gray dashed lines crossing at point (0, 0) represent the reference computations. The colored dots represent previous results. Figure 9a is composed from the results of the quantiles method, where Fig. 9b is prepared with the results of the maximum likelihood method.

The ranges of variability are marked with colored belts and proper values. In the case of the quantiles method (Fig. 9a), the 34% uncertainty of maximum flows obtained with the 20-year-long series gives 16% uncertainty in inundation areas. For the 30-year-long series, this relationships look as follows: 18% uncertainty of maximum flows gives 8% uncertainty of the inundation area. In the case of the 40-year-long series, the 9% uncertainty of maximum flows is related to 4% uncertainty of inundation areas. For the maximum likelihood method (Fig. 9b), the relations are as follows: 39% to 20% for 30-year-long series and 5% to 3% for 40-year-long series.

The results of analyzed tests are presented in Tables 2 and 3 in a slightly different form. In both tables, there are shown ranges of variability calculated as the percentage of

the reference values obtained for 44-year-long series of data. The first one presents the results of the quantiles method. In the second, the results obtained with the maximum likelihood method are shown. The tables include the ranges of variability of maximum flow calculation as well as the ranges of variability in determination of related inundation areas. All measures are estimated for each tested maximum flow, namely the 10-year flood, 100-year flood and 500-year flood. In the case of the quantiles method, the lengths of the data series are 40, 30 and 20 years. In the case of the maximum likelihood method, the tested lengths of data series are 40 and 30 years. The results of analysis of tendencies presented in Tables 2 and 3 are averaged over flood events as well as over the length of the data series. The first averages are shown in the last rows denoted as “EVENT AVERAGE.” The processed results of the second kind are placed in the last columns of each table denoted as “SERIES AVERAGE.”

The results presented in Tables 2, 3 show that the uncertainty measured in this way is a monotonically decreasing function of the length of the data series. For the case of maximum flows determined with quantiles method (Table 2), the range of variability increases from about 10% for a 40-year series to over 30% for a 20-year series. The range

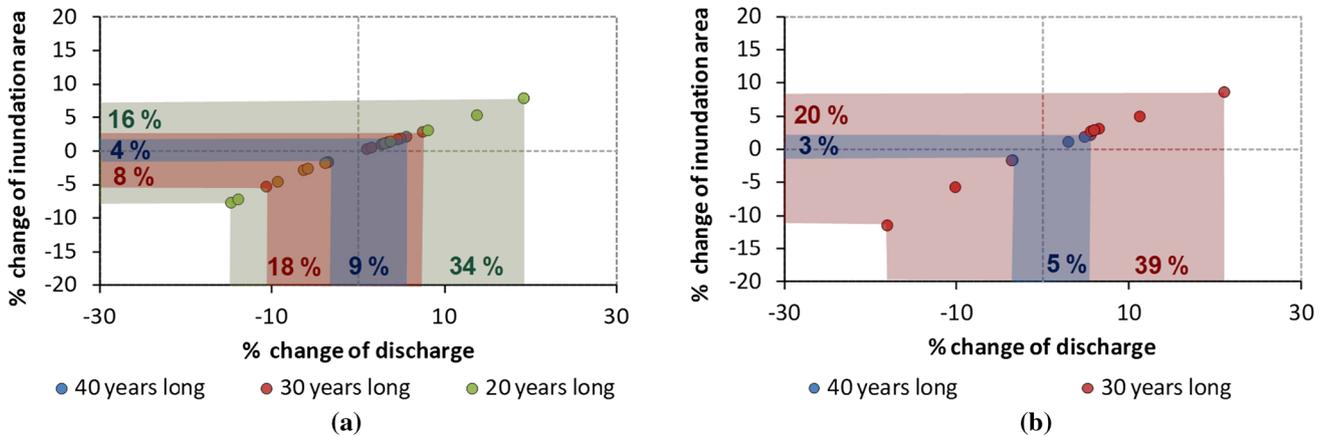


Fig. 9 Relation between uncertainty of discharge assessment and uncertainty of inundation area: **a** quantiles method, **b** maximum likelihood method

Table 2 Quantiles method: variability of max flows and related inundation areas with respect to length of the series

Length of series (years)	Range of variability (%)							
	10-Year flood		100-Year flood		500-Year flood		SERIES AVERAGE	
	Max flow	Inundation area	Max flow	Inundation area	Max flow	Inundation area	Max flow	Inundation area
40	8.92	10.47	9.15	3.82	10.84	5.13	9.64	6.47
30	15.53	20.13	18.11	8.18	20.81	9.97	18.15	12.76
20	26.03	31.80	33.89	15.60	39.83	18.99	33.25	22.13
EVENT AVERAGE	16.83	20.80	20.38	9.20	23.83	11.37		

Table 3 Maximum likelihood method: variability of max flows and related inundation areas with respect to length of the series

Length of series (years)	Range of variability (%)							
	10-Year flood		100-Year flood		500-Year flood		SERIES AVERAGE	
	Max flow	Inundation area	Max flow	Inundation area	Max flow	Inundation area	Max flow	Inundation area
40	3.39	3.76	5.43	2.75	9.03	4.14	5.95	3.55
30	15.37	19.89	39.10	20.06	56.17	26.05	36.88	22.00
EVENT AVERAGE	9.38	11.83	22.27	11.40	32.60	15.09		

of variability of the inundation area for the same method increases from almost 7% to over 20%. The results obtained with the second method, maximum likelihood (Table 3), show that the range of variability for maximum discharges increases from about 6% for the 40-year series to over 35% for the 30-year series. The range of the inundation areas for the same method varies from 4% to over 20%.

There are some interesting results reported in the scientific literature, which could be compared with the present research. Sun and co-authors (Sun et al. 2017) mentioned earlier analyzed only the uncertainty in determination of the maximum discharges. The study was conducted on two Chinese rivers with available observation periods of 70 years long. This study showed that the great increase in uncertainty is related to the increase in return period, which means the decrease in flood frequency. Sun and co-authors (Sun et al., 2017) analyzed six return periods ranging from 20 to 1000 years. In the first gauge station, Cuntan, they obtained uncertainty increasing with return period from 15 to 58%. In the second station, Pingshan, the uncertainty is also increasing with return period and varies in the range 11–45%.

In our example, this factor seems not to be so important. The results presented in Table 2, calculated for the quantiles method, show that the average uncertainty of maximum flow for 10-, 100- and 500-year events increases from 17 to 24%. In the case of the maximum likelihood method (Table 3), this uncertainty changes from 9 to 33%. In both cases, the uncertainty of discharges increases with the return period. If we focus on inundation area, the increase in uncertainty is not so obvious. In the cases of both methods, we can see that the uncertainty of a 100-year event may be lower than the uncertainty of a 10-year event. The only reason may be the special features of the local topography, e.g., local dikes, storage areas, etc. Such elements are present in the analyzed region.

Sun and co-authors (Sun et al. 2017) also noted that the method chosen for estimation of probability distribution is responsible for 3–8% of uncertainty. In the cases analyzed in this paper, the dependence of flow uncertainty on the method chosen is not so simple, because it also depends on the length of the data series. We may see that this uncertainty for the 40-year-long series is greater in the

case of the quantiles method (Table 2). It changes from 9 to 11%. In the case of the likelihood method (Table 3) and the same length of the data series, these values are in the range 3–9%. However, this tendency is inverted if the length of the data series is shortened. For the 30-year series, the quantiles method gives 15–21%, when the second method produces results with 15–57% uncertainty. This variation is lower if we compare the inundation area. For the 40-year series, we obtained uncertainty of 8–20% for the quantiles method (Table 2) and 20–26% for the maximum likelihood method (Table 3).

Merz and Thielen (2005) also presented interesting results for the Rhine River obtained with a 120-year observation series. They observed that the uncertainty of discharges calculated for the 100-year return period is about 11%. In our case, this uncertainty depends on the length of the series. For the quantiles method (Table 2), it varies from 9 to 34% for a series of 40–20 years. For the maximum likelihood method (Table 3), it varies from 5 to 39% for a series 40–30 years. What is interesting, the uncertainty of the inundation area is smaller. In the first example it varies in the range 4–16%, while in the second the range of variability is 3–20%.

As can be seen, the present results are compatible with those presented in the literature, but not exactly the same. The specific features of the local river system play an important role. Additionally, the transition of the discharge uncertainty to uncertainty of the inundation area is not straightforward. The local topography could disturb or even change the tendencies of uncertainty growth or decrease.

Examples of probabilistic maps

Selected probabilistic maps of inundation area are shown in Figs. 10 and 11. The maps are composed for tests with 100-year flood flow. The first figure is composed for the quantiles method, the second for the maximum likelihood method. In both cases, the total zone is presented on the left and the selected piece of the area is zoomed and shown on the right. The colors applied indicate the relative frequency of inundation. The blue and green are rarely flooded cells, while the yellow and red are frequently flooded areas. The

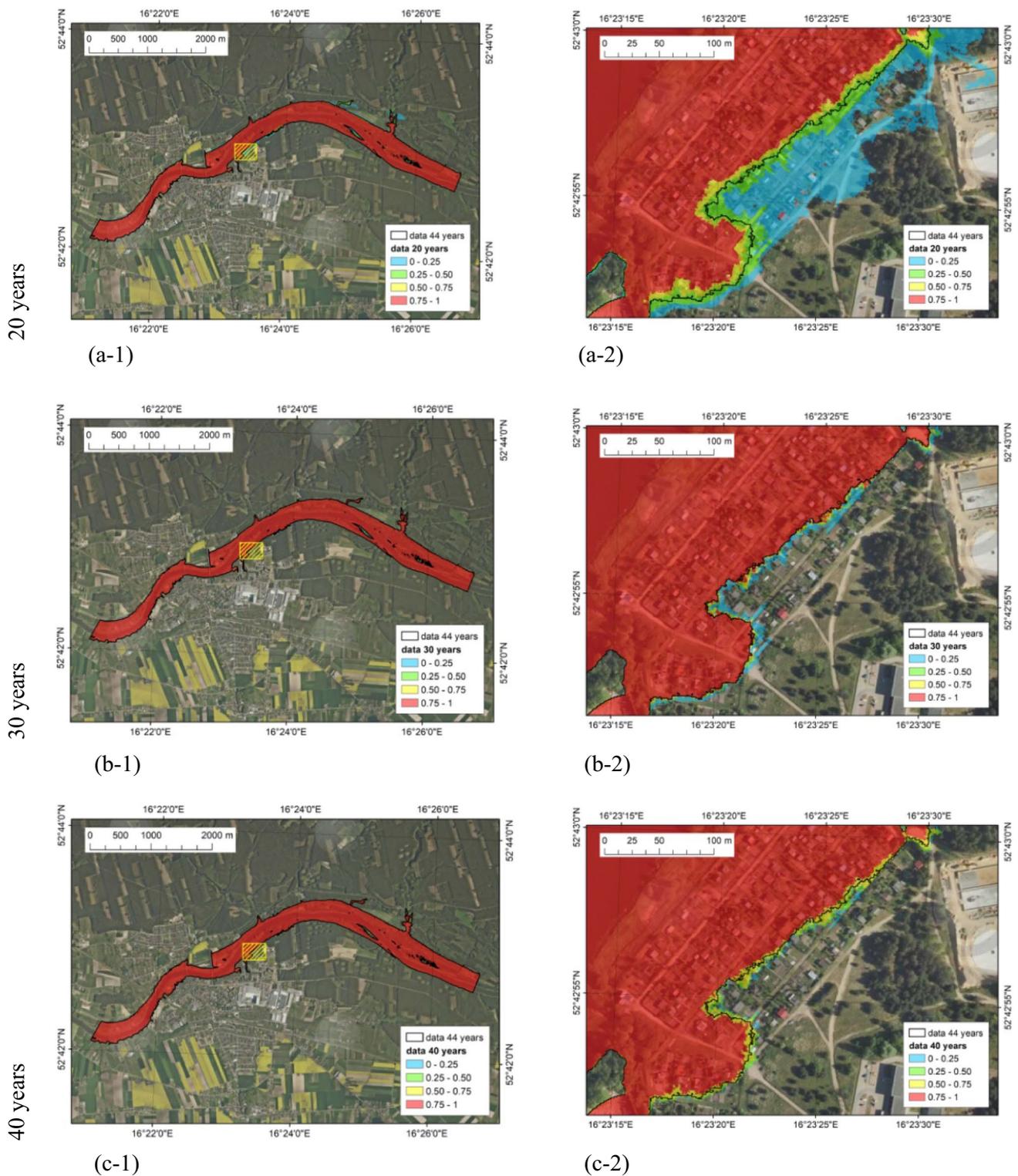


Fig. 10 Relative frequency of flood inundation for 100-year flood and quantiles method

background map is an orthophotomap, which allows us to see the difference of inundation in relation to the town of Wronki infrastructure, e.g., building and roads.

It is clearly seen as the area of probable inundation changes with the length of the time series. The biggest differences are seen for the 20-year-long series for the

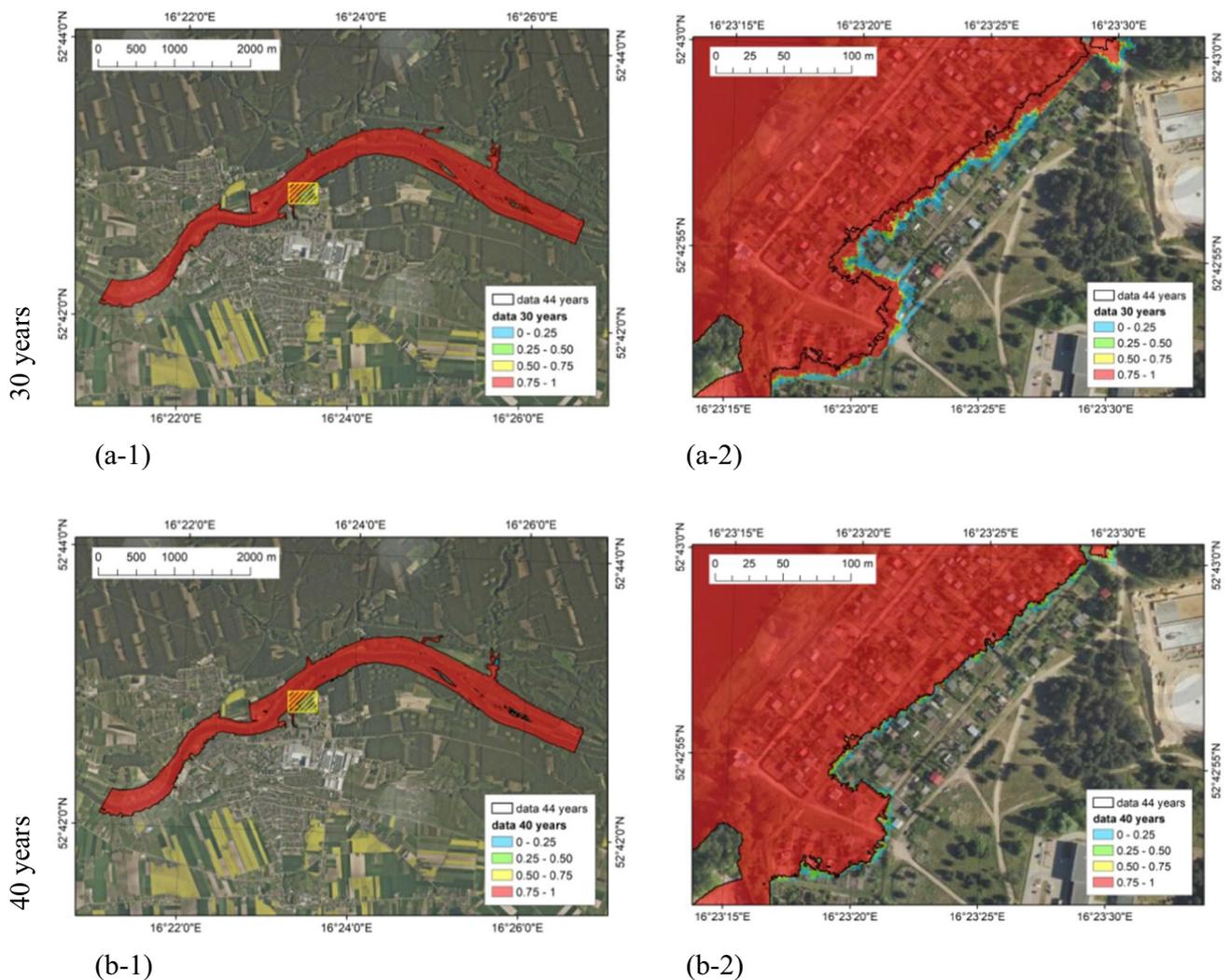


Fig. 11 Relative frequency of flood inundation for 100-year flood and maximum likelihood method

quantiles method (Fig. 10a-2). In this case, the inundation area may extend beyond the reference zone several dozens of meters in the shown region. As can be seen, the region is quite populated and such a difference significantly affects the risk of flooding. In the cases of longer data series (Fig. 10b-2, c-2), the differences in inundation area are smaller but their values vary from a few to over a dozen meters, which may be important in such urbanized areas as those shown in the quoted map.

In the case of the maximum likelihood method, the spread of the inundation area seems to be similar (Fig. 11a-2, b-2). The differences presented in Figs. 8 and 9 are not visible on this scale. However, it is clearly visible that the application of the maximum likelihood method does not protect against the uncertainty related to the length of the data series.

Conclusions

In this paper, the uncertainty assessment of the flood hazard analysis is presented. The chosen case study is the selected reach of the Warta river near the Wronki gauge station. The assessment is performed taking into account two main sources of uncertainty, namely the length of the data series and the method for evaluation of extreme flows. Hydraulic modeling and GIS processing is used to transform the calculated maximum flows into flood inundation zones, water surface and depth maps. The sophisticated scripting tools enabled effective processing of data and results. The obtained results are presented in informative graphs and maps.

The conducted research allowed us to formulate the following conclusions:

- The length of the annual maxima series has the greatest impact on the estimation of design flow. In the case of the quantiles method, the range of variability averaged over 10-, 100- and 500-year flood events changes from about 10% to over 33% with the length of the data series changing from 40 to 20 years. In the second case, the maximum likelihood method, this range varies from about 6% to slightly less than 37% with the tested length of data series 40 and 30 years.
- The impact of the method for estimation of distribution parameters is less important, but still significant. In the case of the quantiles method, the range of variability averaged over the length of the data series changes from about 17% to 24% with the return periods 10-, 100- and 500-years. The results obtained with the maximum likelihood method show variability from about 9% to less than 33% with the same return periods.
- Uncertainty associated with the calculation of design flows propagates into the size of flood hazard zones, but the main tendencies of variability increase or decrease are not kept during this transition. In both methods, the variability in determination of inundation area averaged over the length of the data series decreases with the return period when analyzed from the 10-year flood to the 100-year flood. But this tendency increases if analyzed from the 100-year flood to the 500-year flood.

In general, the presented examples indicate the role and importance of uncertainty in flood hazard assessment. The results proved that the inundation area may be overestimated or underestimated with insufficient data series. The choice of the proper method for evaluation of extreme flows is also crucial. Besides the problem of statistical parameters fitting in first-order regression, the resistance to uncertainties to the insufficient data series is also important. Although the main results present the inundation areas, it is supposed that the impact of these two factors analyzed is similar for the spatial distribution of the depths.

The challenge of the uncertainty assessment in determination of flood hazard maps is the number of necessary trials. In this research, only the minimum number of scenarios was tested. However, even this limited approach indicated a strong need for automation of the river flow modeling and spatial data processing. The response for this requirement is application of the scripting language for management of the whole computational process.

Taking into account the number of applied approaches, methods and data used in the implementation of the Flood Directive, it should be noted that flood hazard analyses in the different EU states are inconsistent. As the European integration covers new areas of economic and environmental management, the need for one uniform methodology in this area of technical activity becomes more and more

crucial. The need for development of robust methodology for uncertainty assessment is obvious. Although the number of uncertainty sources is huge in the case of flood modeling, the length of the data series and the method for evaluation of extreme flows are among the most important.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Albano R, Crăciun I, Mancusi L, Sole A, Ozunu A (2017) Flood damage assessment and uncertainty analysis: the case study of 2006 flood in Ilisua basin in Romania. *Carpathian J Earth Environ Sci* 12(2):335–346
- Arseni M, Roșu A, Bocăneală C, Constantin D-E, Georgescu LP (2017) Flood hazard monitoring using GIS and remote sensing observations. *Carpathian J Earth Environ Sci* 12(2):329–334
- Banasik K, Wałęga A, Weglarczyk S, Więzik B (2017) Update of methodology for calculation of maximum discharges and maximum precipitation with determined exceedance probability for controlled and uncontrolled watersheds and calibration of rainfall—runoff models (in Polish). Association of Polish Hydrologists, Warsaw. <http://www.kzgw.gov.pl/files/zam-pub/20170402-przegląd-i-aktualizacja-map/zal-2-do-OPZ-ujednolicony-1.pdf>. Accessed 22nd Dec 2018
- Bates PD, Pappenberger F, Romanowicz RJ (2014) Uncertainty in flood inundation modelling. In: Beven K, Hall J (eds) *Applied uncertainty analysis for flood risk management*. Imperial College Press, London, pp 232–269
- Beven KJ, Hall J (eds) (2014) *Applied uncertainty analysis for flood risk management*. Imperial College Press, London
- Brunner GW (2016a) HEC-RAS river analysis system hydraulic reference manual. US Army Corps of Engineers. Report No. CPD-69; Hydrologic Engineering Center (HEC), Davis
- Brunner GW (2016b) HEC-RAS river analysis system user's manual version 5.0. US Army Corps of Engineers. Report No. CPD-68; Hydrologic Engineering Center (HEC), Davis
- Calenda G, Mancini CP, Volpi E (2009) Selection of the probabilistic model of extreme floods: the case of the River Tiber in Rome. *J Hydrol* 371(1–4):1–11
- Cameron T, Ackerman PE (2012) HEC-GeoRAS GIS tools for support of HEC-RAS using ArcGIS user's manual. US Army Corps of Engineers, Institute for Water Resources, Hydrologic Engineering Center (HEC). http://www.hec.usace.army.mil/software/hec-geora/documentation/HEC-GeoRAS_43_Users_Manual.pdf. Accessed 7th July 2018
- Chow VT, Maidment DR, Mays LW (1988) *Applied hydrology*. McGraw-Hill Book Company, New Year

- Cook A, Merwade V (2009) Effect of topographic data, geometric configuration and modeling approach on flood inundation mapping. *J Hydrol* 377(1–2):131–142
- Di Baldassarre G, Montanari A (2009) Uncertainty in river discharge observations: a quantitative analysis. *Hydrol Earth Syst Sci* 13(6):913–921
- Di Baldassarre G, Montanari A, Lins H, Koutsoyiannis D, Brandimarte L, Blöschl G (2010) Flood fatalities in Africa: from diagnosis to mitigation. *Geophys Res Lett* 37(22):L22402
- Di Baldassarre G, Laio F, Montanari A (2012) Effect of observation errors on the uncertainty of design floods. *Phys Chem Earth Parts A/B/C* 42:85–90
- Dimitriadis P, Tegos A, Oikonomou A, Pagana V, Koukouvinos A, Mamassis N, Efstratiadis A (2016) Comparative evaluation of 1D and quasi-2D hydraulic models based on benchmark and real-world applications for uncertainty assessment in flood mapping. *J Hydrol* 534:478–492
- Docan DC (2016) Learning ArcGIS for desktop. Packt Publishing. <https://www.packtpub.com/application-development/learning-arcgis-desktop>. Accessed 7th July 2018
- Downey A, Elkner J, Meyers Ch (2002) How to think like a computer scientist. Learning with Python. Green Tea Press, Wellesley. <http://www.greenteapress.com/thinkpython/thinkCSpy.pdf>. Accessed 8th Nov 2017
- Dysarz T (2018a) Development of RiverBox—an ArcGIS toolbox for river bathymetry reconstruction. *Water* 10(9):1266
- Dysarz T (2018b) Application of Python scripting techniques for control and automation of HEC-RAS simulations. *Water* 10(10):1382
- Dysarz T, Wicher-Dysarz J, Sojka M (2015) Assessment of the impact of new investments on flood hazard—study case: the bridge on the Warta River near Wronki. *Water* 7:5752–5767. <http://www.mdpi.com/2073-4441/7/10/5752>. Accessed 23rd July 2018
- Engeland K, Steinsland I, Johansen SS, Petersen-Øverleir A, Kolberg S (2016) Effects of uncertainties in hydrological modelling. A case study of a mountainous catchment in Southern Norway. *J Hydrol* 536:147–160
- European Parliament (2007) Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks. <http://data.europa.eu/eli/dir/2007/60/oj>
- Ewemoje TA, Ewemoje OS (2011) Best distribution and plotting positions of daily maximum flood estimation at Ona River in Ogun-oshun river basin, Nigeria. *Agric Eng Int CIGR J* 13(3), Manuscript No. 1380
- Gąsiorowski D (2013) Analysis of floodplain inundation using 2D nonlinear diffusive wave equation solved with splitting technique. *Acta Geophys* 61(3):668–689
- Gilles D, Young N, Schroeder H, Piotrowski J, Chang YJ (2012) Inundation mapping initiatives of the Iowa Flood Center: statewide coverage and detailed urban flooding analysis. *Water* 4(1):85–106
- Goodell Ch (2014) Breaking HEC-RAS code. A user's guide to automating HEC-RA. h2ls, Portland
- Griffis VW, Stedinger JR (2007) The use of GLS regression in regional hydrologic analyses. *J Hydrol* 344(1–2):82–95
- Hammond M, Robinson A (2000) Python programming on Win32. O'Reilly Media Inc, Sebastopol
- Hirsh RM, Stedinger JR (1987) Plotting position for historical floods and their precision. *Water Resour Res* 23(4):715–727
- Jung Y, Kim D, Kim D, Kim M, Lee SO (2014) Simplified flood inundation mapping based on flood elevation-discharge rating curves using satellite images in gauged watersheds. *Water* 6:1280–1299
- Kolerski T (2018) Mathematical modeling of ice dynamics as a decision support tool in river engineering. *Water* 10(9):1241
- Kundzewicz ZW, Stoffel M, Wyżga B, Ruiz-Villanueva V, Niedźwiedź T, Kaczka R, Ballesteros-Cánovas JA, Pińskwar I, Łupikasza E, Zawiejska J, Mikuś P, Choryński A, Hajdukiewicz H, Spyt B, Janecka K (2017) Changes of flood risk on the northern foothills of the Tatra Mountains. *Acta Geophys* 65:799–807
- Laio F, Ganora D, Claps P, Galeati G (2011) Spatially smooth regional estimation of the flood frequency curve (with uncertainty). *J Hydrol* 408(1–2):67–77
- Laks I, Sojka M, Walczak Z, Wróżyński R (2017) Possibilities of using low quality digital elevation models of floodplains in hydraulic numerical models. *Water* 9(4):283
- Law M, Collins A (2018) Getting to know ArcGIS desktop, 5th edn. Esri Press, Redlands
- Liu Z, Merwade V (2018) Accounting for model structure, parameter and input forcing uncertainty in flood inundation modeling using Bayesian model averaging. *J Hydrol* 565:138–149
- Merz B, Thielen AH (2005) Separating natural and epistemic uncertainty in flood frequency analysis. *J Hydrol* 309(1–4):114–132
- Nones M (2015) Implementation of the floods directive in selected EU member states. *Water Environ J* 29(3):412–418
- Nones M (2017) Flood hazard maps in the European context. *Water Int.* 42(3):324–332
- Pappenberger F, Beven KJ, Ratto M, Matgen P (2008) Multi-method global sensitivity analysis of flood inundation models. *Adv Water Resour* 31(1):1–14
- Parkes B, Demeritt D (2016) Defining the hundred year flood: a Bayesian approach for using historic data to reduce uncertainty in flood frequency estimates. *J Hydrol* 540:1189–1208
- Python Software Foundation (2017a) Python 2.7.14 documentation. <https://docs.python.org/2/index.html>. Accessed 8th Nov 2017
- PythonCOM Documentation Index (2017) Python and COM. Blowing the rest away! <http://docs.activestate.com/activepython/2.4/pywin32/html/com/win32com/HTML/docindex.html>. Accessed 8 Nov 2017
- Refsgaard JC, Storm B (1990) Construction, calibration and validation of hydrological models. Distributed hydrological modeling. Springer, Dordrecht, pp 41–54
- Romanowicz R, Beven K (2003) Estimation of flood inundation probabilities as conditioned on event inundation maps. *Water Resour Res* 39(3):1073
- Sampson CC, Fewtrell TJ, Duncan A, Shaad K, Horritt MS, Bates PD (2012) Use of terrestrial laser scanning data to drive decimetric resolution urban inundation models. *Adv Water Resour* 41:1–17
- Schendel T, Thongwichian R (2015) Flood frequency analysis: confidence interval estimation by test inversion bootstrapping. *Adv Water Resour* 83:1–9
- Schendel T, Thongwichian R (2017) Considering historical flood events in flood frequency analysis: is it worth the effort? *Adv Water Resour* 105:144–153
- Serago JM, Vogel RM (2018) Parsimonious nonstationary flood frequency analysis. *Adv Water Resour* 112:1–16
- Serinaldi F, Kilsby CG (2015) Stationarity is undead: uncertainty dominates the distribution of extremes. *Adv Water Resour* 77:17–36
- Sibson R (1981) A brief description of nearest neighbor interpolation. Interpolating multivariate data. Wiley, New York
- Sun H, Jiang T, Jing C, Su B, Wang G (2017) Uncertainty analysis of hydrological return period estimation, taking the upper Yangtze River as an example. *Hydrol Earth Syst Sci Discuss* 1–26. <https://doi.org/10.5194/hess-2016-566>
- Szydłowski M, Szpakowski W, Zima P (2013) Numerical simulation of catastrophic flood: the case study of hypothetical failure of the Bielkowo hydro-power plant reservoir. *Acta Geophys* 61(5):1229–1245
- Teng J, Jakeman AJ, Vaze J, Croke BF, Dutta D, Kim S (2017) Flood inundation modelling: a review of methods, recent advances and uncertainty analysis. *Environ Model Softw* 90:201–216
- Van Alpen J, Passchier R (2007) Atlas of flood maps, examples from 19 European countries, USA and Japan. Ministry of Transport, Public Works and Water Management, The Hague

- Van der Graaf SC (2016) Natural Neighbour Kriging and its potential for quality mapping and grid design, M.Sc. thesis, Delft University of Technology. <https://repository.tudelft.nl/>. Accessed 7th July 2018
- Walczak Z, Sojka M, Laks I (2013) Assessment of mapping of embankments and control structure on digital elevation model based upon Majdany polder. *Rocz Ochr Śr* 15:2711–2724
- Yuan F, Zhao C, Jiang Y, Ren L, Shan H, Zhang L, Shen H (2017) Evaluation on uncertainty sources in projecting hydrological changes over the Xijiang River basin in South China. *J Hydrol* 554:434–450
- Zandbergen PA (2013) Python scripting for ArcGIS. Esri Press, Redlands



Dealing with sediment transport in flood risk management

Michael Nones¹

Received: 8 November 2018 / Accepted: 21 February 2019 / Published online: 4 March 2019
© The Author(s) 2019

Abstract

Flooding events are rising across European watercourses because of changing climatic conditions and anthropogenic pressure. To deal with these events, the European Floods Directive requires the development of flood risk management plans regularly updated every 6 years, where areas affected by flood risk and relative management strategies should be identified. Along the Directive, sediment transport and morphological changes in freshwater environments like rivers are only marginally considered, leading to a possibly wrong estimate of the impact of floods in the case of watercourses in which sediment transport represents a fundamental component. Using the Secchia River in Italy as a case study, the paper compares two numerical simulations performed with the freeware iRIC suite, imposing the same boundary conditions but comparing fixed and mobile bed. The obtained results pinpoint the importance of considering sediment transport in drawing flooding scenarios for alluvial sandy rivers. As suggested by this example, for the incoming revisions of the flood risk management plans, forecasted by 2021, water managers should account for the dynamic behaviour of surface watercourses, considering sediments not only as a driver of pollutants but also as a key aspect that shapes the environment and should be considered in modelling future scenarios and drawing associated management strategies.

Keywords Flood risk management plans · Floods directive · iRIC suite · Morphodynamics · Numerical model · Secchia River

Introduction

The quality of European rivers is one of the most important concerns for the future, as recognized, under a legislative point of view, by the Water Framework Directive (WFD) at the beginning of this century (EU 2000). Such Directive represents a new holistic, integrated approach to water protection, requiring, among other obligations, the classification of watercourses based on reference conditions. As described by the Directive, in fact, deviations from reference conditions are assessed by means of biological, hydromorphological and physicochemical quality elements, but only rivers classified in high status must achieve hydromorphological characteristics totally or nearly totally corresponding to undisturbed conditions when accounting for sediments (Nones et al. 2017). Thus, following the Annex V of the WFD and

the CIS Guidance n. 13 (CIS 2005), EU Water Authorities categorize water bodies as achieving good, moderate, poor or bad ecological status only based on biological monitoring results, neglecting the consideration of hydromorphological quality elements and sediment transport for stream which are not in a high status (Nardini et al. 2008). On the other side, however, sediments and associated transport processes are of great relevance due to their ecological (e.g. aquatic habitats, shelter area for fish), energy (e.g. reservoir sedimentation) and risk-related (floods and debris flows) effects (Habersack et al. 2017), given that sediment fluxes provide the basilar hydromorphological conditions to support dynamic aquatic ecosystems (Brils 2008; Fryirs and Brierley 2013; Wohl et al. 2015, Gurnell et al. 2016; Hajdukiewicz et al. 2017). As an example, the WFD mentions river continuity not only for biota but also for sediments, although the improvement of such continuity is still a challenging topic for water managers (Habersack et al. 2016).

Inadequate considerations of fluvial morphology and sediment transport have also important implications for the Directive 2007/60/EC on the Assessment and Management of Flood Risk (hereafter called Floods Directive-FD).

✉ Michael Nones
mnones@igf.edu.pl

¹ Department of Hydrology and Hydrodynamics, Institute of Geophysics, Polish Academy of Science, Ksiecia Janusza 64, 01-452 Warsaw, Poland

Developed accordingly with the WFD, this Directive aims to reduce and manage risks that floods pose to human health, environment, cultural heritage and economic activities (EU 2007), but giving open room to methods that can be applied to achieve these goals (Albano et al. 2017). According to the Directive, each member state must follow three steps during its implementation: (1) preliminary flood risk assessment, (2) flood hazard maps and flood risk maps and (3) flood risk management plan for each catchment (Nones 2015). The latter step is crucial, as it institutionalizes an ongoing paradigm shift from flood protection towards flood risk management (Bubeck et al. 2016; Roos et al. 2017). Although the Article 6.5d of the FD suggests drawing upon additional information regarding the impact of sediments and debris floods in the preparation of the flood maps (EU 2007), after the first implementation cycle ended in 2015, several shortcomings and weaknesses were evident, including a lack of theoretical risk management knowledge (Kallis and Butler 2001; EU 2015; Norén et al. 2016), inadequacy in considering hydromorphological alterations and sediments impact on flood risk management (Nones et al. 2017), incomplete understanding of cascading effects of floods (Pescaroli and Nones 2016). Such shortcomings need to be addressed in the short term because, by 2021, the next flood risk management plans (FRMP) should be produced, containing information on progress achieved towards the Directive goals (EU 2007).

Typically, alluvial channel is dynamic systems having erodible boundaries, self-adjusting to a variation of liquid and solid discharges supplied from upstream (see, among many others, Leopold and Maddock 1953; Schumm and Lichty 1965; Yalin 1992; Singh 2003; Nones and Di Silvio 2016; Slater 2016). Because a flood starts when water levels in the main channels are sufficient to exceed the bank height, locally flood risk is driven by changes in river channel stage, which may be impacted upon by variations in both flow magnitude and river channel conveyance (Lane et al. 2007; Neuhold et al. 2009), especially when considering a series of repeated floods instead of a single event (Guan et al. 2016). Geomorphological alterations and sediments moving in rivers impact the flood frequency and can increase it through the following: (1) reduction of the channel capacity (Slater et al. 2015); (2) morphological adjustments in response to a variable sediment supply from upstream (Lane and Thorne 2007); (3) bed aggradation and erosion due to damming and backwater effects (Walter and Merritts 2008; Maselli et al. 2018).

Researches about flood risk are typically associated with extreme hydrological events, assuming only clear water and non-erodible channels in implementing two-dimensional (2-D) hydraulic models to prepare FRMP (de Moel et al. 2009; Alfieri et al. 2014; Nied et al. 2017). However, geomorphic processes and anthropogenic alterations of the topography (Costabile and Macchione 2015) can mediate

and eventually increase the impacts of extreme happenings and backwater effects (Guan et al. 2016; Bohorquez and del Moral-Erencia 2017). Moreover, management practices can change the flow regime of a catchment, determining its geomorphological behaviour and its response to flooding events (O'Connell et al. 2005; Wheater and Evans 2009), especially along floodplains that are dynamical systems subjected to cyclical erosion and deposition phenomena.

River systems show a great spatial interactivity: flood defences built upstream can change the sedimentary load at the watershed scale, causing several problems for humans and infrastructures located downstream, such as increasing the siltation in hydropower reservoir or changing the river bathymetry, increasing the water level locally and consequently the flood risk (Merz et al. 2014; Liu et al. 2018). These interactions can be represented in numerical models for assessing future scenarios, but an engineering understanding of the main processes like long-term morphological changes is quite limited and relatively difficult to schematize (Sayers et al. 2002). Based on that, typically flood risk models do not account for sediment-related processes, driving to possible underestimations of flooded areas and associated depths (Nones and Pescaroli 2016), also because of the lack of detailed sedimentological data on both bedload and suspended load that should be used to accurately calibrate these modelling tools.

In the literature, less attention has been paid in the evaluation of the impact of in-channel morphology on flood risk and inundation extent (Sinnakaudan et al. 2003; Radice et al. 2016), although researches based on historical data and geomorphic analyses highlighted that an increase in the flood risk can be correlated not only with an increase in flow discharge, but also with a reduction in the channel conveyance (James 1999; Stover and Montgomery 2001). High sediment delivery ratio can be the necessary drivers for increasing flood risk, due to the reduction in the cross sections after sedimentation (Korup et al. 2004; Pinter and Heine 2005) or the alteration of the river pattern (Sinnakaudan et al. 2003). Recent works verified the effects of channel adjustments on the hydraulics of flood either by field observations (Rickenmann et al. 2015; Wyzga et al. 2015) or by using numerical modelling (Li et al. 2014; Guan et al. 2015; Staines and Carrivick 2015). These studies suggested that the influence of sediment transport and the subsequent morphological changes of the river can be a key driver of flood hazard.

Because of the complexity of hydro-morphodynamics models, nevertheless, before applying a specific code a thorough study of the water body is required. Indeed, historical geomorphic analyses can inform on selecting the most adequate modelling tools (e.g. geomorphic, hydraulic, morphodynamic), providing a better understanding of how river morphology, in planform and section, evolved. Such preliminary studies permit an estimation of the impact of

considering only clear water or also accounting for the sediment transport in computing the flood risk, helping the water managers in choosing an ad hoc monitoring and modelling approach (Brierley and Hooke 2015; Warner et al. 2018).

Trying to add additional insights to the ongoing debate about sediments and flood risk management in alluvial watercourses, the paper presents an application of the free-ware code iRIC to a small reach of the Secchia River, a lowland tributary of the Po River in Italy. The study is performed considering a single flooding event, highlighting the impact that sediment changes have locally and at the short time-scale. After the description of the study site and the adopted model, the results show that the sand transported along the watercourse deposits on the floodplains during low-flow conditions, reducing the cross section and increasing the likelihood of floods. Aside from the presentation of these specific results, missing points in the Floods Directive, like the consideration of sediment transport, are here discussed, suggesting the adoption of a broader vision in adopting flood risk management strategies.

Materials and methods

Study site and input data

The Secchia River is a right-hand tributary of the Po River and flows through the Emilia-Romagna Region in northern Italy (Fig. 1a). It is around 172 kilometres long and has a drainage basin with a catchment area of about 2300 km², with climatic conditions alternating between aridity in the dry summer months and higher flows during the wet periods, typically in spring and autumn. Being an important waterway since the ancient times, this watercourse has been diverted from the natural course to its present path during the XIII–XIV centuries, and now it flows through a highly anthropized environment, catching pollutants from the agricultural surroundings (Pescaroli and Nones 2018).

To emphasize the importance of sediment transport and deposition/erosion phases on the flood risk, the lowland area close to Ponte Motta is modelled with the free-ware code iRIC, using as geometrical input a digital terrain model (DTM) derived from LIDAR data and bathymetric surveys taken in 2015 (Fig. 1b). The domain is discretized using unstructured grids having cells between 10 m² close to the bridge to 100 m² elsewhere. Regarding the hydrological data, they refer to a flooding event happened in January 2014, having a return period of around 20 years, and discharges and water levels as upstream and downstream boundary conditions, respectively, are imposed. The simulated event has a total duration of 13 h, while the given hydraulic data are hourly, and spans between 250 and 290 m³/s at upstream, corresponding to a downstream level that rises from 8.5

to 9.4 m. The composition of the material transported and deposited along the floodplains is derived from site-specific field samples performed in 2016, corresponding to fine/medium sand.

iRIC model

The present study uses the solver Mflow_02 of the free-ware suite iRIC version 3.0 (<https://i-ric.org/en/>), which allows the user to reproduce the riverine hydro-morphodynamics using unstructured grids and finite element method, which can solve the 2-D vertically averaged unsteady flow and river morphodynamics equations (iRIC Software 2014; Nelson et al. 2016; Nones et al. 2018). By applying the typical characteristics of unstructured meshes, Mflow_02 computes the unsteady flow in areas including hydraulic structures like bridge piers and can model flooding events over complex areas such as floodplains. In addition, the solver allows for computing river bed variations as well as bedload and suspended load along the reach by means of several sediment transport equations.

Among many others, some characteristics of the code are: (1) the domain can be discretized using unstructured meshes having linear triangular elements; (2) the Galerkin finite element method is applied for solving the spatial derivatives; (3) the model is explicit in time; (4) the turbulence field is solved using a $k-\epsilon$ model; (5) the user can model suspended load and bedload separately; (6) the code accounts for a multiple grain size composition of the river bed and transported within the water column. The Galerkin finite element method permits the spatial discretization of the flow continuity and momentum equations, adopting linear triangular elements as shape functions to weight the residuals. The turbulence field is represented by the flow with large and small eddies, by means of the $k-\epsilon$ model (iRIC Software 2014). The kinematic eddy viscosity is directly computed, without solving the transport equations. The turbulent energy is evaluated as a function of the water depth (Nezu and Nakagawa 1993), while the approach suggested by Lomax et al. (2013) is applied for deriving the energy dissipation rate from the water velocity field. The bottom friction is set up using the Manning's roughness coefficient, here changing between the main channel and the floodplains, following a preliminary calibration (Pugliese 2017). To describe the morphodynamics, the model uses a mixed grain size deriving from a few sediment samples made along the Secchia River. Because there are no direct measurements of the sediment transport, it was divided into suspended load and bedload, both assumed in equilibrium with the water flow. This simplification involves a direct adjustment of the solid phase to variations of the liquid phase, without accounting for possible time lags between the two phases. For each grain size, the bedload is computed using the Meyer-Peter–Muller

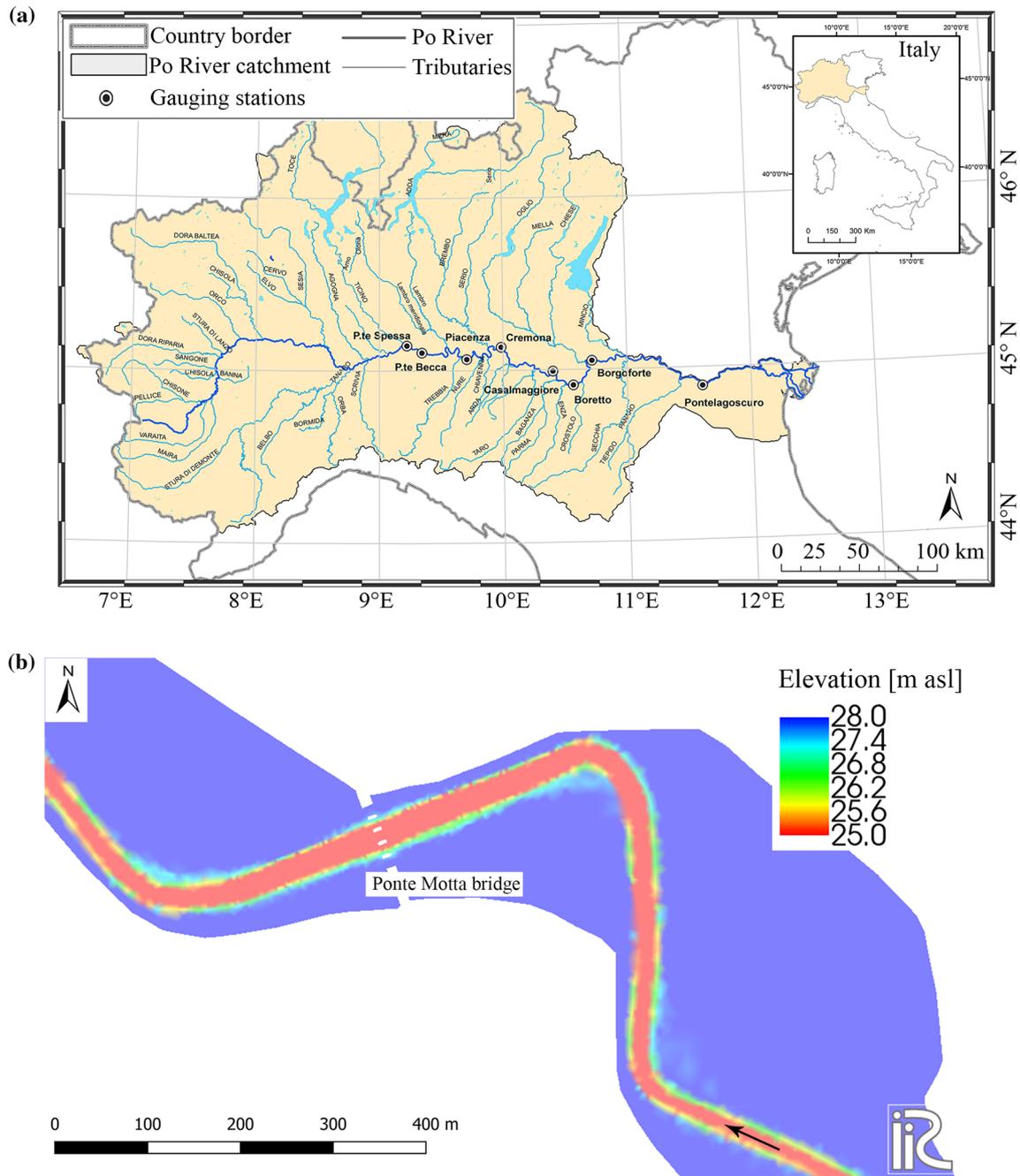


Fig. 1 **a** Map of the Po River basin with the indication of the Secchia River; **b** Digital Terrain Model implemented in the iRIC

formula, while the suspended load is calculated following the Einstein approach (Einstein 1950). This method permits to determine a representative particle diameter of the suspended sediment (van Rijn 1984) and, based on it, to compute the velocity of buoyancy of each grain size. The reference conditions for the concentration are calculated by means of the Rouse formula (Rouse 1937). Moreover, the mixed grain size composition is handled considering the shelter effect (Ashida and Michiue 1972) and assuring the

continuity for all the sediment classes. Additional information on the mathematical background is reported on the software website i-ric.org/en/.

Given that the main aim of the paper is to discuss the importance of considering sediment transport and related phenomena of erosion/deposition in flood risk modelling and associate drawing of FRMP rather than presenting the application of a flooding forecast model, no further details on the modelling structure or the input data are here reported. For

the sake of clarity, it is worth mentioning that the model was calibrated against historical flow data and compared with previous analyses (D'Alpaos et al. 2014). At the same time, sensitivity analyses of the used grid and time discretization were performed, comparing the simulations having a fixed bed with the measured data (Pugliese 2017).

However, given the lack of information about sediment transport, the present application is affected by some uncertainties, which are considered not affecting the overall behaviour.

Results and discussion

To evaluate the importance of sediment transport and river morphodynamics on the characteristics of flooded areas (extension and depth, and consequently on flow velocity), two sets of simulations were performed, using the same hydrological input but a changing bed: fixed and mobile. While the first set uses, as boundary conditions, only the geometry (DTM) and the hydraulics (i.e. discharge and water depth), the second one assumes that the river can mobilize the material from the river bed and floodplains and transport it along the main channel in equilibrium with the flow phase. Imposing a computational time step of 0.1 s for assuring the code stability, each simulation lasts around 4 h to simulate the 13 h of the event, with a few effects of computing the solid phase on the computational workload.

Fixed bed

The first simulation was performed assuming a fixed bed of the river (i.e. no consideration of sediment movements and erosion/deposition along the main channel and floodplains). Based on the boundary conditions described above, the iRIC Mflow_02 code permits to simulate the 2-D temporal evolution in terms of water depth and flux (specific discharge), providing the information for each cell.

As visible from Fig. 2a, following the geometry close to the bridge piers, the water depth reaches values higher than 8 m, as usual in the main channel of the Secchia River during flooding events. The flooding wave, however, has a minor impact along the floodplains upstream and downstream of the bridge, where the water flows very slowly, with a maximum depth of around 50–100 cm. A higher flow current within the main channel is confirmed by Fig. 2b, where the specific discharge (flux) is reported. One can notice that the higher flux is concentrated between the two central bridge piers, causing a higher water level. After overflowing the levees, the flow is practically standing along the floodplains.

Although the Emilia-Romagna Region is highly industrialized, these areas close to the watercourse are not yet inhabited but have still a very high economic value as pastures

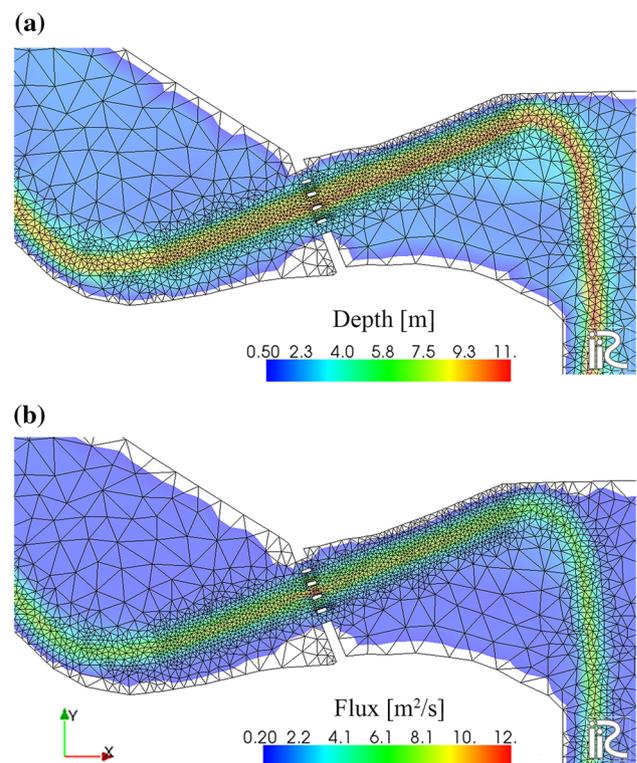


Fig. 2 Simulation performed assuming a fixed bed: **a** water depths and **b** water fluxes, at the end of the simulation

and agricultural field, being this region one of the most agricultural ones in Italy. Therefore, even flooding happenings characterized by a relatively low return period, like the one analysed here, can have a significant impact on the river surroundings with several consequences both on the environment and on the economy (Pescaroli and Nones 2018).

Mobile bed

Accounting for sediment transport and morphological variation during floods can lead to results quite different from considering only clear water flowing over a fixed bed, in particular in the case of active sandy or gravel river beds. In fact, as visible in Fig. 3, even a single flooding event, sufficiently relevant but not extreme, can change the bathymetry of many centimetres, slightly increasing the depth in the main channel due to erosion and the bed elevation of the floodplain following a deposition. With such conditions, at the end of the simulation, water depths over the flooded areas are reduced with respect to the previous simulation (Fig. 3a), even if the water flows over all the domain during the water peak and then stay there because of the increase in the river levees after sand deposition (Fig. 3c). The flux on the floodplains is slightly higher (Fig. 3b) and reaches regions not affected in the case of fixed-bed simulation and

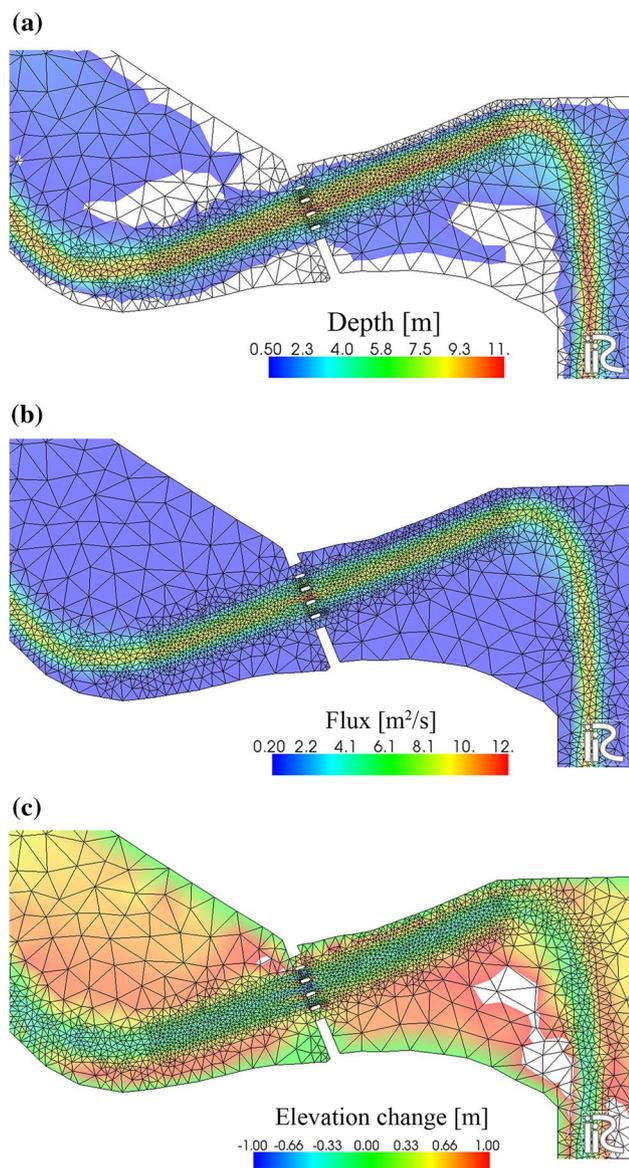


Fig. 3 Simulation performed imposing a mobile bed: **a** water depths, **b** water fluxes and **c** changes of the bed elevation

is therefore possibly not accounted for in drawing flood maps. Moreover, the overflow caused an increase in the elevation of the floodplain of about 50–70 cm, driving to a future knock-on effect.

Indeed, traditionally, sediment transport is considered as a kind of local alteration in time and space with respect to an equilibrium channel morphology, where cross-sectional areas adjust to a characteristic (bankfull) discharge. When sediments are delivered, it is thought to be a matter of time before river capacity re-establishes itself through transport of the solid component downstream along with the river system. However, as suggested by the present study and a few examples in the literature, this overlooks extensive

geomorphological evidence due to floods, potentially rapid over short time periods (Macklin and Lewin 2003). The aggradation of the floodplains can reduce the room for the river in the case of subsequent flooding happenings, potentially increasing the likelihood of future overflowing, especially downstream.

Geometrical changes in the Secchia River and the neighbour inundated areas depend upon both the magnitude of the flooding event and the type and quantity of sediments moved by the flow. In fact, considering fixed or movable conditions drive of significant differences in flooded area extents and associated water depths in the whole area, eventually affecting the resilience of the region interested and the management techniques that should be adopted to address this kind of happenings.

In addition, as visible, the transported material can be picked up from the bed, moved within the current and deposited along the floodplains, potentially altering the entire aquatic environment and affecting the WFD classification, which is based on aquatic flora and fauna (EU 2000), ultimately driving to inconsistency in defining the river status. Therefore, for the next river basin management plans, a coherent modelling of water–sediment interactions is also necessary and desirable.

However, as already mentioned, the present model was calibrated only against hydrological information (water levels measured at the Ponte Motta bridge), while it was not possible to accurately calibrate the sediment transport because of the availability of only a few bed samples but no distributed information about bedload and suspended load. Thanks to very recent flow field and concentration data captured with a horizontal ADCP (Acoustic Doppler Current Profiler) having a frequency of 600 kHz during the period November 2017–January 2018, it is planned to rerun the model calibrating both the liquid and the solid phases, aiming to provide more reliable quantitative information to local water managers and to confirm these preliminary outcomes.

Conclusions

This paper presents a first application of the freeware hydro-morphological model iRIC Mflow_02 on the Secchia River in Italy. Despite some uncertainties due to the lack of proper calibration data on the solid phase, the outcomes clearly pointed out the importance of considering sediment transport and morphological variations (erosion and deposition) in modelling areas affected by flood risk, aiming to produce adequate and reliable FRMP. While based on a single and relatively simple case study, the research presented here highlights important implications for flood risk assessment and river management, which could be valid more generally. As visible by comparing the two simulations made with

fixed and mobile bed conditions, indeed, sediments can have multiple effects, depending on the flood event type and its duration. As a matter of fact, even the modelling of a single flooding happening at the local scale can be affected by the presence of sediments, which are moved by water from the main channel and deposited along the floodplains, changing the channel capacity and therefore driving to a possible increase in the flood risk and the residence time of water on floodplains.

Usually, such considerations are not made by water managers, who simply follow the EU legislation and produce maps and plans in agreement with the WFD/FD requirements, using modelling tools that do not account for the presence of sediments as boundary conditions or drivers of large-scale changes. In many cases, neglecting sediment transport and bed variations cannot affect the characteristics of the flooded areas (water depth, extension, water velocity, etc.). However, in the case of very active rivers, like the lowland ones composed by fine sand that can be easily transported also by low water levels like the one presented here, but even in the case of mountainous gravel-bed streams, neglecting the effects of sediments can result in an underestimation of the flood risk, with important consequences under an economic and social point of view at manifold spatial and temporal scales.

For the next FRMP, which should be produced by 2021, water managers must adopt a broader vision, incorporating the effects of sediments and considering the watercourse that they should model as a dynamic system, affected by many drivers that can change its characteristics. Overcoming the representation of nature as a fixed system should be the best way to reduce the impact of flooding events, moving from post-event (emergency) intervention strategies to flood risks management policies. Ultimately, future studies should address the geomorphological processes in flood risk management in a proper way, using historic geomorphic analyses as a starting point and, based on them, adopting adequate tools (mostly already available) that can predict the medium-term (annual to decadal) response of river beds and banks to sediment delivery, aiming to assess the flood vulnerability. This should include considerations of those types of fluvial systems where aggradation issues are relevant, like the one analysed here.

Acknowledgements The author would like to thank all the colleagues involved in the INFRASAFE project for the fruitful discussions and the preparation works that contributed to the development of the ideas reported in the manuscript and, in particular, Alessio Pugliese. The author is very thankful to the reviewers for the helpful comments received.

Funding This study has been partially developed within the INFRASAFE project (www.infrasafe-project.com), financed by the Emilia-Romagna Region (Italy) under the Por-Fesr 2014–2020 Programme. Part of this work was also supported within statutory activities No.

3841/E-41/S/2018 of the Ministry of Science and Higher Education of Poland.

Compliance with ethical standards

Conflict of interest The author states that there is no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Albano A, Mancusi L, Abbate A (2017) Improving flood risk analysis for effectively supporting the implementation of flood risk management plans: the case study of “Serio” Valley. *Environ Sci Policy* 75:158–172. <https://doi.org/10.1016/j.envsci.2017.05.017>
- Alfieri L, Salamon P, Bianchi A, Neal J, Bates P, Feyen L (2014) Advances in pan-European flood hazard mapping. *Hydrol Process* 28(13):4067–4077. <https://doi.org/10.1002/hyp.9947>
- Ashida K, Michiue M (1972) Studies on bed load transportation for nonuniform sediment and river bed variation. *Disaster Prevention Research Institute Annuals, Kyoto University, No. 14B*, pp 259–273
- Bohorquez P, del Moral-Erencia JD (2017) 100 years of competition between reduction in channel capacity and streamflow during floods. *Remote Sens* 9(7):727. <https://doi.org/10.3390/rs9070727>
- Brierley G, Hooke J (2015) Emerging geomorphic approaches to guide river management practices. *Geomorphology* 215:1–5. <https://doi.org/10.1016/j.geomorph.2015.08.019>
- Brils J (2008) Sediment monitoring and the European Water Framework Directive. *Annali dell’Istituto Superiore di Sanita* 44(3):218–223
- Bubeck P, Aerts J CJH, de Moel H, Kreibich H (2016) Preface: flood-risk analysis and integrated management. *Nat Hazards Earth Syst Sci* 16:1005–1110. <https://doi.org/10.5194/nhess-16-1005-2016>
- CIS (2005) Common implementation strategy for the water framework directive (2000/60/EC). Guidance document n. 13: overall approach to the classification of ecological status and ecological potential
- Costabile P, Macchione F (2015) Enhancing river model set-up for 2-D dynamic flood modelling. *Environ Model Softw* 67:89–107. <https://doi.org/10.1016/j.envsoft.2015.01.009>
- D’Alpaos L, Brath A, Fioravante V, Gottardi G, Mignosa P, Orlandini S (2014) Relazione tecnico-scientifica sulle cause del collasso dell’argine del fiume Secchia avvenuto il giorno 19 gennaio 2014 presso la frazione San Matteo. (Technical-scientific report on the causes of the failure of the Secchia River embankment at the San Matteo village, happened on January 19, 2014) (in Italian)
- de Moel H, van Alphen J, Aerts J CJH (2009) Flood maps in Europe—methods, availability and use. *Nat Hazards Earth Syst Sci* 9:289–301. <https://doi.org/10.5194/nhess-9-289-2009>
- Einstein HA (1950) The bed-load function for sediment transportation in open channel flows. *USDA Soil Conserv Serv Tech Bull* 1026:1–71
- European Union (2000) Directive of the European Parliament and of the council 2000/60/EC establishing a framework for

- community action in the field of water policy. Official Journal, C513, 23.10.2000
- European Union (2007) Directive 2007/60/EC of the European Parliament and of the council of 23 October 2007 on the assessment and the management of flood risks. Official Journal, L288, 6.11.2007
- European Union (2015). EU overview of methodologies used in preparation of flood hazard and flood risk maps. Final report, September 2015
- Fryirs KA, Brierley GJ (2013) Geomorphic analysis of river systems: an approach to reading the landscape. Wiley, Chichester
- Guan M, Wright NG, Sleight PA, Carrivick JL (2015) Assessment of hydro-morphodynamic modelling and geomorphological impacts of a sediment-charged jökulhlaup, at Sólheimajökull, Iceland. *J Hydrol* 530:336–349. <https://doi.org/10.1016/j.jhydrol.2015.09.062>
- Guan M, Carrivick JL, Wright NG, Sleight PA, Staines KEH (2016) Quantifying the combined effects of multiple extreme floods on river channel geometry and on flood hazards. *J Hydrol* 538:256–268. <https://doi.org/10.1016/j.jhydrol.2016.04.004>
- Gurnell AM, Rinaldi M, Bellelli B, Bizzi S, Blamauer B, Braca G, Buijse AD, Bussetini M, Camenen B, Comiti F, Demarchi L, García de Jalón D, González del Tánago M, Grabowski RC, Gunn IDM, Habersack H, Hendriks D, Henshaw A, Klösch M, Lastoria B, Latapie A, Marcinkowski P, Martínez Fernández V, Mosselman E, Mountford JO, Nardi L, Okruszko T, O'Hare MT, Palma M, Percopo C, Surian N, van de Bund W, Weissteiner C, Ziliani L (2016) A multi-scale hierarchical framework for developing understanding of river behaviour to support river management. *Aquat Sci* 78(1):1–16. <https://doi.org/10.1007/s00027-015-0424-5>
- Habersack H, Hein T, Stanica A, Liska I, Mair R, Jäger E, Hauer C, Bradley C (2016) Challenges of river basin management: current status of, and prospects for, the River Danube from a river engineering perspective. *Sci Total Environ* 543(A):828–845. <https://doi.org/10.1016/j.scitotenv.2015.10.123>
- Habersack H, Liébault F, Comiti F (2017) Sediment dynamics in Alpine basins. *Geomorphology* 291:1–3. <https://doi.org/10.1016/j.geomorph.2017.05.014>
- Hajdukiewicz H, Wyzga B, Zawiejska J, Amirowicz A, Oglecki P, Radecki-Pawlik A (2017) Assessment of river hydromorphological quality for restoration purposes: an example of the application of RHQ method to a Polish Carpathian river. *Acta Geophys* 65(3):423–440. <https://doi.org/10.1007/s11600-017-0044-7>
- iRIC Software (2014) Mflow_02 Solver Manuals. iRIC Software. <https://i-ric.org/en/>
- James A (1999) Time and the persistence of alluvium: river engineering, fluvial geomorphology, and mining sediment in California. *Geomorphology* 31(1–4):265–290. [https://doi.org/10.1016/S0169-555X\(99\)00084-7](https://doi.org/10.1016/S0169-555X(99)00084-7)
- Kallis G, Butler D (2001) The EU Water Framework Directive: measures and Implications. *Water Policy* 3:125–142. [https://doi.org/10.1016/S1366-7017\(01\)00007-1](https://doi.org/10.1016/S1366-7017(01)00007-1)
- Korup O, McSaveney MJ, Davies TRH (2004) Sediment generation and delivery from large historic landslides in the Southern Alps, New Zealand. *Geomorphology* 61(1–2):189–207. <https://doi.org/10.1016/j.geomorph.2004.01.001>
- Lane SN, Thorne CR (2007) River processes. In: Thorne CR, Evans EP, Penning-Roswell EC (eds) Future flooding and coastal erosion risks. ICE Publishing, a division of Thomas Telford Ltd, the commercial arm of the Institution of Civil Engineers, pp 82–99. <https://doi.org/10.1680/ffacer.34495.0006>
- Lane SN, Tayefi V, Reid SC, Yu D, Hardy RJ (2007) Interactions between sediment delivery, channel change, climate change and flood risk in a temperate upland environment. *Earth Surf Proc Land* 32(3):429–446. <https://doi.org/10.1002/esp.1404>
- Leopold LB, Maddock T Jr. (1953) The hydraulic geometry of stream channels and some physiographic implications. U.S. Geological Survey, Washington, DC
- Li W, van Maren DS, Wang ZB, de Vriend HJ, Wu B (2014) Peak discharge increase in hyperconcentrated floods. *Adv Water Resour* 67:65–77. <https://doi.org/10.1016/j.advwatres.2014.02.007>
- Liu C, Walling DE, He Y (2018) Review: the International Sediment Initiative case studies of sediment problems in river basins and their management. *Int J Sedim Res* 33(2):216–219. <https://doi.org/10.1016/j.ijsrc.2017.05.005>
- Lomax H, Pulliam TH, Zingg DW (2013) Fundamentals of computational fluid dynamics. Springer, Berlin
- Macklin MG, Lewin J (2003) River sediments, great floods and centennial-scale Holocene climate change. *J Quat Sci* 18(2):101–105. <https://doi.org/10.1002/jqs.751>
- Maselli V, Pellegrini C, Del Bianco F, Mercorella A, Nones M, Crose L, Guerrero M, Nittrouer JA (2018) River morphodynamic evolution under dam-induced backwater: an example from the Po River (Italy). *J Sediment Res* 88(10):1190–1204. <https://doi.org/10.2110/jsr.2018.61>
- Merz B, Elmer F, Kunz M, Mühr B, Schröter K, Uhlemann-Elmer S (2014) The extreme flood in June 2013 in Germany. *La Houille Blanche* 1:5–10. <https://doi.org/10.1051/lhb/2014001>
- Nardini A, Sansoni G, Schipani I, Conte G, Goltara A, Boz B, Bizzi S, Polazzo A, Monaci M (2008) The water framework directive: a soap bubble? An integrative proposal: FLEA (Fluvial Ecosystem Assessment). Centro Italiano per la Riquilificazione Fluviale
- Nelson JM, Shimizu Y, Abe T, Asahi K, Gamou M, Inoue T, Iwasaki T, Kakinuma T, Kawamura S, Kimura I, Kyuka T, McDonald RR, Nabi M, Nakatsugawa M, Simoes FR, Takebayashi H, Watanabe Y (2016) The international river interface cooperative: public domain flow and morphodynamics software for education and applications. *Adv Water Resour* 93:62–74. <https://doi.org/10.1016/j.advwatres.2015.09.017>
- Neuhold C, Stanzel P, Nachtnebel HP (2009) Incorporating river morphological changes to flood risk assessment: uncertainties, methodology and application. *Nat Hazards Earth Syst Sci* 9:789–799. <https://doi.org/10.5194/nhess-9-789-2009>
- Nezu I, Nakagawa H (1993) Turbulence in open-channel flow. IAHR monograph. Belkema, Rotterdam
- Nied M, Schröter K, Lütke S, Nguyen VD, Merz B (2017) What are the hydro-meteorological controls on flood characteristics? *J Hydrol* 545:310–326. <https://doi.org/10.1016/j.jhydrol.2016.12.003>
- Nones M (2015) Implementation of the floods directive in selected EU Member States. *Water Environ J* 29(2):412–418. <https://doi.org/10.1111/wej.12247>
- Nones M, Di Silvio G (2016) Modeling of river width variations based on hydrological, morphological, and biological dynamics. *J Hydraul Eng* 142(7):04016012. [https://doi.org/10.1061/\(asce\)hy.1943-7900.0001135](https://doi.org/10.1061/(asce)hy.1943-7900.0001135)
- Nones M, Pescaroli G (2016) Implications of cascading effects for the EU floods directive. *Int J River Basin Manag* 14(2):195–204. <https://doi.org/10.1080/15715124.2016.1149074>
- Nones M, Gerstgraser C, Wharton G (2017) Consideration of hydro-morphology and sediment in the implementation of the EU water framework and floods directives: a comparative analysis of selected EU member states. *Water Environ J* 31(3):324–329. <https://doi.org/10.1111/wej.12247>
- Nones M, Pugliese A, Domeneghetti A, Guerrero M (2018) Po River morphodynamics modelled with the open-source code iRIC. In: Kalinowska M, Mrokowska M, Rowiński P (eds) Free surface flows and transport processes. GeoPlanet: earth and planetary sciences. Springer, Cham. https://doi.org/10.1007/978-3-319-70914-7_22

- Norén V, Hedelin B, Nyberg L, Bishop K (2016) Flood risk assessment—practices in flood prone Swedish municipalities. *Int J Disaster Risk Reduct* 18:206–217. <https://doi.org/10.1016/j.ijdrr.2016.07.003>
- O’Connell PE, Beven KJ, Carney JN, Clements RO, Ewen J, Fowler H, Harris GL, Hollis J, Morris J, O’Donnell GM, Packman JC, Parkin A, Quinn PF, Rose SC, Shepher M, Tellier S (2005) Review of impacts of rural land use and management on flood generation: impact study report. Department of the Environment, Farming and Rural Affairs R&D Technical Report, FD2114/TR
- Pescaroli G, Nones M (2016) Cascading events, technology and the floods directive: future challenges. In: Proceedings of the 3rd European conference on flood risk management, FLOODrisk 2016, Lyon, France. E3S Web Of Conference, 7, 07003. <https://doi.org/10.1051/e3sconf/20160707003>
- Pescaroli G, Nones M (2018) Integrating cascading effects of floods in policy making: a case study for Emilia Romagna, Italy. In: Proceedings of the 5th IAHR Europe Congress, Trento, Italy. https://doi.org/10.3850/978-981-11-2731-1_291-cd
- Pinter N, Heine RA (2005) Hydrodynamic and morphodynamic response to river engineering documented by fixed-discharge analysis, Lower Missouri River, USA. *J Hydrol* 302(1–4):70–91. <https://doi.org/10.1016/j.jhydrol.2004.06.039>
- Pugliese A (2017) Ponte Motta flooding event of January 2014: calibration and validation of an iRIC model. Personal communication
- Radice A, Longoni L, Papini M, Brambilla D, Ivanov VI (2016) Generation of a design flood-event scenario for a mountain river with intense sediment transport. *Water* 8(12):597. <https://doi.org/10.3390/w8120597>
- Rickenmann D, Badoux A, Hunzinger L (2015) Significance of sediment transport processes during piedmont floods: the 2005 flood events in Switzerland. *Earth Surf Proc Land* 41(2):224–230. <https://doi.org/10.1002/esp.3835>
- Roos MMD, Hartmann TT, Spit TTJM, Johann GG (2017) Constructing risks—internalisation of flood risks in the flood risk management plan. *Environ Sci Policy* 74:23–29. <https://doi.org/10.1016/j.envsci.2017.04.007>
- Rouse H (1937) Modern conception of the mechanics of turbulence. *Trans ASCE* 102:463–543
- Sayers PB, Hall JW, Meadowcroft IC (2002) Towards risk-based flood hazard management in the UK. *Proc Inst Civil Eng Civil Eng* 150(5):36–42
- Schumm SA, Lichty R (1965) Time, space, and causality in geomorphology. *Am J Sci* 263(2):110–119. <https://doi.org/10.2475/ajs.263.2.110>
- Singh VP (2003) On the theories of hydraulic geometry. *Int J Sedim Res* 18(3):196–218
- Sinnakaudan SK, Ab Ghani A, Ahnmad MSS, Zakaria NA (2003) Flood risk mapping for Pari River incorporating sediment transport. *Environ Model Softw* 18(2):119–130. [https://doi.org/10.1016/S1364-8152\(02\)00068-3](https://doi.org/10.1016/S1364-8152(02)00068-3)
- Slater LJ (2016) To what extent have changes in channel capacity contributed to flood hazard trends in England and Wales? *Earth Surf Proc Land* 41:1115–1128. <https://doi.org/10.1002/esp.3927>
- Slater LJ, Singer MB, Kirchner JW (2015) Hydrologic versus geomorphic drivers of trends in flood hazard. *Geophys Res Lett* 42(2):370–376. <https://doi.org/10.1002/2014GL062482>
- Staines KEH, Carrivick JL (2015) Geomorphological impact and morphodynamic effects on flow conveyance of the 1999 jökulhlaup at sólheimajökull, Iceland. *Earth Surf Process Landf* 40(10):1401–1416. <https://doi.org/10.1002/esp.3750>
- Stover SC, Montgomery DR (2001) Channel change and flooding, Skokomish River, Washington. *J Hydrol* 243(3–4):272–286. [https://doi.org/10.1016/S0022-1694\(00\)00421-2](https://doi.org/10.1016/S0022-1694(00)00421-2)
- van Rijn LC (1984) Sediment transport, part II: suspended load transport. *J Hydraul Eng* 110(11):1613–1641. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1984\)110:11\(1613\)](https://doi.org/10.1061/(ASCE)0733-9429(1984)110:11(1613))
- Walter RC, Merritts DJ (2008) Natural streams and the legacy of water-powered mills. *Science* 319(5861):299–304. <https://doi.org/10.1126/science.1151716>
- Warner BP, Gartner JD, Hatch CE (2018) Fluvial geomorphic assessment and river corridor mapping as flood risk management tools in Massachusetts, USA. *J Flood Risk Manag* 11(S2):S1100–S1104. <https://doi.org/10.1111/jfr3.12328>
- Wheater H, Evans E (2009) Land use, water management and future flood risk. *Land Use Policy* 265(1):S251–S264. <https://doi.org/10.1016/j.landusepol.2009.08.019>
- Wohl E, Bledsoe BP, Jacobson R, Poff NL, Rathburn SL, Walters DM, Wilcox AC (2015) The Natural Sediment Regime in Rivers: broadening the Foundation for Ecosystem Management. *Bioscience* 65(4):358–371. <https://doi.org/10.1093/biosci/biv002>
- Wyzga B, Zawiejska J, Radecki-Pawlik A (2015) Impact of channel incision on the hydraulics of flood flows: examples from Polish Carpathian rivers. *Geomorphology* 272:10–20. <https://doi.org/10.1016/j.geomorph.2015.05.017>
- Yalin MS (1992) River mechanics. Pergamon Press, Oxford



Investigations into solar flare effects using wavelet-based local intermittency measure

Sumesh Gopinath¹ · P. R. Prince¹

Received: 20 January 2018 / Accepted: 29 January 2019 / Published online: 7 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

The present study analyzes the efficiency of local intermittency measure based on wavelet transforms in identifying solar flare effects on magnetograms. If we observe the flare-time features in geomagnetic components, most often, disturbances associated with other solar phenomena will enhance or mask the solar flare signatures. Similarly, diurnal and high-latitude geomagnetic variabilities will suppress solar flare effects on magnetograms. The measurements of amplitudes taken directly from temporal variations of weak geomagnetic components have certain limitations regarding the identification of the proper base and peak values from which the deviation due to solar flare has to be measured. In such situations, local intermittency measure based on cross-wavelet analysis can be employed which could remarkably identify the flare effects, even if the signatures are weak or masked by other disturbance effects. The present study shows that local intermittency measure based on wavelet analysis could act as an alternate quantification technique for analyzing solar flare effects on geomagnetic activity.

Keywords Solar flare effects · Wavelets · Local intermittency measure

Introduction

The solar flares are spontaneous outbursts from the solar atmosphere, often associated with changes in brightness (Carrington 1859; Hodgson 1860). They are generally considered as powerful manifestations of energy release from the Sun. They affect the solar wind–magnetosphere–ionosphere (SW–M–I) system similar to coronal mass ejections (CMEs) and high-speed solar winds. The solar flare phenomena usually have different phases such as quiet phase, impulsive phase and declining phase. Among them, the pre-flare brightening (associated with local lighting up of H_{α}) during quiet phase often lasts for a few minutes. It is followed by an explosive phase, which displays features of rapid rise in brightness or area of the flare. They are accompanied by momentary bursts in gamma rays, hard X-rays, EUV or microwave radiation. The impulsive phase usually lasts only for a few minutes and is followed by the declining phase where thermal radiation dominates. The individual

spikes associated with solar flares will have timescales of few seconds or less (Tandberg-Hanssen and Emslie 1988).

The geomagnetic disturbance during solar flare is usually seen on a magnetogram as a small ‘spiky structure’ called crochet, and its physical characteristics are quantified in terms of geomagnetic solar flare effect (GSFE). It is defined as increase in both the magnitude of geomagnetic field and ionization in lower ionosphere (of sunlit hemisphere of the Earth), which gets detected as a spike in the magnetogram. The ionospheric current system behind the formation of GSFE is believed to be confined in the E- and D-regions of the ionosphere (Van Sabben 1961; Richmond and Venkateswaran 1971). Ohshio et al. (1967) investigated the global variations of GSFE and reported that the region of ionospheric current due to solar flares is slightly below the layer for solar quiet (S_q) current system, while the S_q currents mainly flow in the E-region. The currents that create the GSFE depend on several factors such as the tidal and thermal flows, the geomagnetic field at the ionospheric layers and the distribution of the ionospheric conductivity, to produce an independent current system (Volland and Taubenheim 1958; Van Sabben 1961). Richmond and Venkateswaran (1971) stated that the SFE consists of both fast and slow components, which they have been attributed to EUV radiation (100–1000 Å) and soft X-rays (1–100 Å),

✉ P. R. Prince
princerprasad@gmail.com

¹ Department of Physics, University College, Trivandrum, Kerala 695034, India

respectively. During the time of solar flares, the UV- and X-ray flare radiation (1–1000 Å) will be absorbed by the ionosphere causing variations in ion and electron density which in turn lead to an enhancement in the ionospheric conductivity of the D- and E-regions.

Table 1 Stations considered for the LIM analysis of geomagnetic disturbances during solar flare event on August 08–10, 2011

Station	Acronym	Geomagnetic latitude (N)	Geomagnetic longitude (E)
Alibag	ABG	10.36	146.54
Jaipur	JAI	18.35	150.16
Arti	ARS	52.73	131.99
Port-aux-Francais	PAF	−57.3	128

As the solar flare strikes the Earth, the intensity of SFE has been related to the local time such that the maximum value will be measured around local noon for different latitudes (Liu et al. 1996). The magnitudes of H due to SFE follow the latitudinal variation pattern similar to that of $S_q(H)$. While studying the SFEs on H - and Y -components at Huancayo, Rastogi (2001) reported that during normal electrojet periods, a solar flare produces a positive impulse in H (and Y) while a negative impulse in Z . Meanwhile, during counter-electrojet times, a solar flare will produce a negative jerk in H - and Y -components while a positive jerk in Z field. Hence, the study showed that a SFE is only an augmentation of both zonal and meridional components at the ionospheric currents existing at pre-flare stage. During electrojet periods, the zonal currents are eastward, producing positive ΔH , while the meridional currents are away from the magnetic equator, which gives rise to negative ΔY .

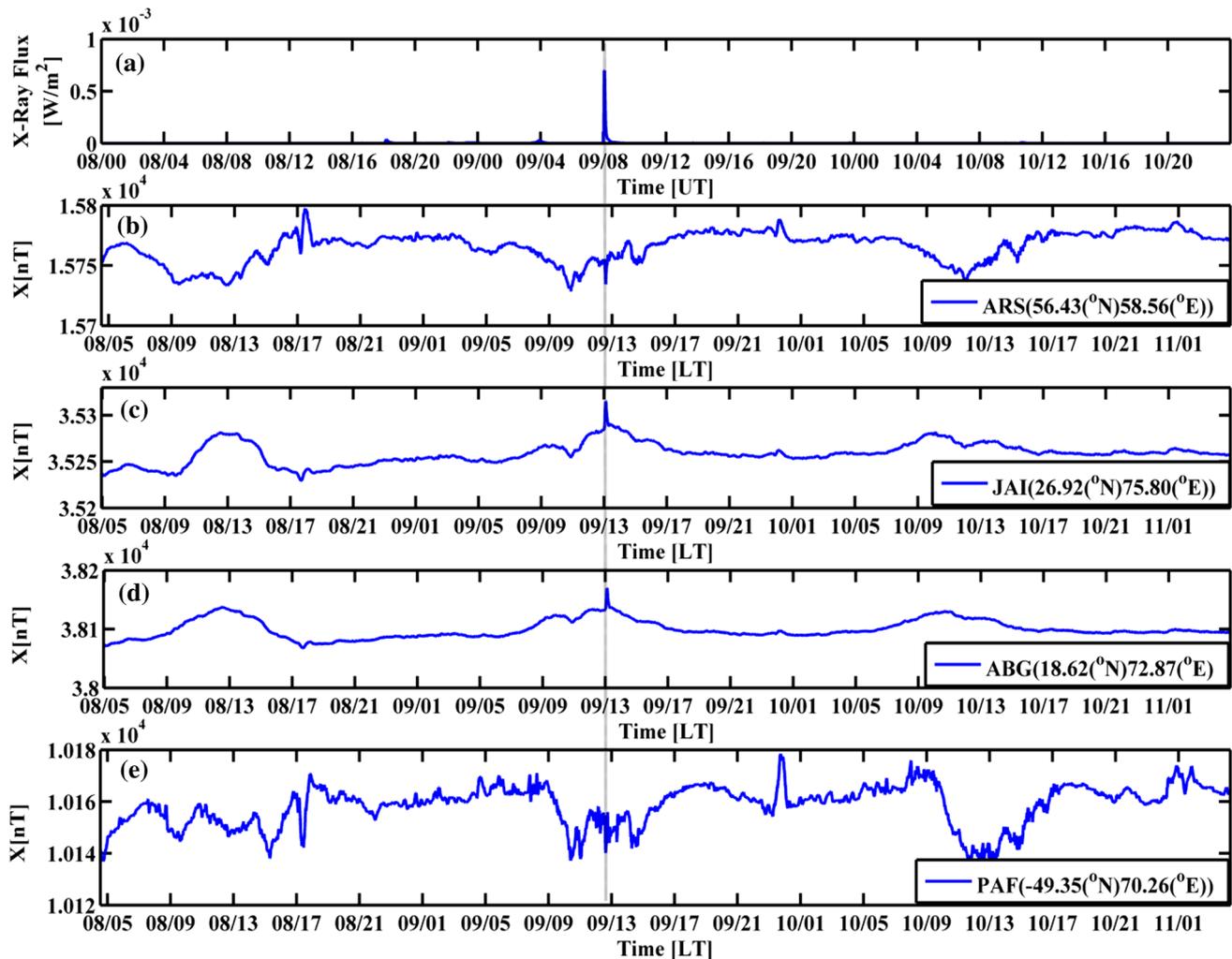


Fig. 1 X-ray flux and geomagnetic X -component variations during the period August 08–10, 2011, with X-class solar flare event on August 09, 2011: **a** X-ray flux in the range 1–8 Å and X -component

variations at **b** Arti (ARS), **c** Jaipur (JAI), **d** Alibag (ABG) and **e** Port-aux-Francais (PAF)

But, during the counter-electrojet periods, the zonal current will be reversed westward giving a negative ΔH , while the meridional current associated with the electrojet will also be reversed to equatorward direction giving a positive jerk in ΔY (Rastogi 1962, 2001; Rastogi et al. 2013). Recently, Okeke and Okpala (2005) have studied the SFE on May 06, 1998, and argued that every GSFE is unique with a characteristic signature of the event. They also argued that the SFE current system is not a simple augmentation of the S_q current system as a significant phase difference exists between both of them. Hence, for studying such complex augmentation processes, nonlinear methods such as wavelet techniques have to be employed, which could identify the deterministic and stochastic patterns that exist in geomagnetic components during flare time.

During earlier times, the period of SFE was identified from high-frequency fade-outs, and the magnitude of disturbance in geomagnetic components during that period was taken for the analysis (Nagata 1966). A prominent technique used for the quantification of SFE is based on the

comparison between S_q and disturbed current systems. It considers the enhancement due to the solar flare on the geomagnetic component as defined by $\Delta B_{\text{SFE}} = B_{\text{peak}} - B_{\text{base}}$, where B_{peak} is the value of geomagnetic component recorded at the peak time of the flare and B_{base} is the value of geomagnetic component recorded just before the start time of the flare (Volland and Taubenheim 1958; Okeke and Okpala 2005; Ugonabo et al. 2016). Dmitriev and Yeh (2008) have studied flare-associated magnetic variations in X - and Y -components by subtracting background field, calculated using linear interpolation technique, from each magnetic component. Meza et al. (2009) have considered the difference between geomagnetic field B affected by SFE and that during quiet times. According to them, the quiet condition is marked by a straight line that connects two points: one recorded in the last quiet moments and the other immediately after the SFE has disappeared.

The conventional methods that measure the amplitudes directly from temporal variations of geomagnetic components are having certain limitations regarding the

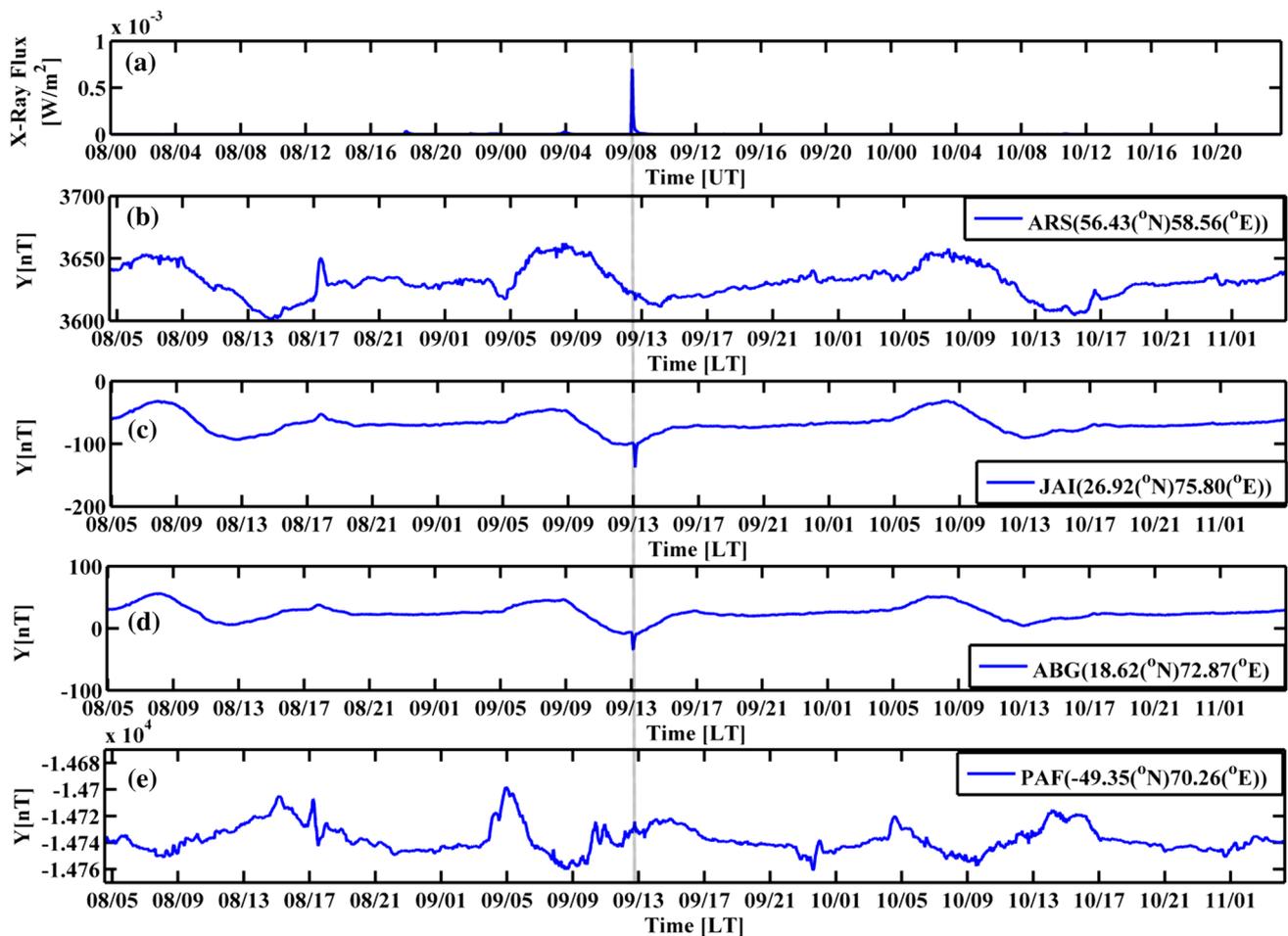


Fig. 2 X-ray flux and geomagnetic Y -component variations during the period August 08–10, 2011, with X-class SFE on August 09, 2011: a X-ray flux in the range 1–8 Å and Y -component variations at **b** Arti, **c** Jaipur, **d** Alibag and **e** Port-aux-Francais

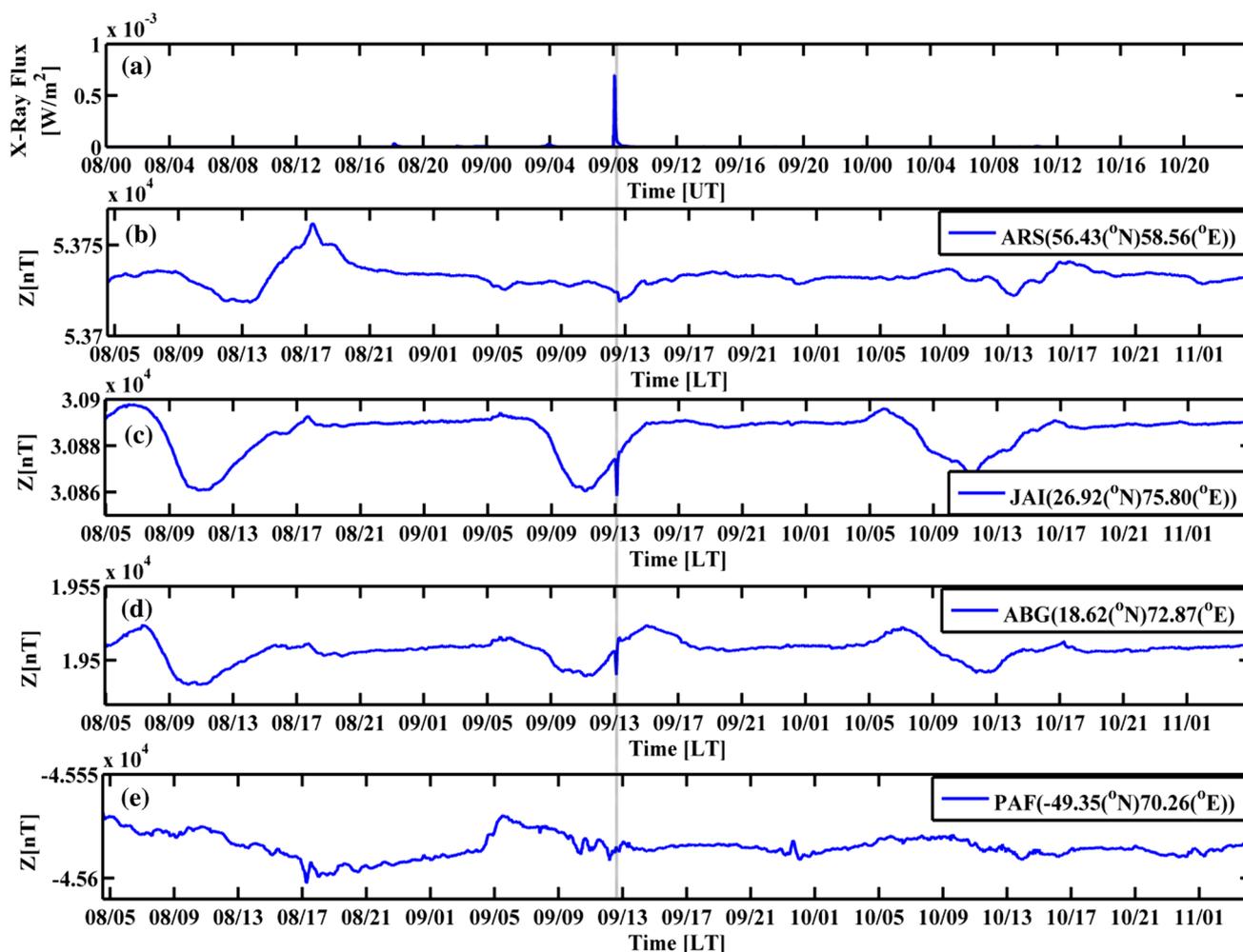


Fig. 3 X-ray flux and geomagnetic Z-component variations during the period August 08–10, 2011, with X-class SFE on August 09, 2011: a X-ray flux in the range 1–8 Å and Z-component variations at **b** Arti, **c** Jaipur, **d** Alibag and **e** Port-aux-Francais

identification of proper base value and peak value of the variation in the geomagnetic component. Most SFEs produce irregular variations on geomagnetic components, and hence, measuring the amplitude directly will be biased to a large extent. Since these methods are not employing any nonlinear functions, they have limited ability in providing accurate information on the intermittent changes occurring in the associated physical processes as well as their space–time characteristics. The disadvantages of linear methods also include limitations in identifying and interpreting hidden patterns in the behavior of the data.

The filtering of intermittent events from geomagnetic time series using a wavelet-based method was proposed by Kovács et al. (2001). Later, De Michelis and Tozzi (2005) have reported ordinary wavelet transform-based local intermittency measure (LIM) for the detection of geomagnetic jerks. In the present study, the efficiency of

novel cross-wavelet local intermittency measure in identifying solar flare effects which usually appear as crochets on magnetograms has been analyzed. If the SFE variations on the geomagnetic components during the solar flare events are analyzed, most often, we can see fluctuations associated with other solar effects shadowing the variations due to solar flares. These solar effects could be due to CMEs or high-speed solar wind streams (HSSs) which are usually developed in association with solar flares or vice versa. These effects also accompany solar flares and create their signatures on magnetograms. In most cases, the fluctuations produced by them mask the SFE variations, and to identify whether the variation is really caused by a SFE or not is quite difficult. In these situations, cross-wavelet LIM analysis can be employed which could remarkably identify the SFEs on magnetograms even if the signatures are naturally weak or masked by other disturbance effects.

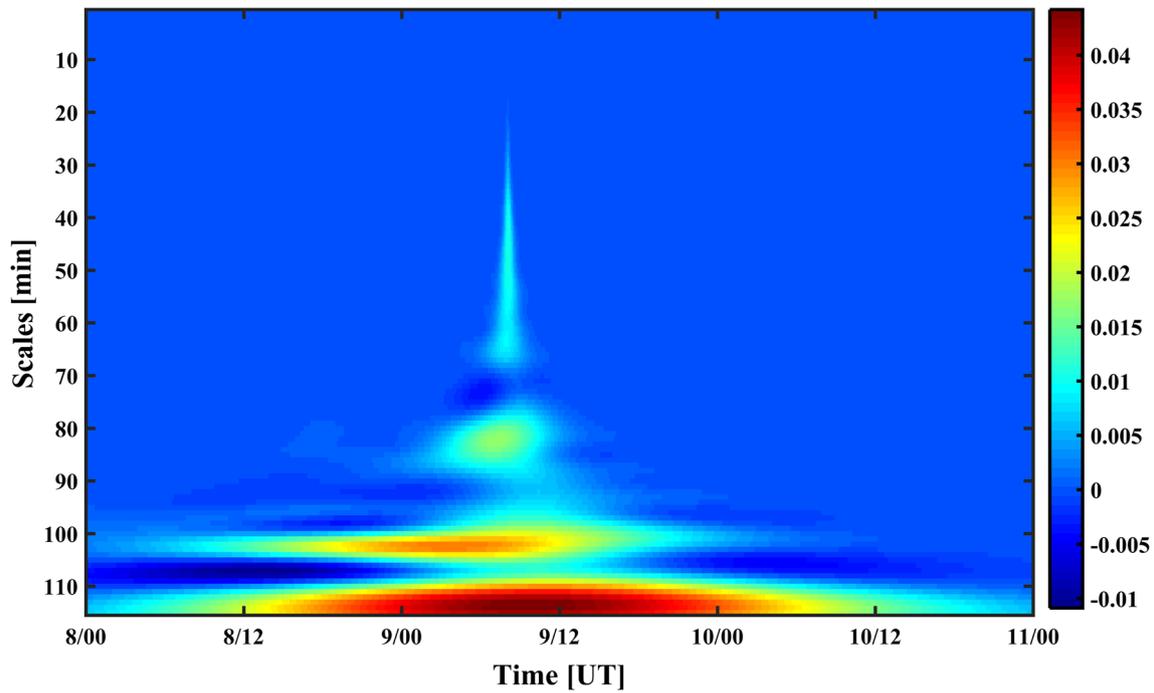


Fig. 4 Cross-wavelet transform of X-ray flux with X-component at Alibag during the period August 08–10, 2011. The scalogram intensity is shown in the right color panel

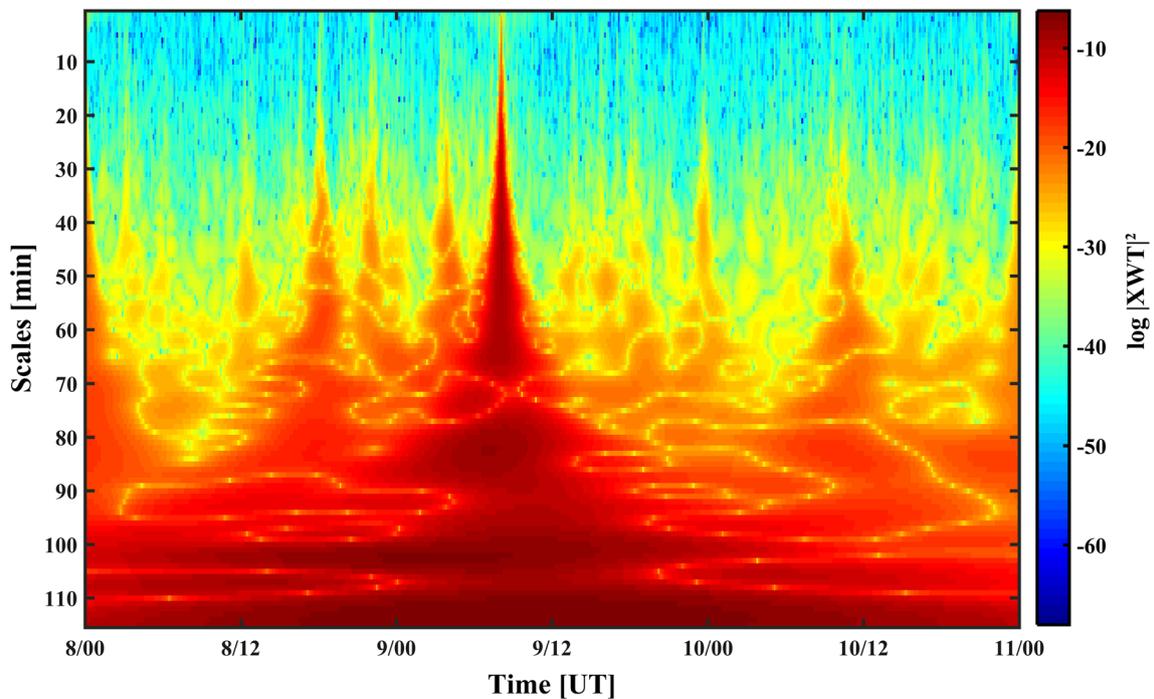


Fig. 5 Scalogram $|XWT|^2$ of X-ray flux with X-component at Alibag during the period August 08–10, 2011. The scalogram intensity is shown in log arbitrary values in the right color panel

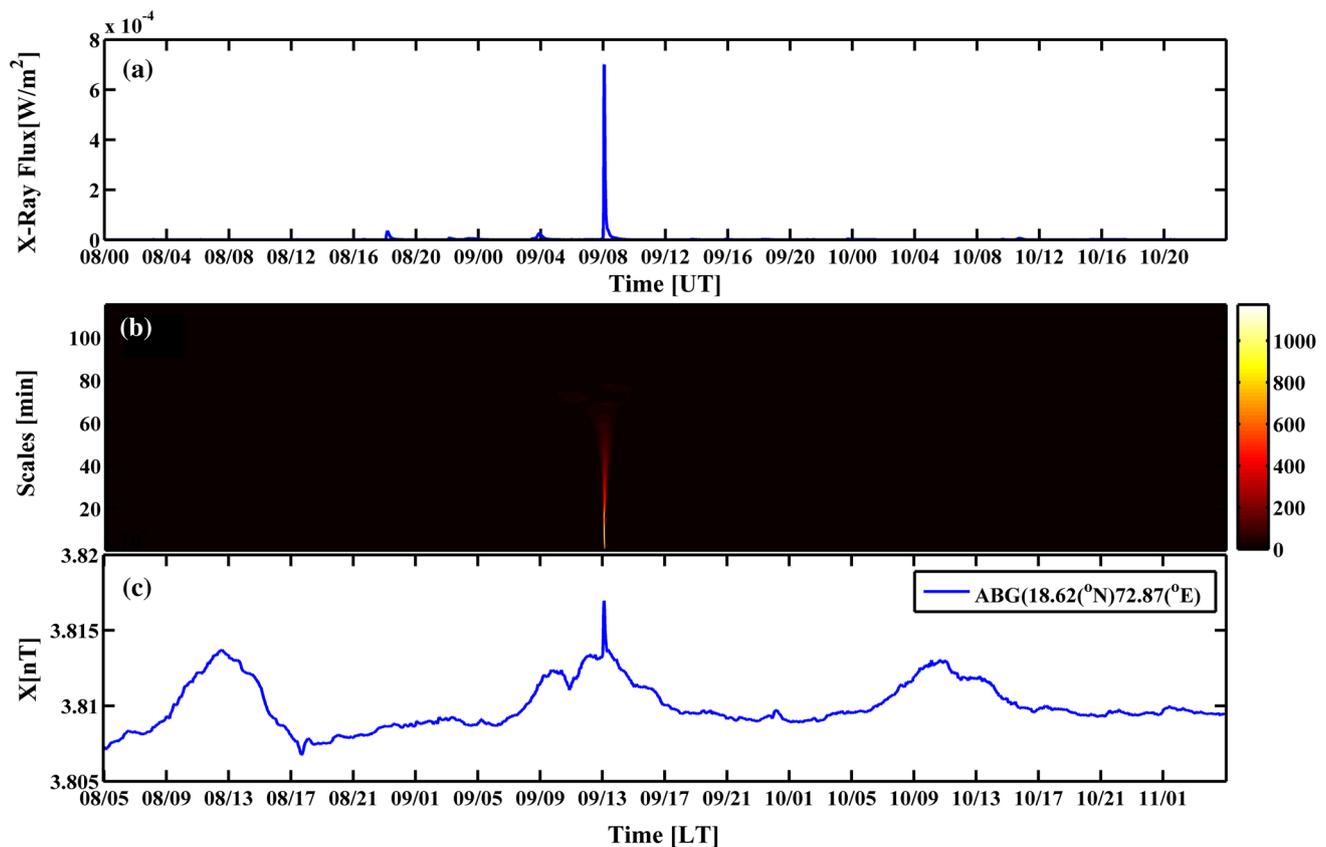


Fig. 6 LIM scalogram during the solar flare event on August 09, 2011, where **a** solar X-ray flux, **b** LIM scalogram and **c** geomagnetic X-component at Alibag

Data

The 1-min solar X-ray flux during each solar flare event has been taken from Geostationary Operational Environmental Satellite (GOES) data archive. On each GOES, there are two X-ray sensors which provide solar X-ray fluxes for the wavelength bands of 0.5–4 Å (short channel) and 1–8 Å (long channel). The data come from the National Oceanic and Atmospheric Administration Space Weather Prediction Center (NOAA SWPC) and are archived at the National Geophysical Data Center (NGDC) (<http://www.swpc.noaa.gov/products/goes-x-ray-flux>). The minute data of geomagnetic X-, Y- and Z-components have been collected from the INTERMAGNET which is a global network of geomagnetic observatories, observing the Earth's magnetic field (www.intermagnet.org).

For the present study, the flare on August 09, 2011, is considered, since it is the first largest Earth-directed X-class (X6.9) solar flare in the solar cycle 24 as measured by the GOES. The flare starts at 07:48 UT, peaks at 08:05 UT and ends at 08:08 UT. The choice of the stations, as shown in Table 1, is in such a way that geomagnetic stations at local noon where flare signatures become prominent are

considered. Since the flare is having its peak at 08:05 UT, the available geomagnetic stations which are nearly at noon will fall in the Indian sector. The data from the magnetic observatories at Alibag (ABG, 10.36°N and 146.54°E geomagnetic) and Jaipur (JAI, 18.35°N and 150.16°E geomagnetic) are hence chosen for the study. For high-latitude variations, data from a magnetic observatory far from north that comes under the Russian region, called Arti (ARS, 52.73°N and 131.99°E geomagnetic), are taken. On the southern hemisphere, data from an observatory at Port-aux-Francais (PAF, −57.3°N, 128°E geomagnetic) are taken.

Figures 1, 2 and 3 show the fluctuations in geomagnetic components X, Y and Z for 3 days (August 08–10, 2011) where the X-class flare event is on August 09, 2011. From the fluctuations in the X-component as shown in Fig. 1, positive excursions due to SFE are seen at ABG and JAI (low- and midlatitude stations), while the high-latitude station, ARS, shows a negative excursion. Another high-latitude station, PAF, also shows a negative excursion. While the excursions seen at ABG, JAI and ARS are strong, the excursion at PAF is obscured. From the fluctuations in the Y-component as shown in Fig. 2, negative excursions due to SFE are seen at ABG and

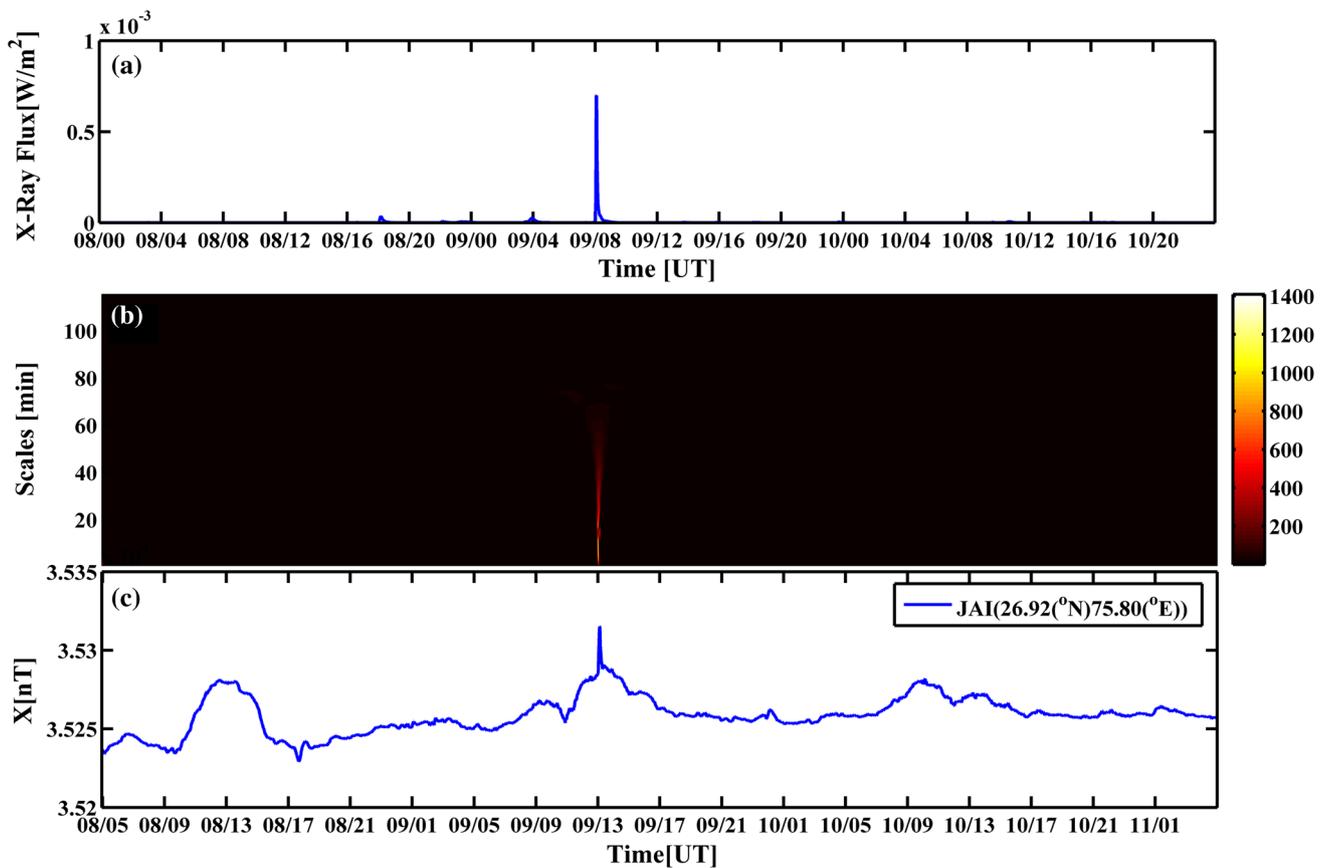


Fig. 7 LIM scalogram during the solar flare event on August 09, 2011, where **a** solar X-ray flux, **b** LIM scalogram and **c** geomagnetic X-component at Jaipur

JAI (low- and midlatitude stations), while the high-latitude station ARS does not show a positive excursion as expected. At the same time, the high-latitude station, PAF, shows a positive excursion as expected. While the excursions seen at ABG and JAI are strong, ARS and PAF show extremely weak variations which are hard to analyze by linear methods. In Fig. 3, it is clear that the fluctuations in the Z-component at ABG and JAI (low- and midlatitude stations) show negative excursions due to SFE as expected, while the high-latitude station ARS shows a slight negative excursion. In this case, the excursions at ABG and JAI during flaring time clearly fall on the excursion of the Z-component such that it is difficult to find a proper base value by linear methods. Figure 3 also shows that the excursions produced at the ARS and PAF are extremely weak which are difficult to analyze by linear methods.

Methods

Wavelet analysis can be considered as a promising tool in the time–frequency localization of geomagnetic spikes associated with solar flares. The wavelet transform is the

decomposition of a function, say $f(t)$, into time–scale contributions by convolving it with a set of localized time and scale basis functions called as daughter wavelets, which are constructed by translation and dilation of a function, the mother (or analyzing) wavelet (Meyer 1992; Venugopal et al. 2006). In other words, the signals are represented as a linear combination of scaled and shifted versions of a single wavelet function called mother wavelet.

Now, starting with a mother wavelet, say Ψ , a family $\Psi_{a,b}$ of wavelet daughters can be obtained by scaling and translating Ψ ,

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right), \quad (1)$$

where a is a scale factor that controls the width of the wavelet and b is a translation parameter controlling the location of the wavelet. The analytical equation of wavelet transform can be written as the inner product of the geomagnetic signal $f(t)$ and the time-shifted and scaled version of $\Psi_{a,b}(t)$ (Mallat 1998). The wavelet transform can be represented as,

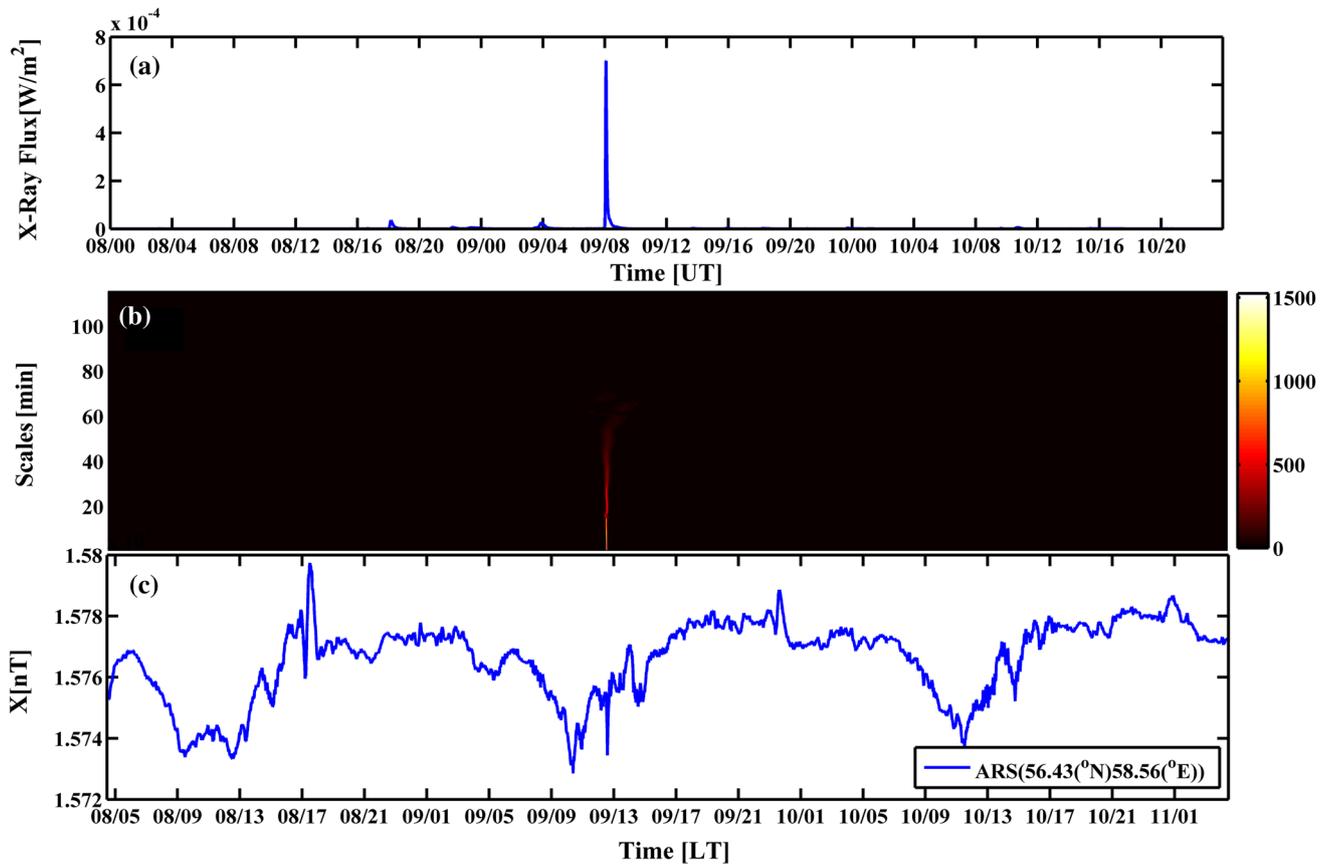


Fig. 8 LIM scalogram during the solar flare event on August 09, 2011, where **a** solar X-ray flux, **b** LIM scalogram and **c** geomagnetic X-component at Arti

$$WT(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \Psi\left(\frac{t-b}{a}\right) f(t) dt. \tag{2}$$

The local intermittency measure (LIM) is defined as,

$$WT_{LIM}(a, b) \equiv \frac{|WT(a, b)|^2}{\langle |WT(a, b)|^2 \rangle_b}, \tag{3}$$

where $WT_{LIM}(a, b)$ is the ratio of the squared wavelet amplitude $|WT(a, b)|^2$ to its average value $|WT(a, b)|^2_b$ obtained averaging over b (Farge et al. 1990; Farge 1992; Meneveau 1991; Bruno et al. 1999; Consolini and de Michelis 2005). Since the matter of interest in the present analysis is the flare-induced spike detection, directly seeking LIM on geomagnetic time series will not give desired results. Hence, cross-wavelet transform (XWT) of solar X-ray flux and geomagnetic components is employed, and LIM is calculated from the cross-wavelet transform coefficients (Giménez de Castro et al. 2016). The cross-wavelet transform can be represented as,

$$XWT(a, b) = \left| WT_{sig1}(a, b) \times WT_{sig2}^*(a, b) \right|, \tag{4}$$

where the absolute value of product of the continuous wavelet transform (CWT) of the first signal and complex conjugate (represented by asterisk) of the CWT of the second signal are taken. Now, for finding the cross-wavelet LIM, Eq. (3) can be modified as,

$$XWT_{LIM}(a, b) \equiv \frac{|XWT(a, b)|^2}{\langle |XWT(a, b)|^2 \rangle_b}. \tag{5}$$

Here, intermittency is considered as change in ionospheric equilibrium, from a calm or quiescent phase to an active or disturbance phase due to the sudden impact of solar flares (Giménez de Castro et al. 2016). It is to be noted that $XWT_{LIM}(a, b) = 1$ signifies the condition where function $f(t)$ does not show any signatures of intermittency since each region has the same energy spectrum where it corresponds to ordinary Fourier spectrum (De Michelis and Tozzi 2005). If $XWT_{LIM}(a, b) > 1$, that is, at time b the

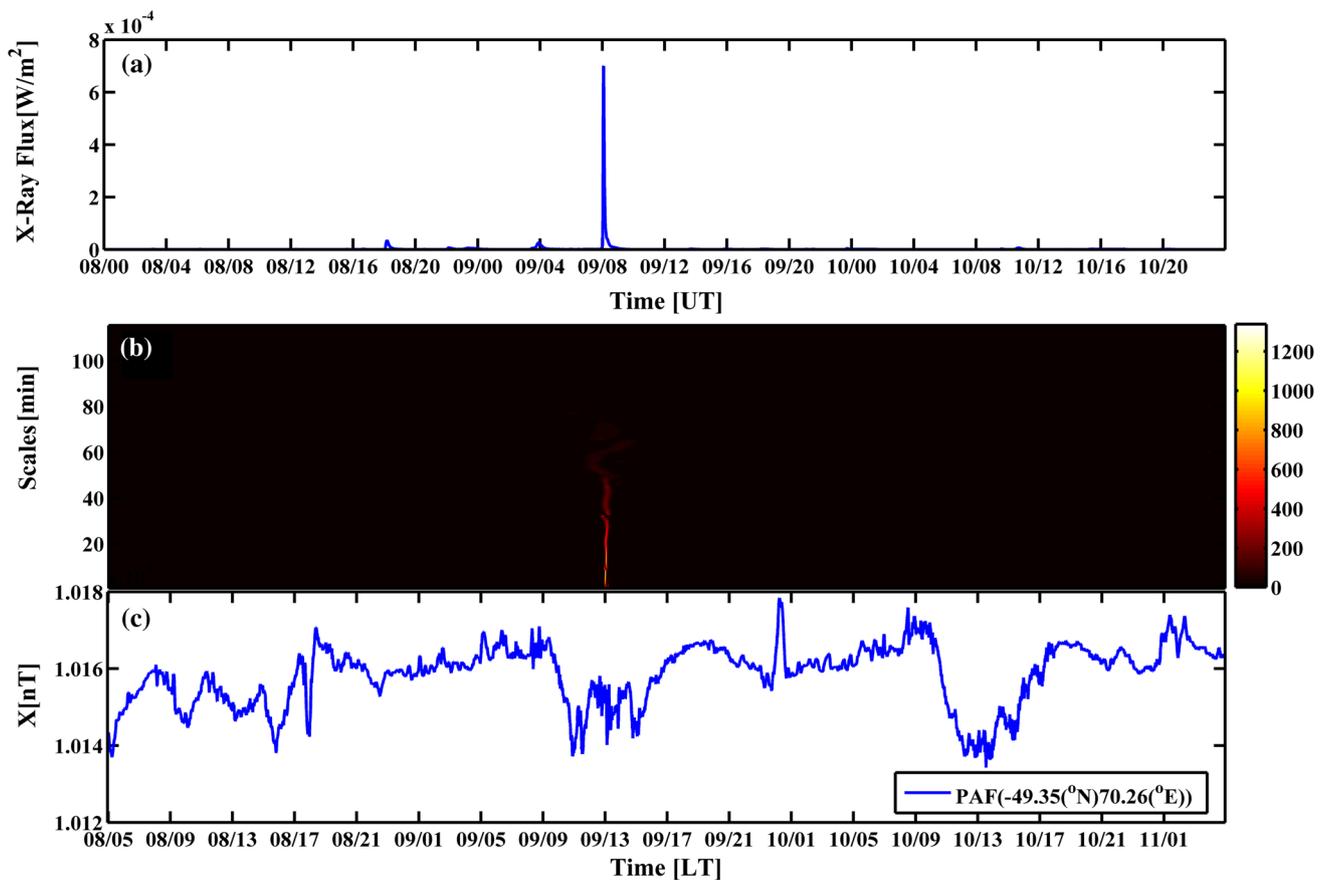


Fig. 9 LIM scalogram during the solar flare event on August 09, 2011, where **a** solar X-ray flux, **b** LIM scalogram and **c** geomagnetic X -component at Port-aux-Francais

variations at scale a have more energy than they would in case of the Fourier spectrum. In other words, the condition $XWT_{LIM}(a, b) > 1$ identifies those regions of the function $f(t)$ which are having more energy than they should have if they were normally distributed (or a deviation from Gaussian distribution of fluctuations). Thus, the condition $XWT_{LIM}(a, b) > 1$ helps to locate the geomagnetic spikes in magnetograms associated with solar flares in time and scale. The wavelet coefficients contain the amplitude and phase information of the signal under analysis over the whole time–scale plane. The absolute value of the coefficient matrix indicates the wavelet energy of the signal. From the cross-wavelet LIM scalogram, it is also possible to extract the flare signature as a signal $g(t)$ by employing inverse continuous wavelet transform (ICWT), which can be written as,

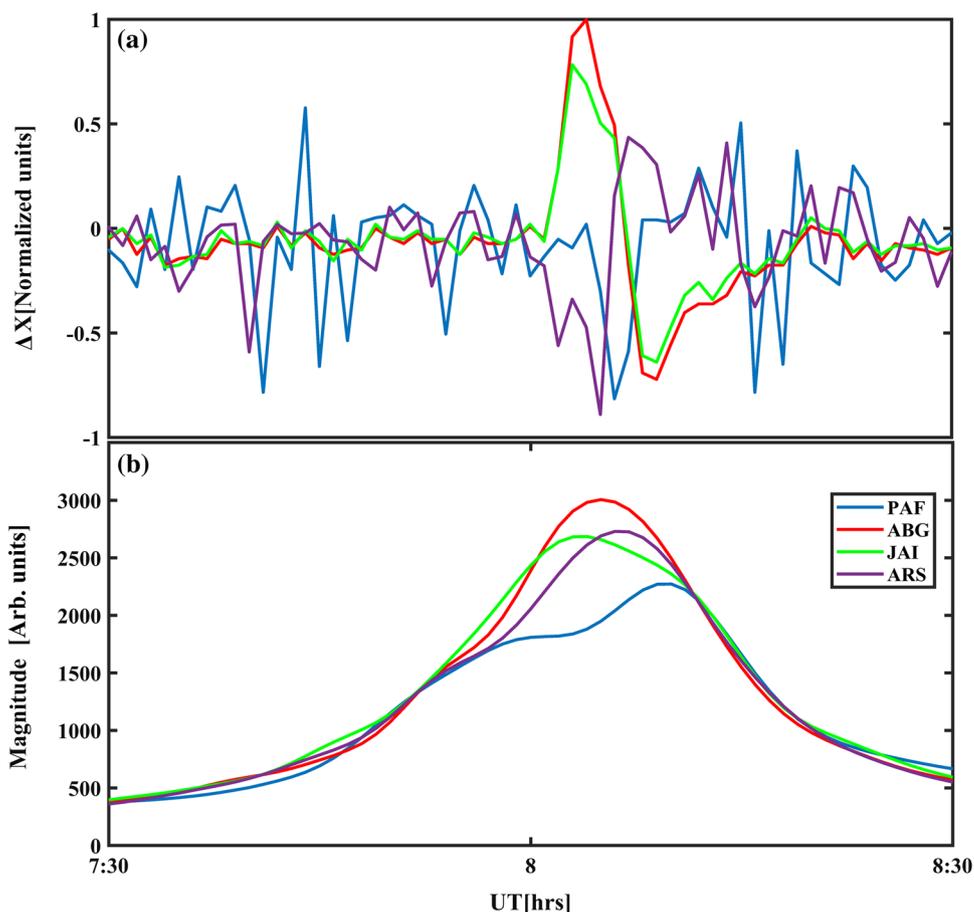
$$g(t) = \frac{1}{N_\psi} \int_0^\infty da \int_{-\infty}^\infty XW_{LIM}(a, b) \Psi\left(\frac{b-t}{a}\right) db, \quad (6)$$

where N_ψ is a normalization constant which depends on the chosen wavelet and $XWT_{LIM}(a, b)$ are the cross-wavelet coefficients related to flare signatures (after applying the LIM condition). For the present analysis, continuous bump wavelet is used since it is analytic and will have only a single peak frequency (band-limited) which makes it suitable for detecting flare signatures (Meignen et al. 2012). For LIM intensities having a bilobed (or multi-lobed) structure, a Gaussian fit is made, and the peak amplitudes are calculated from the fitted function. In the present study, the LIM analysis is performed by taking the first difference of geomagnetic series.

Results and discussion

Figures 4 and 5 show cross-wavelet transform scalogram and cross-wavelet logarithmic intensity scalogram $|XWT_{a,b}|^2$, respectively, of the X -component at ABG against solar X-ray

Fig. 10 **a** The first difference series of the geomagnetic X -component and **b** inverse wavelet transform of LIM coefficients during the event time on August 09, 2011, for different stations such as Port-aux-Francais, Alibag, Jaipur and Arti

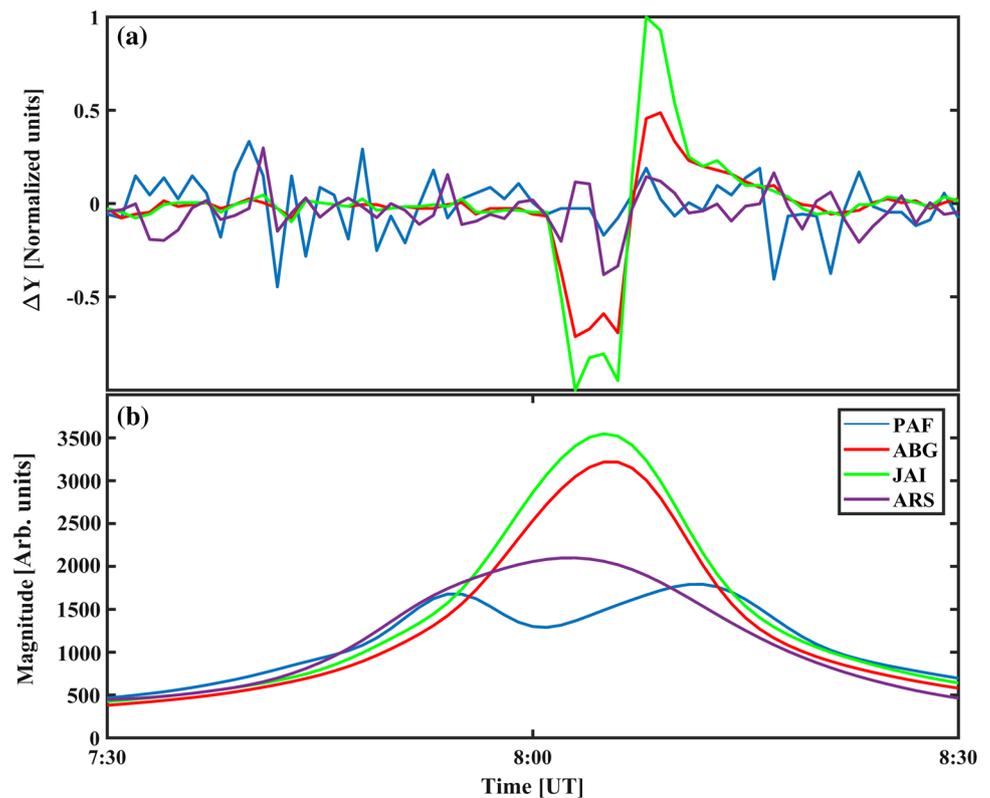


flux during the period August 08–10, 2011. Even though LIM has been introduced to detect coherent structures that are responsible for intermittency in turbulent fluid flows, in the present work LIM is having a more general meaning. According to the present scenario, LIM is useful in the identification and localization in time of the active events which are characterized by energy above the threshold. The term ‘intermittence’ is used in the sense of ‘punctuated equilibrium’ as suggested by De Michelis and Tozzi (2005). This means that a certain calm phase is interrupted by a disturbance which generates a singularity or a significant variation in equilibrium of the ionosphere. The crochets can be considered as abrupt changes which can be interpreted as crisis events in temporal trends of secular variation of Earth’s geomagnetic field components. Thus, local intermittency arises as a result of change in ionospheric equilibrium at each geomagnetic location, from a calm phase to a disturbed phase due to the impact of solar flares. The scale–time structure of scalogram $|XWT_{a,b}|^2$ is highly complex, which portrays the existence of different coherent scale–time structures in

coincidence with the intensifications in both geomagnetic X -component and X-ray flux. The high-intensity structure in the scalogram near 09/08 UT shows the flare-related coherent structure, but it is not easy to directly extract it from others. By applying the LIM method in this scalogram, we can extract the exact intermittent structure that acts as a fingerprint of flare-time excursion in the geomagnetic component.

Figures 6, 7, 8 and 9 show LIM scalogram of the geomagnetic X -components with X-ray flux at ABG, ARS and PAF, respectively. According to the LIM condition, only the flare-time intermittent structure exists and all other structures must vanish. For that, the wavelet coefficient matrix will only have values satisfying the LIM condition to facilitate the identification of intermittent processes. The scale–time structure extracted using the LIM technique coincides with the flare-time excursion of geomagnetic component as seen in all Figs. 6, 7, 8 and 9. The singularities due to solar flare effects are clearly detectable, while the other unwanted singularities are suppressed. Thus, it is indeed possible to say that LIM scalogram could simultaneously enhance the

Fig. 11 **a** The first difference series of the geomagnetic Y -component and **b** inverse wavelet transform of LIM coefficients during the event time on August 09, 2011, for different stations such as Port-aux-Francais, Alibag, Jaipur and Arti



localization of intermittencies both in time and in scale. It is to be noted that the continuous wavelet transform local intermittency measure (WT_{LIM}) of geomagnetic X -component alone (not shown) will reveal several intermittent structures associated with anomalous fluctuations in the geomagnetic series of which one will be the flare-time signature. In contrast, the cross-wavelet transform local intermittency measure (XWT_{LIM}) selects only those singularities related to solar flare effect.

By considering only those areas of the scalogram, which is related to the observed flare-induced intermittency, it is possible to extract the scale–time structure by means of the inverse continuous wavelet transform (ICWT). The amplitude variation of reconstructed signals from LIM coefficients characterizes the coupling efficiency of geomagnetic components with X-ray flux. In other words, the coupling efficiency acts as a quantifier of the amount of energy that perturbs the S_q current system during solar flares. Figure 10a shows the first difference series of geomagnetic X -components from ABG, JAI, ARS and PAF. Figure 10b shows the inverse wavelet transform analysis of LIM coefficients which give the magnitudes of coupling between geomagnetic X -components of ABG,

JAI, ARS and PAF with X-ray flux during the solar flare event of August 09, 2011. It is seen that the ABG station is having maximum amplitude when compared with other stations. The JAI and ARS are having opposite excursions in the ΔX series, while they are having almost equal magnitudes in the inverse transform. PAF is having the lowest amplitude of excursion among the four stations. The variations of geomagnetic Y -component at different stations are captured by the LIM analysis as shown in Fig. 11. Here, JAI shows higher amplitude than all the other stations. The amplitude of ABG is more than that of ARS and PAF. Figure 12b shows the variations of geomagnetic Z -component at different stations where JAI shows higher intensity than all the other stations. The weak and opposite intensities in ΔZ from PAF and ARS create bimodal intensity distributions. To quantify bimodal intensities, a Gaussian fitting is performed such that this could give a decent approximation of the resultant intensity. Figure 13 shows a sample Gaussian fit on the reconstructed signal from the X -component of PAF station. Although not shown, similar Gaussian fit is performed on all other cases where bimodal intensity distributions arise. Finally,

Fig. 12 **a** The first difference series of the geomagnetic Z-component and **b** inverse wavelet transform of LIM coefficients during the event time on August 09, 2011, for different stations such as Port-aux-Francais, Alibag, Jaipur and Arti

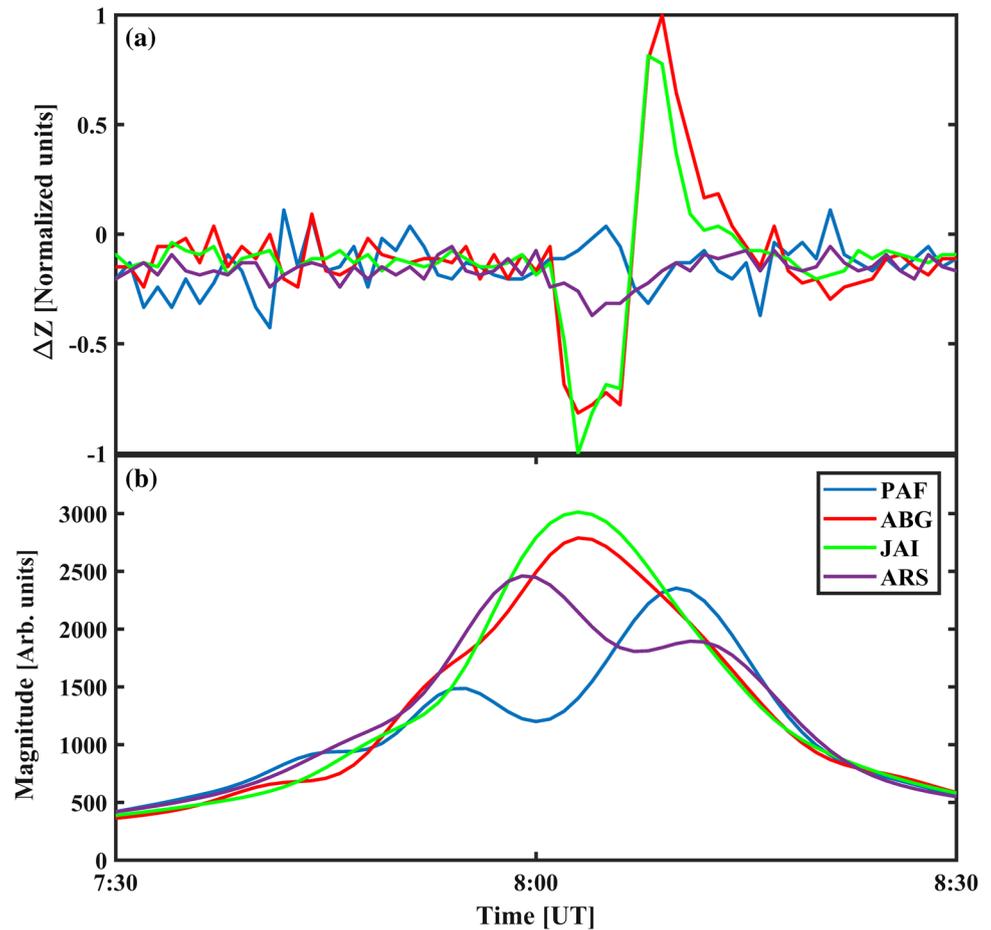


Fig. 14 shows the respective variations of magnitudes of geomagnetic X-, Y- and Z-components from PAF, ABG, JAI and ARS stations using LIM analysis. As we can see, the stations JAI and ABG show intense solar flare effect when compared with high-latitude stations such as ARS and PAF. Considering the high-latitude stations, ARS shows much higher solar flare effect than PAF. Figure 15 shows the conventional analysis of solar flare effects on geomagnetic X-component where ΔG_{SFE} from each station is also shown. Table 2 shows the resultant intensities of SFEs computed through conventional and LIM methods. The variations of X_{LIM} and $|\Delta X_{\text{SFE}}|$ are consistent in the sense that ABG is having the highest value of X_{LIM} and $|\Delta X_{\text{SFE}}|$, while PAF is having the lowest value of X_{LIM} and

$|\Delta X_{\text{SFE}}|$. But the variations of Y_{LIM} versus $|\Delta Y_{\text{SFE}}|$ are dissimilar in case of lowest value stations. Such a variation arises due to the fact that the flare-time structures are not prominent for conventional analysis in case of PAF and ARS (as evident from Figs. 2, 3). Z_{LIM} versus $|\Delta Z_{\text{SFE}}|$ are dissimilar in case of highest value stations as both ABG and JAI show equal intensity in $|\Delta Z_{\text{SFE}}|$. But, it is seen that Z_{LIM} could differentiate the amplitude variations of ABG and JAI where both stations are having the same $|\Delta Z_{\text{SFE}}|$ values. Thus, the present study shows that LIM analysis using cross-wavelet transform could detect and quantify the variations of geomagnetic activity during solar flares in a different way and could supplement the conventional analysis of solar flare effects on geomagnetic activity.

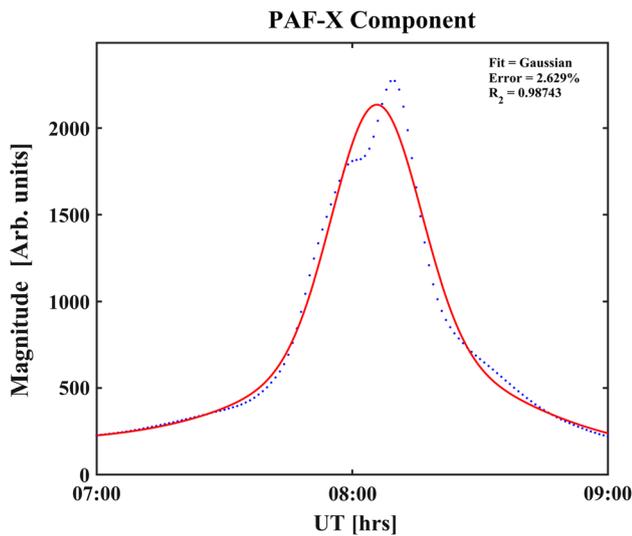


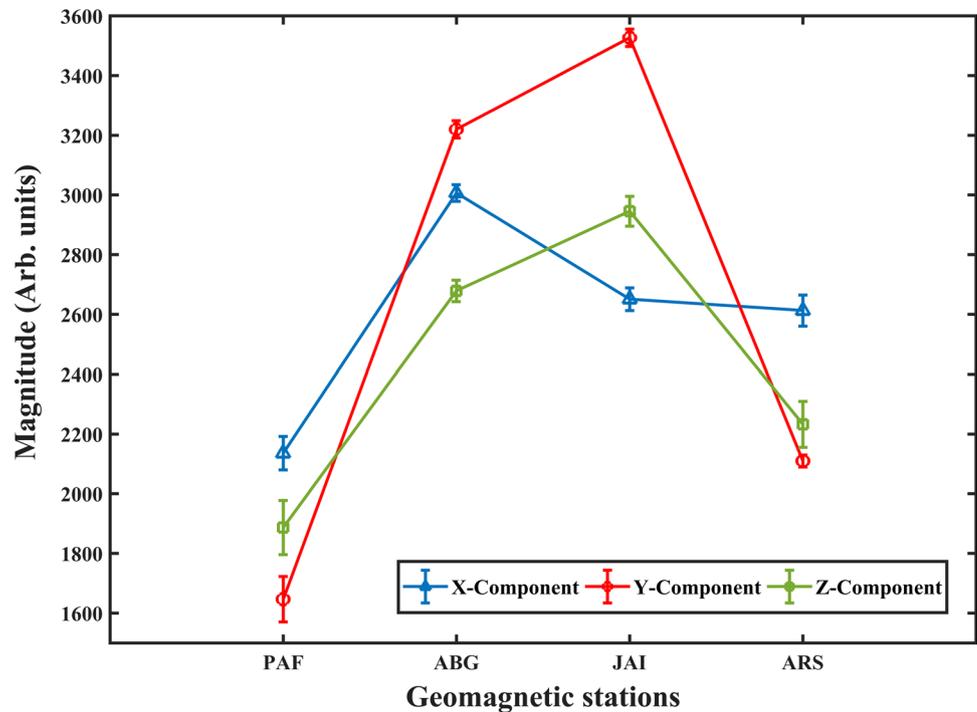
Fig. 13 A sample Gaussian fit on the reconstructed signal from LIM wavelet coefficients from the X-component of Port-aux-Francais (PAF) station

Conclusion

From the present analysis, it is seen that the cross-wavelet LIM analysis identifies the solar flare effects on magnetograms and quantifies them in an alternate way. It is evident that the LIM method accounts for time shifts in

geomagnetic components, since the coupling of geomagnetic excursions with X-ray fluxes is highly dependent on the time correlation of the flare signature on X-ray flux as well as on geomagnetic signal. Thus, if time correlation is poor, the resultant intensity will also be proportionally diminished. If strength and span of the excursion is the same, then the geomagnetic station that shows maximum time correlation with X-ray flux has the highest amplitude of local intermittency. Similarly, the method has the potential to quantify both the positive and negative excursions alike, and hence, a direct comparison of magnitude is always possible. The conventional linear method of taking the magnitude of spikes directly is having so many difficulties regarding the choice of the baseline of geomagnetic fluctuations, the identification of exact variation due to solar flare effect from noisy geomagnetic signals (especially signals from high-latitude stations), the quantification of low-magnitude excursions from stations that are off-noon and so on. It is also an advantage that LIM could differentiate the amplitude variations where both stations are having nearly the same $|\Delta G_{SFE}|$ values. The major limitation of the present study is that only one event is used, and hence, a more detailed analysis incorporating several events is required to completely understand and verify the usefulness of LIM in identifying solar flare effects on magnetograms. But, this study shows undoubtedly that the wavelet LIM method is still promising as it could act as a

Fig. 14 Variations in magnitudes of geomagnetic X-, Y- and Z-components from PAF, ABG, JAI and ARS stations using LIM analysis



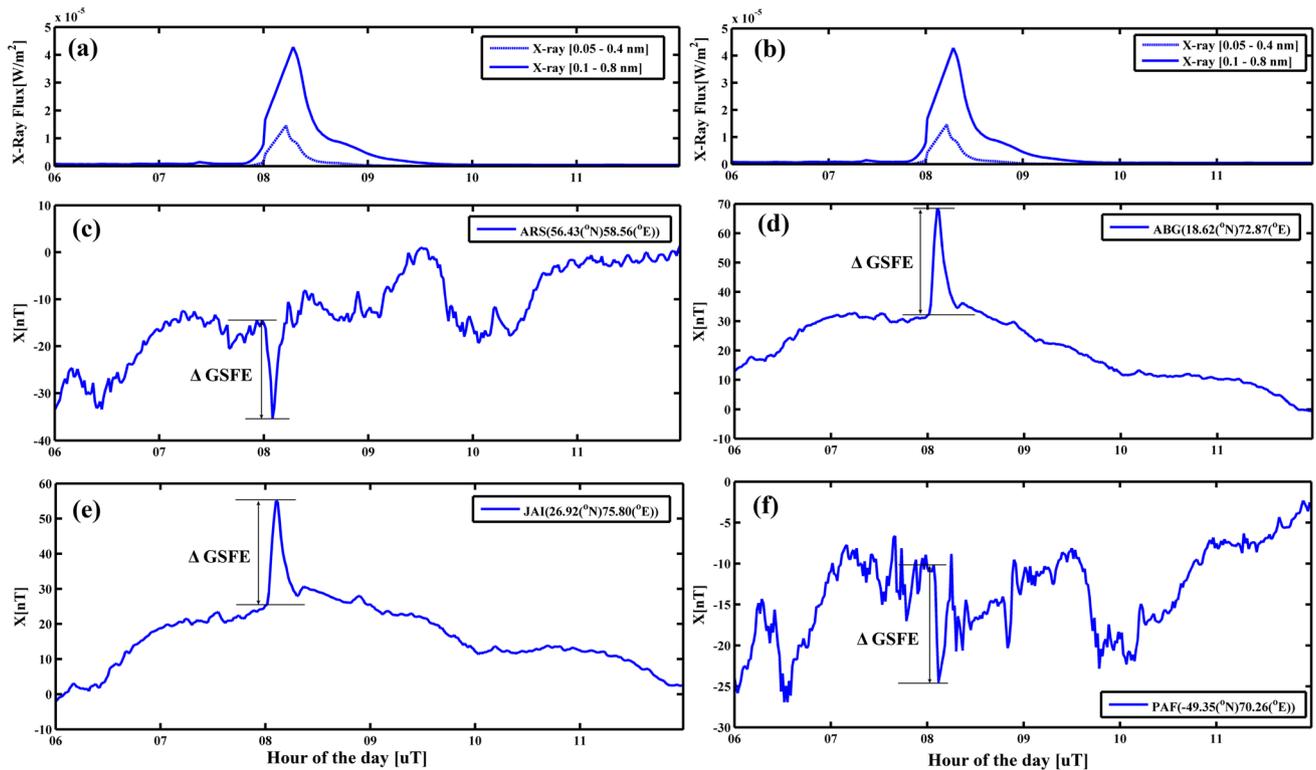


Fig. 15 Conventional analysis of solar flare effects on the X-component of geomagnetic stations where **a, b** show X-ray flux, while **c** ARS, **d** ABG, **e** JAI and **f** PAF show geomagnetic X-component vari-

ations during August 09, 2011. The ΔG_{SFE} is marked on all graphs **c–f**, where base value is chosen arbitrarily as the most possible value just before the excursion commences

Table 2 A comparison between LIM peak amplitudes of geomagnetic components and $|\Delta G_{SFE}|$ by conventional analysis from ABG, JAI, ARS and PAF stations for the event day, August 09, 2011

Station	X_{LIM}	Y_{LIM}	Z_{LIM}	$ \Delta X_{SFE} $	$ \Delta Y_{SFE} $	$ \Delta Z_{SFE} $
ABG	3007	3220	2791	37.3	29.0	16.0
JAI	2686	3550	3014	30.7	38.9	16.0
ARS	2731	2100	2219	18.5	5.5	4.1
PAF	2108	1647	1876	14.6	16.0	2.9

supportive technique for the conventional linear method for analyzing solar flare effect on geomagnetic activity.

Acknowledgements The authors are thankful to NOAA Space Weather Prediction Center (SWPC) for providing GOES X-ray flux archived at the NOAA National Geophysical Data Center (NGDC) (<http://www.swpc.noaa.gov/products/goes-x-ray-flux>). The authors are also thankful to INTERMAGNET for providing 1-min data of geomagnetic components (www.intermagnet.org). One of the authors SG acknowledges a Junior Research Fellowship (Ac.EVI(4)/17465/JRF/2017) from the University of Kerala, Trivandrum.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Bruno R, Bavassano B, Pietropaolo E, Carbone V, Veltri P (1999) Effects of intermittency on interplanetary velocity and magnetic field fluctuations anisotropy. *Geophys Res Lett* 26(20):3185–3188
- Carrington RC (1859) Description of a singular appearance seen in the Sun on September 1. *Mon Not R Astron Soc* 20:13–15
- Consolini G, De Michelis P (2005) Local intermittency measure analysis of AE index: the directly driven and unloading component. *Geophys Res Lett* 32:1–4
- De Michelis P, Tozzi R (2005) A Local Intermittency Measure (LIM) approach to the detection of geomagnetic jerks. *Earth Planet Sci Lett* 235:261–272
- Dmitriev AV, Yeh HC (2008) Geomagnetic signatures of sudden ionospheric disturbances during extreme solar radiation events. *J Atmos Sol Terr Phys* 70:1971–1984
- Farge M (1992) Wavelet transforms and their applications to turbulence. *Annu Rev Fluid Mech* 24:395

- Farge M, Holschneider M, Colonna JF (1990) Wavelet analysis of coherent structures in two-dimensional turbulent flows in Topological Fluid Mechanics. Cambridge University Press, Cambridge
- Giménez de Castro G, Simões PJA, Raulin JP, Guimarães OM (2016) Analysis of intermittency in submillimeter radio and hard X-Ray data during the impulsive phase of a solar flare. *Sol Phys* 291:2003–2016
- Hodgson R (1860) On a curious appearance seen in the Sun. *Mon Not R Soc* 20:15–16
- Kovács P, Carbone V, Vörös Z (2001) Wavelet-based filtering of intermittent events from geomagnetic time-series. *Planet Space Sci* 49:1219–1231
- Liu JY, Chiu CS, Lin CH (1996) The solar flare radiation responsible for sudden frequency deviation and geomagnetic fluctuation. *J Geophys Res* 101:10855–10862
- Mallat S (1998) A wavelet tour of signal processing. Academic Press, New York
- Meignen S, Oberlin T, McLaughlin S (2012) Multicomponent signal denoising with synchrosqueezing. In: 2012 IEEE Statistical Signal Processing Workshop (SSP 2012), pp 660–663
- Meneveau C (1991) Analysis of turbulence in the orthonormal wavelet representation. *J Fluid Mech* 232:469–520
- Meyer Y (1992) Wavelets and operators. Cambridge University Press, Cambridge
- Meza A, Van Zele MA, Rovira M (2009) Solar flare effect on the geomagnetic field and ionosphere. *J Atmos Sol Terr Phys* 71:1322–1332
- Nagata T (1966) Solar flare effect on the geomagnetic field. *J Geomagn Geoelectr* 18:197–219
- Ohshio M, Fukushima N, Nagata T (1967) Solar flare effects on geomagnetic field. *Rep Ionos Space Res Jpn* 21:77–114
- Okeke FN, Okpala KC (2005) Geomagnetic solar flare effect in H, Y, and Z fields in the Northern Hemisphere. *NJSR* 1:39–99
- Rastogi RG (1962) Longitudinal variation in the equatorial electrojet. *J Atmos Terr Phys* 24:1031–1040
- Rastogi RG (2001) Electromagnetic induction due to solar flares at equatorial stations. *J Atmos Sol Terr Phys* 63:599–604
- Rastogi RG, Chandra H, Yumoto K (2013) Unique examples of solar flare effects in geomagnetic H field during partial counter electrojet along CPMN longitude sector. *Earth Planets Space* 65:1027–1040
- Richmond AD, Venkateswaran SV (1971) Geomagnetic crochets and associated ionospheric current systems. *Radio Sci* 6:139–164
- Tandberg-Hanssen E, Emslie AG (1988) The physics of solar flares. Cambridge University Press, Cambridge
- Ugonabo OJ, Okeke FN, Ugwu BI (2016) Solar flare effects (SFE) on geomagnetic fields across latitudes. *Int J Phys Sci* 11:112–120
- Van Sabben D (1961) Ionospheric current systems of ten IGY-solar flare effects. *J Atmos Terr Phys* 22:32–42
- Venugopal V, Roux SG, Fofoula-Georgiou E, Arneodo A (2006) Revisiting multifractality of high-resolution temporal rainfall using a wavelet-based formalism. *Water Resour Res* 42(6):W06D14
- Volland H, Taubenheim J (1958) On the ionospheric current system of the geomagnetic solar flare effect (SFE). *J Atmos Terr Phys* 12:258–265



Precipitation and other propagation impairments effects at microwave and millimeter wave bands: a mini survey

Sarat Kumar Kotamraju¹ · Ch Sri Kavya Korada¹

Received: 3 November 2017 / Accepted: 19 December 2018 / Published online: 7 February 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

The current past has seen a sensational increment in the utilization of satellites for the applications like navigation, entertainment, media transmission, remote sensing, mobile communications, weather forecasting, defense and other purposes. These applications are assigned in the microwave and millimeter wave bands, which offer higher information transfer possibility in lesser time and use very small antennas and devices by ensuring secured and effective communications. However, beyond the 10 GHz range of frequencies these applications are generally subjected to signal losses due to various atmospheric parameters like rain, clouds, fog, hail ice and other applicable phenomena. The main factor for the signal degradation is the rainfall. The attenuation caused by rain increases with frequency, as there is increased absorption of the RF energy at higher frequencies due to water drops present along the path of the transmission; hence, the signal attenuation is more in higher-frequency bands. The other factors that induce losses in the signal are the clouds, gases present in the lower atmosphere and the different layers in the atmosphere that cause scintillation and the system losses and cable losses. This survey article abridges all outcomes related to propagation impairments and attenuation aspects at microwave and millimeter wave frequencies covering the studies of various researchers in last three decades. In addition, few of the models developed by various researchers were listed along with model parameters which are useful for the propagation engineers and others who are interested in this specialization.

Keywords Propagation impairments · Rain attenuation · Cloud attenuation · Earth–space paths · Microwave and millimeter frequencies

Introduction

The execution of all the satellite frameworks working in the higher-frequency bands basically relies upon the spread attributes of medium of transmission in general, i.e., space and atmosphere. Effects in light of the ionosphere can be slighted at frequencies more than 10 GHz. Tropospheric impacts in any case can cause loss of signal on Earth–space paths for considerable rates of time, which prompts diminish in the quality and availability of actual signals. Likely, the most basic tropospheric impacts are precipitation attenuation, depolarization and tropospheric scintillations which are delineated here under: (1) link accessibility, when there is no

significant power available for effective communication at the receiver; (2) degraded performance, identified with intervals when service quality is lesser than the required level; and (3) unwanted signal interference from other sources operation in similar frequencies or from cross-polarization obstruction (intersystem or intra-framework) caused due to depolarization.

Rain attenuation

In general, the technical parameters that usually impact the propagation are operating frequency, polarization, latitude, longitude, Earth station altitude from mean sea level, noise temperature and finally local meteorological parameters. The most important propagation impairments and phenomena those are impacting the communication links through the satellite systems for frequencies beyond 10 GHz are outlined as follows.

✉ Sarat Kumar Kotamraju
kksarat@gmail.com

¹ Department of Electronics & Communication Engineering,
Koneru Lakshmaiah Education Foundation, Vaddeswaram,
Guntur, AP 522502, India

Whenever radio wave propagates via rain droplets, fog, snow, clouds, gases, hail, melting layer or ice crystals, sky noise emission, it severely suffers from loss of power because of hydrometeor scattering as shown in Fig. 1. Although this raindrop scattering is one of the important criteria and factor to be considered for frequencies greater than 10 GHz, the absorption caused by hydrometeor is the most influential phenomenon causing the significant power loss in the range of frequencies between 10 and 30 GHz. The effect of hydrometeor absorption and scattering together realizes a relation between power loss (in dB) which is equal to the square of the operational frequency and was discussed by Stutzman (1993). This constitutes the guideline impediment of working at millimeter frequencies. To the extent of satellite links are concerned, elevation angle and polarization assume a vital role in calculation of losses due to rain. On the other hand, as rain attenuation thus depends unfavorably on the precipitation rate and the raindrop evaluate movement, it impacts seriously tropical areas.

Attenuation due to rain can be acquired specifically through estimation or predicted from the aspects of rainfall rate (mm/hr) and drop size distribution (DSD). The spectrum of raindrop sizes decides the precipitation rate and the radar reflectivity along with microwave radiation discharge of raindrops. In this way, it is conceivable to evaluate the rain attenuation by methods for indirect estimations performed with rain gauges, radiometers and radars. The methodology for predicting the attenuation depends on the relationship between specific attenuation and rain rate, built up through the modeling of the rain microstructure, e.g., the terminal falling speed, temperature, shape, size, of the raindrops.

This review paper has been formulated using the outcomes of various research papers by highlighting the

significant points from each source considering precipitation and other impairments.

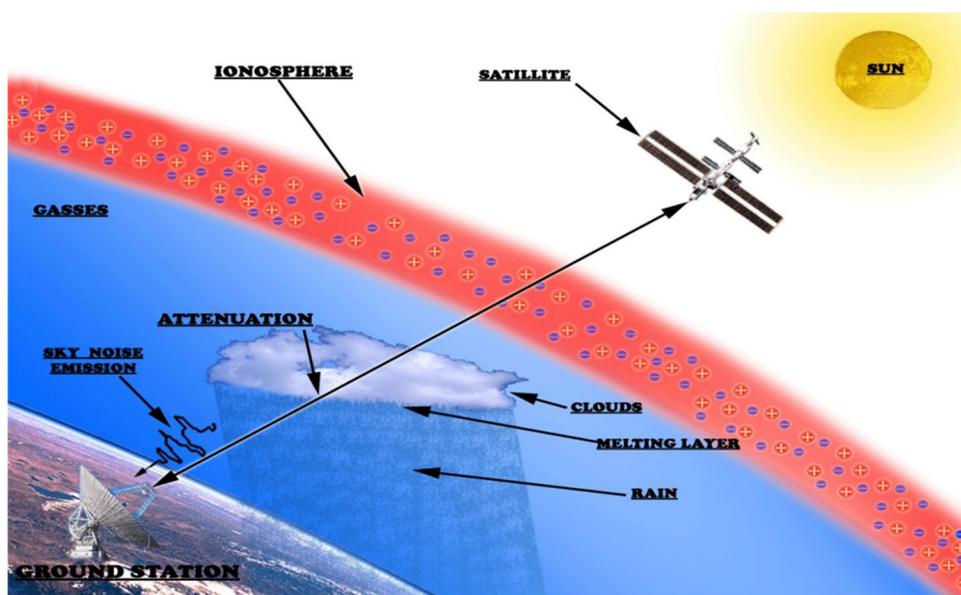
The absorption of microwave radio frequency signal is mainly due to rain or snow and is particularly predominant in frequencies over 10 GHz. It likewise alludes to the degradation of an actual signal caused by the electromagnetic obstruction of the main edge of a storm front, and it will happen during uplink or downlink of signal. It does not need a rain at an area for it to be influenced by rain fade. The actual signal may go through rain or snow somewhere while it is traveling over distances during its propagation. About 90% signal attenuation may likewise be caused mainly by rain, and the part of total attenuation is due to clouds, snow or ice on the downlink reception apparatus reflector, radome or horn. Therefore, numerous analysts are concentrating their consideration on setting up a solid strategy to anticipate rain-initiated attenuation. It has been realized that droplets can assimilate and disperse the vitality from the occurrence microwaves.

Figure 2 shows the climatic regions of the world and tropical regions between 30°N and 30°S. The regions are classified as tropical wet, dry, arid, humid and subtropical regions. Similarly, for the climatic regions of India, various regions were classified with long-term data analysis based on the sources from Indian Meteorological Department and is shown in Fig. 3. The further sections review about the propagation and attenuation research carried out using measurements campaigns.

Other propagation impairments

In addition to rain attenuation, the gaseous absorption which is caused mainly due to oxygen and water vapor contributes

Fig. 1 Earth–satellite slant path affected by propagation impairments. Sarat Kumar et al. (2008)



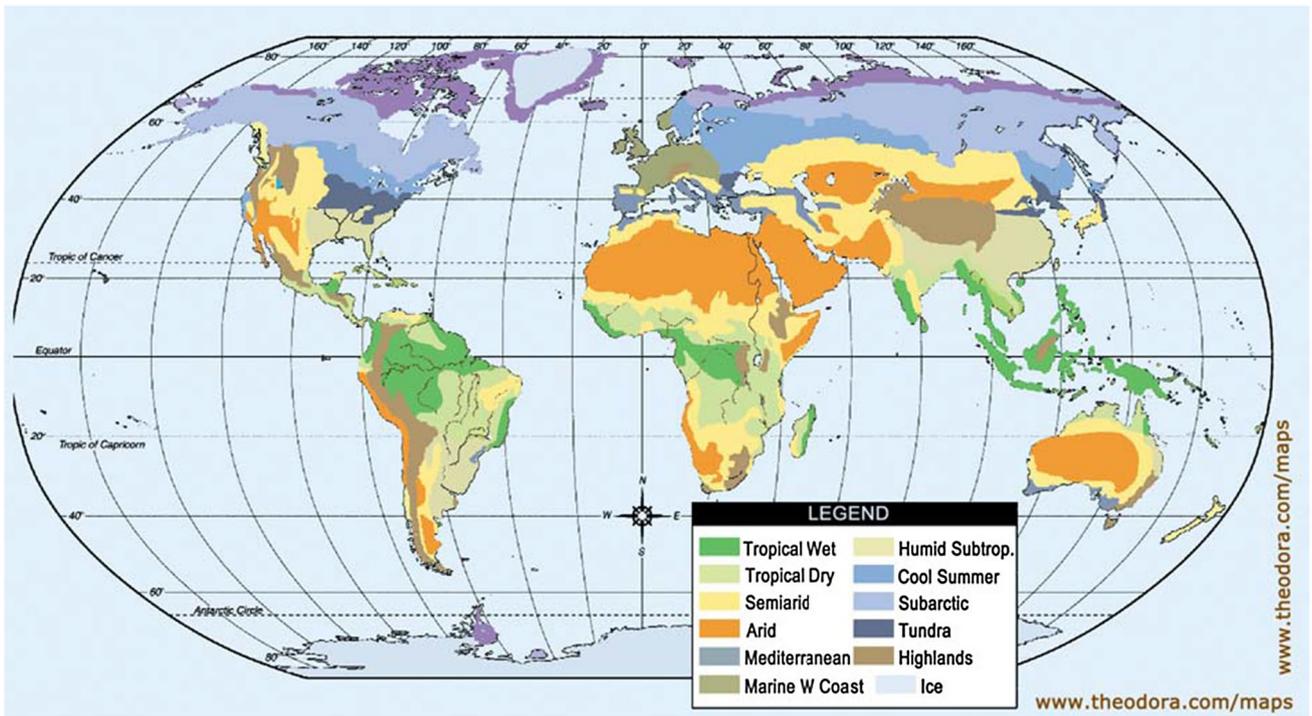


Fig. 2 Climatic regions of the world and tropical regions between 30°N and 30°S. Source: <http://theodora.com/maps>

Fig. 3 Climatic regions of India, classified with long-term data analysis. Source: Peel et al. (2007)



for the total attenuation, whereas the proportion of attenuation caused by the gases is very little, when compared to rain attenuation. From Fig. 4, it is clearly evident that frequency dependence of water vapor and oxygen absorption is exhibited regarding specific attenuation. The gaseous attenuation occurs is 22.5 GHz which is playing a role in Ka band of communications. Also, for oxygen absorption, the other maxima are observed at windows, one at 60 GHz and the other at 119 GHz frequencies. Further, the water vapor impact at particular frequency of 183 GHz can be seen in graph 1 and other maxima at 320 GHz. A total attenuation can be computed by combining all the specified attenuation for which the related temperature and humidity play a role and is specified in ITU-R.P.676-8 (2007).

Several efforts have been made by many researchers to estimate the level of degradation of terrestrial and satellite signals in Nigeria based on ITU models (Ajayi et al. 1996; Ajewole et al. 1999; Ojo et al. 2008, 2009; Ojo and Omotosho 2013; Ojo and Owolawi 2015). However, most of the investigations are based on the cumulative distribution of rain-induced attenuation, while very limited investigations have been focused on fade dynamics of rain attenuation statistics.

The main important cause of cloud attenuation is due to liquid water content present in the clouds. Models have been developed for computation of the cloud attenuation factor, and some of them can be seen at Salonen and Uppala (1991) and ITU-R.P.840-3 (1999).

This melting layer plays a role for the slant path attenuation at low elevation angles and the results are presented in Zhang et al. (1994) and this can be described as above certain height from the ground level, where ice particles be converted into water droplets and that particular point be usually called as melting layer.

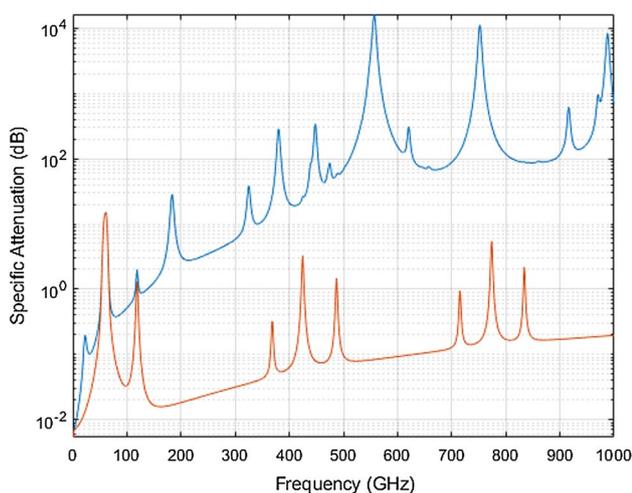


Fig. 4 Frequency versus specific attenuation, graph 1: global reference atmosphere and graph 2: dry atmosphere. Panagapoulous et al. (2004)

There is quite possibility for noise from external sources for a radio wave which is affecting the transmitting wave from space to ground station. Any regular retaining medium in the air which communicates with a radio wave will not just deliver a signal sufficiency diminishment, and it also contributes to thermal noise. As a result, receiver antenna temperature gets affected because of all these noise sources. Increase in attenuation results in the emissions noise (ITU-R.P.618-7 2002). Signal scattering from hydrometeors gives rise to noise considered to be more critical than attenuation noise at receiver.

The signal depolarization is mainly due to phase variation and the differential attenuation caused by non-spherical scatterers like rain droplets and ice particles. The importance of signal depolarization is mainly significant for systems reusing the frequencies due to transmission of perpendicularly polarized signals for the purpose of better utilization of spectrum. Part of the transmitted power in one polarization meddles with the other signal, and it is due to depolarization (Kanellopoulos and Clarke 1981).

Scintillations are nothing but the rapid variation in the signal strength with reference to the time and are mainly caused due to turbulence in the atmosphere. The tropospheric scintillations play an important role if the variations are observed for long time, and in such instances, it may also contribute to the overall attenuation. Intersystem interference particularly occurs between two satellites operating at similar frequency ranges or when they are in adjacent orbital positions.

Significant contributions—propagation research

Foreseeing the insights of rain attenuation from nearby rain rate measurements is surely well known. The rain rate measurements are themselves sensibly all-around described for the greater part of mid-latitudes, equatorial regions and less so for the subtropical, tropical and central areas. The measurements made at other regions where impact of terrain, restricted atmospheres and the typical year-to-year changes in precipitation all consolidates to present some unchangeable inconstancy in the phenomenon.

Europe has an expansive and all-around respected propagation investigate group, the greater part of whom work for academic and government laboratories, which is facilitated and financed by the European Space Agency (ESA) in a profoundly viable program that has been working since the mid-1970s. European research is additionally led through COST programs: COST 205, COST 255, and COST 280, IC0802. All these programs' purpose is to develop new models and organize the propagation campaigns across Europe. The details of the research

campaigns cover the *Impact of the Atmosphere on Earth–Satellite Radio Propagation at Frequencies Above 10 GHz* (Fedi 1985) and *Radio wave Propagation Modeling for Satcom Services at Ku Band or more* (Martellucci 2002) and COST Action 280 program covering the propagation impairment mitigation for millimeter wave radio systems with concentration on fade dynamics and satellite diversity aspects. The ESA satellite OLYMPUS gives coherent signals at 12.5, 20 and 30 GHz frequencies; the 20 GHz signal is switching between horizontal and vertical polarization. The Italian satellite ITALSAT F1, worked by Alenia Spazio, conveys a 20 GHz telemetry reference point and refined spread signals at 40 and 50 GHz frequencies.

Japanese analysts have spearheaded a lot of comprehension of satellite path propagation at frequencies over 10 GHz, especially through crafted by T. Oguchi at Communications Research Laboratory (CRL). Early Japanese utilization of Ku band, and the accessibility of reference points and transponders on the CS arrangement and BSE shuttle, gave chances to rain attenuation and depolarization estimations and uplink power control tests. Hogg and Chu (1975) discussed degradations produced by rain in satellite communication. Davies and Mackenzie (1981) reviewed the slant path propagation measurements using SHF/EHF links (Oguchi 1983), a most exhaustive review on the various aspects of rain and hydrometeors and their interaction with microwaves. Raina and Uppal (1981) made investigations identified with rain attenuation over New Delhi locale utilizing microwave radiometer amid the rainstorm time period and demonstrated examination between the attenuation appropriation and rain rate which demonstrates that for 0.01% of time, the attenuation quality surpassed 13.0 dB in relation to the surface precipitation rate of 130 mm per hour. For 0.1% of time (monsoon of 1978, 1979), the attenuation estimation of 8.0 dB supposedly corresponds to the surface precipitation rate of 80 mm/hr.

Nackoney and Davidson (1982) measured satellite–Earth rain attenuation at 11.7 GHz, utilizing CTS beacon signals, and (Crane 1985) has evaluated rain attenuation prediction models which are widely followed worldwide and his focus is mainly based on point rain rate distribution. Capsoni et al. (1987a, b) made a comprehensive methodology for the prediction of wave propagation at frequencies above 10 GHz, and Vogel (1982) conducted experiments at 11.2 GHz on rain fades at low elevation angles.

Arnold et al. (1981) presented the rain attenuation results obtained at three locations for over 2 years and analyzed the elevation angle, frequency, seasonal and diurnal effects along Earth–space paths (Raina and Uppal 1984); Mauri et al. (1986) provided the results of depolarization measurements at Italian locations. Arbesser-Rastburg and Brussaard (1993) detailed the propagation research in Europe

using Olympus satellite and reviewed past research carried with ATS-6 Satellite, OTS, Eutelsat and SIRIO Satellite. Sarkar et al. (1989) presented a brief description of the salient features of rain rate and rain cells extension over India. Barbara et al. (1993) performed experiments in Assam valley using six communication links in 6–11 GHz frequency range. Erwin et al. (2000) presented the analysis of CDFs of the received signal strength at Ka band using 18.7 GHz beam of ITALSAT Satellite by considering different environmental types.

Manabe and Yoshida (1995) described the rain attenuation characteristics of the radio lines and the possibility of outage free links using the experiments carried out in Tokyo at various frequencies. Dynamic rain attenuation model (van de Kamp 2002a, b) using simulation studies was described that how it can be used for fade mitigation techniques. Grémont and Filip (2004) emphasized the procedure for deriving the theoretical distributions for all propagation impairments and relevant quantities required for overall design of communication systems. Feldhake and Ailes-Sengers (2002) made comparison using 3 years of data from NASA ACTS experiment by using various rain attenuation models, and Crane (2002) examined the impact of rain droplets on the ACTS propagation terminal antenna which is an application of rain attenuation over Earth–satellite paths. Charles and Jaeger (2002) carried out propagation experiment at a diverse climate through North America for 5 years and analyzed the results of rain rate and attenuation for different frequencies.

Hasanuddin et al. (2002a, b) used several small antenna terminals in Japan for estimating the rain attenuation 10–14 GHz of frequencies. Khan et al. (2003) estimated the propagation losses for a 97 GHz line-of-sight terrestrial links. Sum et al. (2003) carried out analysis of rain attenuation at 23 GHz for terrestrial links, and Hatsuda et al. (2004) designed a measuring system useful for studying the site diversity characteristics of short- and long-distance links. Khajepuri et al. (2004) made studies in Iran and proposed a new method for calculating the rain fade and its variation for different rain rates at microwave range of frequencies. Maitra and Chakravarty (2005) conducted measurements of rain attenuation with Ku band signal from the geostationary satellite operating in India and developed a model applicable for rain attenuation. del Pino et al. (2005) used ITALSAT beacon at 50 GHz slant path measurements which are carried out at Spain, and results are compared with ITU-R and several attenuation models.

Suryana et al. (2005a, b) introduced 2-year measurements on simultaneous attenuation and tropospheric scintillations at an Indonesian location utilizing 12.247 GHz evenly enraptured JCSAT-3 satellite continuous-wave beacon signals. From the examination comes about, they discovered that on

a precipitation rate $R_{0.01}$ of 120 mm/h, the rain weakening $A_{0.01}$ measured was around 17 dB.

Many workers have calculated slant path attenuation using the point rainfall statistics (Manning 1984; Moupfouma, 1993). Pontes et al. (1990) made slant path radiometric measurements in Brazil; McCarthy et al. (1994) showed the cumulative distributions for slant path attenuation measurements using radiometers at 11.6 GHz in Cameroon, Kenya and Nigeria. Rachan and Prasit (1998) reported rainfall measurements and presented the results of 12 GHz fade duration statistics; Polonio and Riva (1998) made ITALSAT propagation measurements for the frequency bands 20, 40 and 50 GHz on attenuation, scintillation and fade duration.

Timothy et al. (2000) presented few of the comparisons from fade duration and interfade durations measured with a dual-slope Ku band radiometer as well as Ku band INTEL-SAT-IV beacon receiver. August et al. performed spectral analysis for long-term data of 49 years and tried to find its relation to fade dynamics and rainfall rate. Watson and Hu (1994) presented a methodology for prediction of attenuation for low fade margins. Nelson and Stutzman (1996) developed a simple model for estimating fade slope statistics for any frequency in the range of 10–30 GHz. Helmken et al. (1997) examined the fade duration measured at locations using ACTS propagation campaigns. Allnutt and Haidara (1998) presented cumulative statistics of diurnal fade durations and fade duration data from radiometric measurements in Equatorial Africa. Kumar et al. (2004) presented the results of 19.9 GHz radiometric propagation studies conducted over a period of 1 year at Amritsar, for determining the rain-induced zenith path attenuation, and presented the measurement results of rain-induced attenuation in the vertically polarized signal propagating at 19.4 GHz using LOS link at Amritsar, India.

Ventouras and Wrench (1999) presented statistical analysis of measured slant path characteristics at 20 GHz and 40 GHz in Southern England, and special emphasis on the seasonal and diurnal variations of attenuation statistics. Hansson (1990) developed a model for predicting rain attenuation from measured rainfall rate statistics using a new concept of moving rain cells. Experiments are conducted for locations in Europe, North America and East Asia for frequencies of 10–20 GHz and elevation angles in the range of 10–60 degrees.

Brooker (2004) patented the method and system for compensating the atmospheric fading in a communication system; Czarnecki (2000) proposed a fade countermeasure system for satellite links and simulated using data recorded using real-time events. Zhenwei et al. (2003, 2004) developed a rain prediction model for China based on the experimental data of Earth–space paths. Singliar et al. (2006) presented a comparative study based on measured point rain

intensity for various attenuation models for the location in Hungary.

Fiebig (2004) presented a time-series generator for use to model signal attenuation caused by meteorological effects in satellite-to-Earth links and Earth–satellite links. Arapoglou et al. (2004) developed a tool to illustrate the overall impact of rain fades on satellite communication systems. Chu and Chen (2005) presented the effects of rain on the performance of Ka band link in Taiwan. Pontes et al. (2005) presented a review using 18 years of rain attenuation data obtained from the measurements. Felix et al. (2006) analyzed the unavailability of DTH signal operating on Ku and Ka bands in Brazil using mathematical simulations, and Mohamad et al. (2006) conducted a study to identify the effect of rain attenuation on fixed wireless access system.

Research has been done by utilizing the satellite information got from TRMM and rain gauges, and the raindrop estimate conveyances were broken down by Baquero et al. (2005), with the end goal of disdrometer rain recovery. Hendratoro and Zawadzki (2003) analyzed the effect of vulnerability in the estimation of number thickness for little drops utilizing disdrometer frameworks on the exactness of the Y–Z and Y–R relations, and it is an application to ponder for figuring of rain weakening from radar reflectivity factor estimations. Lin and Chen (2002) inferred an exact equation for the figurings of electronic cross section and specific rain attenuation for rain drops in the recurrence scope of 0.6–100 GHz. da Silva Mello et al. (2002) conducted a 2-year campaign for the terrestrial links operating at 15 and 18 GHz frequencies. Timothy et al. (2002) drawn excellent conclusions using the method of moments technique for an application study of DSD. Walther and Terje (2003) portrayed the drop estimate dispersions relying upon the rain occasions, and other particular rain lessening forecast techniques have been produced utilizing the precipitation rate, the middle, or mode drop sizes of DSDs as a parameter for the arrangement. Relationship of debilitating estimations (Awang and Din 2004) using 38 GHz microwave joints at Malaysia with the DSD appear from Malaysia, Singapore, Brazil and Nigeria were impacted using lognormal model to illustrate. Maitra (2004) give huge data on how DSD can assume an imperative part in deciding the rain constriction, additionally featured significance of broad estimation necessities in the tropical area, and Park et al. (2002) presented another model for raindrop measure circulation, in light of the estimations performed in Chungnam National University, Korea, and the outcomes were acquired in great coordination with estimations. Maekawa et al. (2004) exhibited for the main year the consequences of rain lessening attributes of satellite connections utilizing equatorial atmospheric radar.

Chaisang and Hemmakorn (2000) reported an article on the study of rain rates and rain attenuation using THAICOM

downlink satellite signal. Yee et al. (2001) portrayed the estimation of microwave rain constriction over a 10-year time frame (1988–1998) in Singapore. The rain lessening of (vertically and on a level plane energized) microwaves at 15, 21 and 38 GHz spreading over a way length of around 1.1 km was recorded, and the model was produced in view of information gathered over a 10-year time span and through T-grid calculation of aggregate annihilation cross areas of raindrops. Abdul Rahim et al. (2002) explained rain attenuation which depends on information gathered over earthspace paths from the mid latitudes.

Ajose et al. (1995) investigated the problem of attenuation, phase rotation and depolarization of electromagnetic plane waves because to tropical rain using finite element method. Dissanayake (2002) built up cloud attenuation models on the premise of long-haul perception of overcast cover information and normal properties of a few sorts of mists that contain the fluid water which likewise tended to the issues identified with anticipating distinctive sorts of spread weaknesses and also joined them together to decide the effect on the satellite connections over an extensive variety of blackout probabilities. Panagopoulos and Kanellopoulos (2003) discussed a method to predict cumulative distribution of differential rain attenuation on the two converging links and examined the signal-to-noise ratio under many conditions. Panagopoulos et al. (2004) theoretically reviewed about the propagation impairments and mitigation techniques. Martellucci et al. (2000) investigated the spatial correlation of water vapor and liquid water content using radiometric observations. Liou (2000) presented the influence of atmospheric parameters on Ka band propagation links.

Polaries Baptista et al. (1989, 1990) using simple algorithms obtained the results for the combined effect of various propagation impairments and also studied the total attenuation statistics by using the data collected with five radiometers at three different frequencies. Brussard and Rogers (1990) have discussed theoretical aspects about the propagation consideration in satellite communication systems.

Amaya (2002), Amaya and Rogers (2002) conducted an analysis to evaluate the contribution of cloud and gases and also developed world maps of attenuation. Chandrasekar et al. (2003) used the space-born Ku band radar operating at 13.8 GHz global attenuation maps. Emiliani et al. (2004) developed the tools to plot the contour maps for parameters like rain rate and rain attenuation for a tropical location (Columbia) using ITU-R models. Begum and Otung (2008) have derived the cumulative distribution of rain rate for 1-min integration using 40 years of rainfall data. On top of the derived statistics, rain attenuation at 18.7 GHz is predicted using ITU-R model and a new rain attenuation prediction model is produced by implementing the reduction factor in prediction. Hence, the rainfall pattern, rain rate

distribution and rain attenuation characteristics at Bangladesh are calculated. Using the NSS-6 satellite, Panchal and Joshi (2016) studies have been carried out on performance analysis and simulation of rain attenuation models at Ku and Ka band over urban regions of India.

Attenuation effects—due to hydrometeors

At the point when the melting procedure begins, ice is bit-by-bit transformed into water, and the properties of hydrometeors change in terms of physical, dynamical, electrical and morphological aspects. Ekpenyong and Srivastava (1970), Hardaker (1992), Zhang et al. (1994), Stewart et al. (1984), Locatelli and Hobbs (1974), Mitra et al. (1990) developed models for melting layer phenomenon. In these models, liquefying particles are thought to be circular fit as a fiddle, and a couple of other polarimetric models are Russchenberg and Lighthart (1996), D'Amico et al. (1998) that additionally take into account the assessment of differential propagation quantities.

Karhu et al. (1993) calculated the mean cumulative distributions of rain attenuation using Leitao–Watson model for widespread rain and convective rains also the melting layer effects which are discussed. Mondal et al. (2001), Mondal and Sarkar (2003) gave the 0 °C isotherm height, which is valuable parameter required for the attenuation estimation at higher-frequency bands over Earth–space paths, and the experiments are conducted in India. The varieties of 0 °C isotherm stature for various stations corresponding to different seasons are displayed. The outcomes on H_i can be used for the estimation of attenuation of microwave and millimeter wave because of rain. Damodar (2005) considered rain information more than $1.5 \times 1.5^\circ$ lattice that has been figured over India, which is long-haul information measured by ECMWF for over 30 years.

Michelson and Liu (2009) developed a channel model for LEO propagation impairments, which predicts the rain attenuation along with the cloud, gases, fog and scintillation from the available parameters like rain rate, rain height, temperature, humidity, wind speed, wind direction and estimates of turbulence parameters from ITU-R. The output predictions of this model are analyzed to estimate the outage probability, fade dynamics and power spectrum and these predictions of fade slope distributions are compared with the ACTS program for Earth–GEO links and found to be satisfactory and these results will be very useful for planning Earth–LEO satellites at Ka band.

Mello and Pontes (2012) proposed a method for attenuation predictions, which will take the whole rainfall rate distribution rather than the conventional models, which rely on the percentage rain rate. This method is implemented in predicting the attenuation statistics both for satellite

communication and for terrestrial communication. The comparative analysis with different models developed at different regions shows that the proposed method was more effective than the ITU-R model in terrestrial links and in satellite links below 0.02% of the time.

Tropospheric scintillation studies

The rapid variation of signals because of little-scale refractive index inhomogeneities actuated by tropospheric turbulence along the propagation paths is generally known as tropospheric scintillations. It is clearly observed that this impact can be observed in hot season that too around noon and is because of strong turbulence in clouds. It is clearly understood that the scintillations effect is said to be increased due to increment in carrier frequency and also lowering the angle of elevation and size of the antenna. The effect of scintillations is even though less when compared to rain, their importance is treated as significant in long-term scenarios. Scintillations are for the most part connected with the existence of clouds, particularly cumulus and cumulonimbus mists.

Karasawa et al. (1988) derived a new prediction method for tropospheric scintillation from 14/11 GHz wave propagation experiment. Karasawa and Mastudo (1991) investigated the relation between rain attenuation and scintillation at low elevation angles using the experimental data during 1983 where scintillation cases are maximum. Hugues (1999) proposed a new method for estimation of scintillation effects on slant paths using radiosonde data. Catalán and Vilar (2002) tended to the long-standing issue of isolating the tropospheric scintillations for the prevailing pattern of attenuation in a satellite connection. Kassianides and Otung (2003) proposed a novel dynamic model of tropospheric scintillation, and results obtained from the two sites and three beacon frequencies showed good agreement with the theoretical predictions. Ong et al. (2003) gave a preliminary model for the designs of free-space optical links, taking rain and scintillation effects into an account. Mandeep et al. (2007, 2008) analyzed the measurements data and presented the results related to tropospheric scintillation.

Maitra et al. (2012) analyzed the propagation effects and the depolarization effects from the Ku band satellite data at 11.172 GHz and by incorporating the drop size distributions (DSD) of the rain obtained by the disdrometer in the SAM prediction model; the estimates are in good agreement with the measured data. The DSD of rain is further analyzed to get the correlation between the co-polar and cross-polar components of the signal and found the relation over the entire rain events except with the rain events having larger diameter of the rain drops, which is critical for the

enhancement of cross-polar component during pre-monsoon than in monsoon periods.

Mandeep and Islam (2012) evaluated the performance of the scintillation models over Malaysia region using the measured beacon signals from Superbird-C satellite. The ITU-R, Karasawa, Van de Kamp, Ortgies, Marzano MPSP, DPSP and Tervonen–Kamp–Salonen scintillation models are studied and implemented using the recordings at Tronoh, Malaysia. These models are developed based on the non-rainy measurement data and surface temperature parameters, which are in good relation to the scintillation. Thus, the model-based predictions present the accurate estimation than the empirical methods. Immadi et al. (2014a, b, c, 2015a, b, 2017a, b) fulfilled experimental studies on Ku band propagations for Vijayawada location. The equipment like disdrometer, Ku band antenna and micro-rain radar is used in the studies.

Models developed for rain attenuation

There have been a few endeavors to figure models that will allow propagation engineers to foresee the execution of satellite links working in the higher band of frequencies. Taking this further, if the systems so planned do not meet the particular necessities, the further point is to enable ameliorative strategies to be investigated.

Attenuation because of rain is a critical propagation problem which must be considered while designing the satellites. The attenuation due to rain relies upon DSD, temperatures, maximum speed and state of the raindrops. A few models were developed for estimation of the aggregate attenuation. Each of these models seems to have focal points and weaknesses relying upon particular application.

Attenuation due to rain can be measured precisely by methods for satellite reference point signs and radiometers. In any case, since engendering tests are completed just in few places on the Earth and for a predetermined number of frequencies and connection geometry, these outcomes cannot be straightforwardly connected to all destinations.

Therefore, a few attenuation models in view of physical actualities and utilizing accessible overall meteorological information have been created to give satisfactory. Specifically, the estimation of the rain-initiated attenuation starts with the aggregate value of precipitation as revealed by the numerous analysts: a few techniques have been produced and tried against accessible information to relate the local meteorological parameters to the signal attenuation measurements.

This area audits probably the most imperative rain forecast models grew all inclusive; among them, the best performing models are quickly depicted here under Table 1

which gives the model name and Table 2 which gives run-down of parameters utilized as a part of the models.

Angeletti and Lisi (2012) proposed a system model, combining the fade mitigation techniques used in satellite communications in order to provide the best link availability in Ka band communication satellites. The proposed system is a combination of the power control technique and the adaptive modulation and coding technique, where the combination results in maintaining the link at low SNR

and providing the minimum data rate to the users. This approach is highly recommended for providing broadband services using Ka band spectrum.

Yussuff and Khamis (2012) derived the relation between the rain rate and the effective slant path of rain and have modified the effective length used in ITU-R model by the derived ones. The rain rate is taken from the TRMM satellite data, and this rain rate is used to calculate the effective height of the rain. The proposed model is compared with the

Table 1 Various rain attenuation models and relevant parameters

#	Model name	λ	H_s	θ	f	r	k, α	P_0	$Rp(p)$	$R_{0.1}$	$R_{0.01}$	h_R	$H_{0(p)}$	H_{-15}	A_p
1	Assis—Einloft-improved model (Costa 1983)	✓	✓	✓			✓		✓			✓			
2	Australian model (Flavin 1996)	✓	✓	✓	✓	✓	✓	✓	✓						
3	Brazil model (Recommendation CCIR 1992)			✓	✓		✓		✓						
4	Bryant model (Bryant et al. 2001)			✓	✓		✓		✓						
5	Crane global model (Crane and Blood 1979)			✓	✓	✓			✓				✓		
6	Crane two-component model (Crane 1985)	✓	✓	✓			✓		✓						
7	EXCELL model (Capsoni 1987a, b)	✓	✓	✓			✓		✓			✓			✓
8	Garcia model (del Pino et al. 2005)	✓	✓	✓			✓		✓		✓				
9	The ITU-R models (ITU-R.P Series)	✓	✓	✓	✓		✓				✓	✓			✓
10	Karasawa model (Karasawa et al. 1988)	✓	✓	✓			✓					✓			✓
11	Leitao–Watson model (Watson et al. 1985)			✓	✓	✓			✓			✓			✓
12	Matricciani model (Matricciani 1991)	✓	✓	✓	✓	✓			✓						
13	Misme–Waldteufel model (Misme and Waldteufel 1980)		✓	✓			✓	✓	✓				✓	✓	
14	Simple attenuation model (Stutzman and Dishman 1982)	✓	✓	✓			✓		✓						
15	Svjatogor model (Svjatogor 1985)		✓	✓			✓		✓						
16	DAH model (Dissanayake et al. 1997, 2001)	✓	✓	✓	✓		✓								✓

Description of model parameters is included in Table 2

Table 2 Description of various model parameters

Symbol	Description
λ	Wave length (in m)
H_s	Altitude of the Earth station [km]
θ	Slant path elevation angle [deg]
f	Frequency of the link [GHz]
r	The specific attenuation derived using the frequency-dependent coefficients [dB/km]
k, α	Frequency- and polarization-dependent coefficients given by ITU-R to calculate Specific attenuation due to rain [ITU-R]
P_0	Probability of rain falling at a point
$Rp(p)$	Rainfall rate exceeded for any given percentage of the average year, p
$R_{0.1}$	The rain intensity 0.1% of an average year [mm/h]
$R_{0.01}$	The rain intensity 0.01% of an average year [mm/h]
h_R	Rain height [km]
H_0	The 0° isotherm height [km]
H_{-15}	Average – 15 °C isotherm height [km]
$p(A_s)$	The probability of attenuation
$p(R_p)$	The probability of the rain rate
h_0	Yearly average 0 °C isotherm height above mean sea level [km]
A_p	The predicted attenuation [dB]

global models like DAH, SAM, Mandeep and ITU-R and found that the proposed model (modified ITU-R) was nearer to the measured ones and the mean square error, standard deviation and root-mean-square error are less than the other models.

Cerqueira and Assis (2013) have carried the analysis of the data collected in 8 geographical sites in Amazon region, South America. In this work, mainly three conclusions have been drawn from the measurements: firstly, the relation between the rain event duration and the rain rate is found and they are in good correlation with each other; secondly, the rain rate and the average rain event duration are also in correlation and the equations for the relations have been derived, which are useful to estimate the duration of the rain event, when the rain rate is known. The analysis can be used to design a satellite system in Ku band and Ka band. Yussuff and Khamis (2013) compared the measured signal attenuation values, recorded at 3 stations in Nigeria with the global models like ITU-R, SAM and DAH. The global models are not in good agreement with the measured attenuation. Based on the regression coefficients, derived using the statistical methods applied to the measured data, the team has modified the coefficients used in ITU-R and found that the prediction given by the proposed model was in good agreement with less error percentage than the global models.

Kestwal et al. (2014) have compared the measured rain attenuation at Almora, Uttarakhand, with the globally applied model ITU-R. The rainfall data of 20 years have been taken and are converted into rain rate for calculating the rain attenuation using SAM model. The results show that the ITU-R model is performing well at low frequencies and at low rain rates and deviates from the measured attenuation values at higher frequencies and higher rain rates. So an exponential fit is used for the results for evaluation of the correlation between the rain rate and the rain attenuation. The other work by the team is to use the terrestrial communication link power to estimate the rain rate data, which will be helpful in designing the rain rate maps, where the equipment is not available. Das and Maitra (2015) tested the model, proposed by them in 2012, where the model is developed on the basis of the data collected at tropical region, Kolkata. This model is tested with the other tropical site data and found to be suitable in predicting the attenuation. On the other hand, the model is tested with the temperate region Spino d'Adda data, where the region is climatically different from tropical region and found to be not suitable for temperate regions. So the parameters required for the model μ and σ have to be derived from the data collected from temperate regions such that the model is applicable globally.

Semire et al. (2015) proposed two models for prediction of the long-term rain attenuation and statistical prediction of site diversity gain. The rain attenuation model is derived from the global model of ITU-R, modifying the parameters

of horizontal and vertical reduction factors by analyzing the rain attenuation data and atmosphere radar data taken from Malaysian Meteorological Department. The site diversity gain model is derived from the rain rate attenuation statistics, calculated from TRMM measurements and analyzing the data for 5 sites, placed at a distance of 1 km to 50 km randomly from the measured site. The proposed models performed well when compared with the existing models. The models can be used to predict the rain attenuation and diversity gain in tropical regions, which have climatic conditions like Malaysia.

Shrestha and Choi (2017) presented the measured rain attenuation for Ku and Ka band frequencies for a period of 3 years (2013–2015). The measured signals are at 12.25 GHz, 19.8 GHz and 20.73 GHz. Rain attenuation is estimated by using the global ITU-R model and SAM model from rain rate measurements from disdrometer. The comparison between the measured and estimated attenuation values by ITU-R and SAM model were presented and the results are not matched with measured ones. In order to improve the ITU-R frequency scaling model, prediction error is introduced by considering the absolute values of deviation. More testing and analysis have to be done to this method to be applicable in other regions.

Rakshit et al. (2017) analyzed the rain decay parameter, used in global SAM model and modeled for the region of Kolkata. In this work, the variation of decay parameter with respect to rain rate is investigated for the data over 2 years of time (2008–2009). The median of decay parameter and median absolute deviation of gamma are calculated using the rain rate based on the curve-fit parameters. Decay parameter can now be dynamically calculated for different rain rates and is used in the SAM model for predicting the rain attenuation. The proposed model is evaluated using the measured data in the year of 2010, and the model is accurate in predicting the rain attenuation at different rain rates than the global SAM and ITU-R models. The technique embedded in this model is very effective when predicting from the ground-based rain measurements.

Arslan et al. (2018) proposed a technique to estimate the amount of rainfall based on the satellite link attenuation measurement. Two years of Ku band satellite measurements are used to model the relationship between the attenuation in microwave link and the rain rate. Simultaneously, the rain rate has been estimated by the radar reflectivity provided by the NOAA operated by Next-Generation Weather Radar (NEXRAD) and also the ground-measured rainfall accumulation by the rain gauge. The rainfall accumulation estimated by the ITU-R A-R relationship using Ku band satellite signal attenuation agrees well with the estimates from radar Z-R relationship for 6 mm rainfall accumulation and above, for 1 day (24 h). For rainfall accumulations less than 6 mm, the ITU-R estimates are 30–43% less than the radar estimates.

This technique has to be refined to have fair estimates from the satellite signal attenuations.

Halder et al. (2018) carried out the rain attenuation estimation using two different ground-based recordings from disdrometer and radiometer at two tropical locations: Kolkata (22.57 N, 88.36 E, India) and Belem (1.45 S, 48.49 W, Brazil). The radiometer measures brightness temperature measurements at 14 different frequencies in 2 frequency bands of 22.24–31.40 GHz and 50–59 GHz, whereas the disdrometer measures the rain drop size and number of rain drops, from which the rain attenuation has been calculated. The results show that the estimated rain attenuation values below 30 mm/h rain rate are matching with the two types of ground-based recordings: radiometer and disdrometer, whereas at higher rain rates, the DSD estimated rain attenuation matches the ITU-R model estimates. This approach will be useful in estimating the rain attenuation in different rain conditions over tropical region at Ka band.

Conclusions

Rain attenuation models developed worldwide using number of measurement campaigns are discussed in this review paper. This will help the propagation engineers and serve as a single source of reference. An idea of compilation and usage of all theoretical models is made for rain attenuation studies, for the purpose of predicting attenuation in an effective way. Various models were discussed and listed with the variability of the rain rate, rain height and other meteorological parameters.

Attenuation measurements and prediction of results from simulations on Ku and Ka band satellite signal are fulfilled by various researchers. Rain attenuation prediction by ITU-R and the other models showed noticeable deviation to the measured values. Unlike the monotonic decrease predicted by these models, exceedance of rain rate and attenuation in India and other tropical regions showed the presence of breakpoints. It is often necessary to have a good understanding of a model's implementation to use it correctly, otherwise it may produce errors or, worse, wrong results when supplied with parameters outside of an expected validity range.

The most important issue that exists today affecting the empirical models is certainly that not enough information exists in propagation measurement databases to allow for the complete modeling of the attenuation phenomenon throughout all the different climatic regions, worldwide. Therefore, researchers use proxy procedures to fill voids in sites for which data are unavailable.

The lack of rain measurement data from tropical regions for verification or for modeling purposes causes predictions obtained from existing models to deviate from real

measurements. Only a few tropical countries in the world have submitted rain and attenuation measurements to ITU study groups for inclusion in the world database. However, these submissions are not sufficient to reduce the uncertainty that exists when a system designer uses a prediction method to estimate rain margins in other tropical countries.

An inaccurate prediction model can also lead to underestimating the effect of rain on the links, severely impacting end users because of lower-than-expected availability. The severity of the impact on a service depends on the criticality of the applications being utilized by underlying end user. Residential users accessing the Internet may be more permissive than users of mission critical applications, such as telemedicine or tele-education, which normally cannot be rescheduled without impacting the community.

Unless the measurements from the local stations, the results obtained always run the risk of overestimating or underestimating the phenomenon. It has to be noted that the statistical nature of the rain phenomenon and the inherited statistical behavior of the attenuation phenomenon are such that the variability between yearly cumulative functions (measured via comparison of equiprobable values in empirical cumulative distribution functions for each year) and between seasons (measured via comparison of equiprobable values from empirical cumulative distribution functions for different seasons, i.e., summer-to-fall or spring-to-summer, as in temperate zones, or rainy and dry seasons, as in tropical zones) is comparable to the deviation between predictions and measurements, therefore putting all models at the same order of precision. This uncertainty is reduced by the analysis of long-term data sets (> 10 years) and by comparisons including the expected variability based on what is observed.

Even though various campaigns are done all over the world, the attenuation is the phenomenon which varies from location to location and the experimental results obtained from the researchers are not directly useful for the prediction of attenuation at other locations. A model suitable to the particular location by taking into account the local parameters for predictions is very much required.

The major conclusions can be drawn from all the studies specified are:

- Attenuation due to rain and other hydrometeors is a crucial limitation in satellite link design
- Several observational and non-exact rain constrictor expectation models that have been produced depend on the deliberate information acquired from calm districts not reasonable for tropical locales, i.e., high precipitation areas.
- The ITU-R model when compared with the other models either under estimates or over estimates especially for a tropical region measurement sites.

- Other debilitations are because of vaporous ingestion, cloud, troposphere refractive impacts, glitter and wet receiving wire, making crumbling the higher-frequency bands.
- About 0.5–40 dB attenuation during heavy rainfall.
- More propagation studies have to be conducted in future for better prediction of propagation losses. There is a continued need of obtaining and improving the global models for rain attenuation over the Earth–space paths.
- The oxygen lessening is practically consistent at tropical destinations (the temperature and weight varieties are little). In most northerly locales, the weakening varieties of oxygen and water vapor are of a similar request. (The measure of water vapor depends firmly on temperature, while the oxygen constriction depends contrarily on the temperature.)
- The utilization of meteorological information gathered without rain can prompt an underestimation of aggregate lessening as for the qualities acquired utilizing radio-sonde estimations gathered both in rain and in non-rain conditions.
- Studies and campaigns at various locations and at different frequencies were done and many a researchers from their measurements the impact of frequency dependence on rain attenuation at microwave frequencies are presented.
- In addition, the procedures and models are needed for attenuation predictions for overall design of communication systems.
- The campaigns like ACTS, COST programs are very much useful for consolidation of various models and parameters along with fade dynamics and propagation modeling. This in turn helped to resort the development of techniques for site diversity and adaptive mitigation techniques.
- Various researchers conducted the experiments, and propagation data resulted from these are useful for development of rain attenuation maps.
- Rain rate and rain attenuation models are developed based on extraction of attenuation for various locations.

The future campaigns proposed related to propagation impairments may provide the outcomes like:

- Collecting the multi-years Ku, Ka and V band propagation data
- Performing multiple site analysis for rain attenuation and scintillation studies
- Fade dynamics and modeling
- Adaptive control techniques for uplink and downlink

Acknowledgements The authors particularly thank the funding given from Science and Engineering Research Board, Ministry of Science and Technology (DST), Government of India, under EMR grants with F. No: EMR/2015/000100. The authors likewise thank the administration of Koneru Lakshmaiah Education Foundation (KL University) for supporting and empowering this work by giving the facilities in Center for Applied Research in Electromagnetics (CARE), Department of Electronics and Communication Engineering.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Abdul Rahim SK, Sum CS, Din J, Rahman TA, Aziz ZAA, Awang A (2002) Rain attenuation study over terrestrial and earth satellite links in Malaysia. In: International union of radio science XXVIIIth general assembly. Maastricht, Holland, pp 1483–1487
- Ajayi GO, Feng SR, Reddy BM (1996) Handbook on radio propagation related to satellite in tropical and subtropical countries 58
- Ajewole MO, Kolawole LB, Ajayi GO (1999) Theoretical study of the effect of different types of tropical rainfall on microwave and millimeter-wave propagation. *Radio Sci* 34(5):1103–1124
- Ajose SO, Sadiku NO, Goni U (1995) Computation of attenuation, phase rotation, and cross-polarization of radio waves due to rainfall in tropical regions. *IEEE Trans Antennas Propag* 43(1):1–5
- Allnutt JE, Haidara F (1998) Ku-band diurnal rain fade statistics from three, two-year, earth–space experiments in equatorial Africa. In: Proceedings of URSI commission F open symposium climatic parameters in radiowave propagation prediction, pp 27–29
- Amaya C (2002) Impact of clouds and gases on satcom links at Ka and EHF bands. In: 20th AIAA international communication satellite systems conference and exhibit, American Institute of Aeronautics and Astronautics, Montreal, Quebec, Canada
- Amaya C, Rogers DV (2002) Characteristics of rain fading on Ka-band satellite-earth links in a Pacific maritime climate. *IEEE Trans Microw Theory Tech* 50(1):41–45
- Angeletti P, Lisi M (2012) A systemic approach to the compensation of rain attenuation in Ka-band communication satellites. *Int J Microw Sci Technol* 2012:1–7
- Arapoglou P-DM, Panagopoulos AD, Chatzarakis GE, Kanellopoulos JD, Cottis PG (2004) Diversity techniques for satellite communications: an educational graphical tool. *IEEE Antenna Propag Mag* 46(3):109–114
- Arbesser-Rastburg BR, Brussaard G (1993) Propagation research in Europe using the OLYMPUS satellite. *Proc IEEE* 81(6):865–875
- Arnold H, Cox D, Rustako A (1981) Rain attenuation at 10–30 GHz along earth–space paths: elevation angle, frequency, seasonal, and diurnal effects. *IEEE Trans Commun* 29(5):716–721
- Arslan CH, Aydin K, Urbina JV, Dyrud L (2018) Satellite-link attenuation measurement technique for estimating rainfall accumulation. *IEEE Trans Geosci Remote Sens* 56(2):681–693
- Awang MA, Din J (2004) Comparison of the rain drop size distribution model in tropical region. In: RF and microwave conference, 2004. RFM 2004. Proceedings, IEEE, pp 20–22
- Baquero M, Cruz-Pol S, Bringi VN, Chandrasekar V (2005) Rain-rate estimate algorithm evaluation and rainfall characterization in tropical environments using 2DVD, rain gauges and TRMM data. In: International geoscience and remote sensing symposium, vol 2, p 1146

- Barbara AK, Devi M, Timothy KI, Sharma S (1993) Microwave propagation in relation to atmospheric parameters over different terrains of Assam Valley. In: Geoscience and remote sensing symposium, 1993. IGARSS'93. Better understanding of earth environment., International, IEEE, pp 261–263
- Begum S, Otung IE (2008) Characterization of rain attenuation in Bangladesh and application to satellite link design. *Radio Science* 43(01):1–16
- Brooker RL (2004) U.S. Patent No. 6,813,476. U.S. Patent and Trademark Office, Washington, DC
- Brussaard G, Rogers DV (1990) Propagation considerations in satellite communication systems. *Proc IEEE* 78(7):1275–1282
- Bryant GH, Adimula I, Riva C, Brussaard G (2001) Rain attenuation statistics from rain cell diameters and heights. *Int J Satell Commun* 19(3):263–283
- Capsoni C, Fedi F, Paraboni A (1987a) A comprehensive meteorologically oriented methodology for the prediction of wave propagation parameters in telecommunication applications beyond 10 GHz. *Radio Sci* 22(3):387–393
- Capsoni C, Fedi F, Magistrone C, Paraboni A, Pawlina A (1987b) Data and theory for a new model of the horizontal structure of rain cells for propagation applications. *Radio Sci* 22(03):395–404
- Catalán C, Vilar E (2002) Simultaneous analysis of downlink beacon dynamics and sky brightness temperature. Part II: extraction of amplitude scintillations. *IEEE Trans Antennas Propag* 50(4):535–545
- Cerqueira JL, Assis MS (2013) Rainfall rate duration study for performance assessment of satellite communication links. *Int J Microw Sci Technol* 2013:1
- Chaisang A, Hemmakorn N (2000) Satellite due to rain fading at Klong downlink availability of Thaicom 3 Yai ground station. In: TENCON 2000. Proceedings, vol 1, IEEE, pp 106–108
- Chandrasekar V, Fukatsu H, Mubarak K (2003) Global mapping of attenuation at Ku-and Ka-band. *IEEE Trans Geosci Remote Sens* 41(10):2166–2176
- Charles EM, Jaeger BE (2002) Rain attenuation model comparison and validation. *Online J Space Commun* (2):1–29. <http://www.spacejournal.org/>
- Chu CY, Chen KS (2005) Effects of rain fading on the efficiency of the Ka-band LMDS system in the Taiwan area. *IEEE Trans Veh Technol* 54(1):9–19
- Costa E (1983) An analytical and numerical comparison between two rain attenuation prediction methods for earth–satellite paths. In: ESA wave propagation and remote sensing, (SEE N83-29494 18-32), pp 213–218
- Crane RK (1985) Evaluation of global and CCIR models for estimation of rain rate statistics. *Radio Sci* 20(4):865–879
- Crane RK (2002) Analysis of the effects of water on the ACTS propagation terminal antenna. *IEEE Trans Antennas Propag* 50(7):954–965
- Crane RK, Blood DW (1979) Handbook for the estimation of microwave propagation effects—link calculations for Earth-Space Paths. Environmental Research and Technology Rpt. No. 1, DoC. No. P-7376-TRL
- Czarnecki M (2000) Compensation of rain attenuation for ka-band satellite systems. In: 13th international conference on microwaves, radar and wireless communications. MIKON-2000, vol 2. IEEE, pp 439–442
- D'Amico MM, Holt AR, Capsoni C (1998) An anisotropic model of the melting layer. *Radio Sci* 33(3):535–552
- Damodar M (2005) Applications of propagation models to design geostationary satellite links operating in Ka-band over indian rain zones. In: Proceedings of the 28th URSI general assembly, India
- da Silva Mello LAR, Costa E, de Souza RSL (2002) Rain attenuation measurements at 15 and 18 GHz. *Electron Lett* 38(4):197–198
- Das D, Maitra A (2015) Rain attenuation prediction during rain events in different climatic regions. *J Atmos Solar Terr Phys* 128:1–7
- Davies PG, Mackenzie EC (1981) Review of SHF and EHF slant path propagation measurements made near Slough (UK). In: IEE Proceedings H (microwaves, optics and antennas), vol 128, no. 1. IET Digital Library, pp 53–65
- del Pino PG, Riera JM, Benarroch A (2005) Slant path attenuation measurements at 50 GHz in Spain. *IEEE Antennas Wirel Propag Lett* 4(1):162–164
- Devika SV, Kotamraju SK, Kavya KCS, Kumar VS, Suhas K, Vinu K, Anudeep B (2016) A circularly polarized Ka-band antenna for continuous link reception from GSAT-14. *Indian J Sci Technol* 9(38)
- Devika S, Karki K, Kotamraju SK, Kavya K, Rahman MZ (2017) A new computation method for pointing accuracy of Cassegrain antenna in satellite communication. *J Theor Appl Inf Technol* 95(13)
- Dissanayake A (2002) Ka-band propagation modeling for fixed satellite applications. *Online J Space Commun* 2:1–5
- Dissanayake A, Allnut J, Haidara F (1997) A prediction model that combines rain attenuation and other propagation impairments along Earth Satellite Paths. *IEEE Trans Antennas Propag* 45(10):1546–1558
- Dissanayake A, Allnut J, Haidara F (2001) Cloud attenuation modeling for SHF and EHF applications. *Int J Satell Commun* 19(3):335–345
- Dutton EJ, Dougherty HT, Martin RF Jr (1974) Prediction of European rainfall and link performance coefficients at 8 to 30 GHz. Institute for Telecommunication Sciences, Boulder
- Ekpenyong BE, Srivastava RC (1970) Radar characteristics of the melting layer: a theoretical study. University of Chicago, Department of the Geophysical Sciences, Chicago
- Emiliani LD, Agudelo J, Gutierrez E, Restrepo J, Fradique-Mendez C (2004) Development of rain-attenuation and rain-rate maps for satellite system design in the Ku and Ka bands in Colombia. *IEEE Antennas Propag Mag* 46(6):54–68
- Erwin K, Fontan FP, Angeles Vazquez Castro M, Buonomo S, Arbesser-Rastburg BR, Baptista JPVP (2000) Ka-band propagation measurements and statistics for land mobile satellite applications. *IEEE Trans Veh Technol* 49(3):973–984
- Fashuyi MO, Afullo TJ (2007) Rain attenuation prediction and modeling for line-of-sight links on terrestrial paths in South Africa. *Radio Sci* 42(5):1–15
- Fedi F (1985) Special issue on the COST 205 project on earth–satellite radio propagation above 10 GHz. Association of Elettrotec. ed Electron. Italiana
- Feldhake GS, Ailes-Sengers L (2002) Comparison of multiple rain attenuation models with three years of Ka band propagation data concurrently taken at eight different locations. *Online J Space Commun*. <http://www.spacejournal.org/>
- Felix LEM, Pontes MS, Dhein NR, Migliora CG (2006) Unavailability due to rain of VSAT networks operating in Ka and Ku bands in Brazil. *Int J Satell Commun Netw* 24(3):203–213
- Fiebig U-C (2004) A time series generator modeling rain fading. Institute of communication and navigation, German Aerospace Center, DLR Report—2004
- Flavin RK (1996) Satellite link rain attenuation in Brisbane and a proposed new model for Australia. In: Telstra Research Laboratories, Report, (8375)
- Goldhirsh J, Dockery D (2004) Propagation characteristics for coastal region of South Korea and their impact on communication systems. In: Military communications conference, 2004. MILCOM 2004. 2004 IEEE, vol 1. IEEE, pp 460–465

- Grémont BC, Filip M (2004) Spatio-temporal rain attenuation model for application to fade mitigation techniques. *IEEE Trans Antennas Propag* 52(5):1245–1256
- Halder T, Adhikari A, Maitra A (2018) Rain attenuation studies from radiometric and rain DSD measurements at two tropical locations. *J Atmos Solar Terr Phys* 170:11–20
- Hansson L (1990) New concept used to predict slant path rain attenuation statistics. In: *IEEE proceedings H (microwaves, antennas and propagation)*, vol. 137, no. 2. IET Digital Library, pp 89–93
- Hasanuddin ZB, Fujisaki K, Ishida K, Tateiba M (2002a) Measurement of Ku-band rain attenuation using several VSATs in Kyushu Island, Japan. *IEEE Antennas Wirel Propag Lett* 1(1):116–119
- Hasanuddin ZB, Fujisaki K, Ishida K, Tateiba M (2002b) Measurement of Ku-band rain attenuation using several VSATs in Kyushu Island, Japan. *IEEE Antennas Wirel Propag Lett* 1(1):116–119
- Hatsuda T, Aoki Y, Echigo H, Takahata F, Maekawa Y, Fujisaki K (2004) Ku-band long distance site-diversity (SD) characteristics using new measuring system. *IEEE Trans Antennas Propag* 52(6):1481–1491
- Helmken H, Henning RE, Feil J, Ippolito LJ, Mayer CE (1997) A three-site comparison of fade-duration measurements. *Proc IEEE* 85(6):917–925
- Hardaker PJ (1992) A study of the melting layer in single polarisation radar echoes with application to operational weather radar (Doctoral dissertation, University of Essex)
- Hendratoro G, Zawadzki I (2003) Derivation of parameters of YZ power-law relation from raindrop size distribution measurements and its application in the calculation of rain attenuation from radar reflectivity factor measurements. *IEEE Trans Antennas Propag* 51(1):12–22
- Hogg DC, Chu TS (1975) The role of rain in satellite communications. *Proc IEEE* 63(9):1308–1331
- Hugues V (1999) Prediction of tropospheric scintillation on satellite links from radiosonde data. *IEEE Trans Antennas Propag* 47(2):293–302
- Immadi G, Kotamraju SK, Khan H, Venkata Narayana M, Jaya Krishna Pooja M (2014a) Implementation of data logging as a part of propagation impairment studies of Ku band satellite signal with the establishment of low cost experimental setup. *Int J Appl Eng Res* 9(3):355–367
- Immadi G, Kotamraju SK, Khan H, Venkata Narayana M, Hemavasavi K, Pooja Naga Sai K, Sirisha N (2014b) Estimation of Ku band satellite signal propagation impairment due to rain in tropical environment using ITU-R. *Int J Appl Eng Res* 9(20):7149–7168
- Immadi G, Kotamraju SK, Khan H, Venkata Narayana M (2014c) Rain rate-radar reflectivity relationship for drop size distribution and rain attenuation calculation of Ku band signals. *Int J Eng Technol* 6(2):815–824
- Immadi G, Kotamraju SK, Venkata Narayana M, Khan H, Sreemadhuri A, Sravya Chowdary K, Vineela P (2015a) Measurement of tropospheric scintillation using Ku band satellite beacon data in tropical region. *J Eng Appl Sci* 10(4):1568–1575
- Immadi G, Kotamraju SK, Venkata Narayana M, Rajkamal K, Habibulla K, Viswanath G, Avinash I (2015b) Measurement of rain attenuation for Ku band satellite signal in tropical environment using DAH, SAM models. *RPN J Eng Appl Sci* 10(4):1717–1725
- Immadi G, Narayana MV, Kotamraju SK, Kavya K, Maneesha S, Ravali K, Sravani C (2017a) Computation of effects of troposphere on Ku band down link signal in tropical regions. *J Theor Appl Inf Technol* 95(9):2078–2087
- Immadi G, Narayana MV, Kotamraju SK, Sarvani T, Manasa T, Yaswant CV, Kalyan JA (2017b) Computation of attenuation due to rain for Ku band frequencies using DSD for the tropical region. *J Theor Appl Inf Technol* 95(10)
- Islam MR, Chebil J, Tharek AR (1999) Frequency scaling of rain attenuation from 23-to 38-GHz microwave signals measured in Malaysia. In: *Microwave conference, 1999 Asia Pacific*, vol 3. IEEE, pp 793–796
- ITU-R.P.618-3-5-7-8-9 (1994-1997-2002-2004-2007) Propagation data and prediction methods required for the design of earth-space telecommunication systems. Recommendations International Telecommunications Union
- ITU-R.P.676-4-7 (1999-2007) Attenuation by atmospheric gases in the frequency range 1–350 GHz. Recommendations International Telecommunications Union
- ITU-R.P.840-3 (1999) Attenuation due to clouds and fog. Recommendations International Telecommunications Union
- John Philip B, Kotamraju SK, Sri Kavya KC, Madhumitha R, Pavan Kumar A (2017) Performance evaluation of attenuation time series generators over Indian region. *J Adv Res Dyn Control Syst* 2017(Special Issue 2):48–55
- Kalyan SSS, Kavya KCS, Kotamraju SK (2017) A reconfigurable beam steering linear phased array antenna for Ku band satellite communication using graphing method. *J Adv Res Dyn Control Syst* 2012:63–71
- Kanellopoulos JD, Clarke RH (1981) A study of the joint statistics of rain depolarization and attenuation applied to the prediction of radio link performance. *Radio Sci* 16(2):203–211
- Karasawa Y, Matsudo T (1991) Characteristics of fading on low-elevation angle earth-space paths with concurrent rain attenuation and scintillation. *IEEE Trans Antennas Propag* 39(5):657–661
- Karasawa Y, Yamada M, Allnutt JE (1988) A new prediction method for tropospheric scintillation on earth-space paths. *IEEE Trans Antennas Propag* 36(11):1608–1614
- Kassianides CN, Otung IE (2003) Dynamic model of tropospheric scintillation on earth-space paths. *IEE Proc Microw Antennas Propag* 150(2):97–104
- Karhu S, Salonen E, Hyvonen R, Uppala S, Baptista JP (1993) Prediction of rain attenuation at low-availabilities using models and data of widespread and convective rains. In: *Eighth international conference on antennas and propagation*, vol 1993. IET, pp 56–59
- Kestwal MC, Joshi S, Garia LS (2014) Prediction of rain attenuation and impact of rain in wave propagation at microwave frequency for tropical region (Uttarakhand, India). *Int J Microw Sci Technol* 2014:1
- Khajepuri F, Ghorbani A, Amindavar H (2004) A new method for estimating rain attenuation at high frequencies. In: *Proceedings of the 3rd International conference on computational electromagnetics and its applications*, 07803-8562-4/2004. IEEE, pp 76–79
- Khan SA, Tawfik AN, Gibbins CJ, Gremont BC (2003) Extra-high frequency line-of-sight propagation for future urban communications. *IEEE Trans Antennas Propag* 51(11):3109–3121
- Kilaru A, Kotamraju SK, Avlonitis N, Kavya KCS (2016) Rain rate intensity model for communication link design across the Indian region. *J Atmos Solar Terr Phys* 145:136–142
- Kim JC, Schall DE (1999) On the improvement of low elevation angle satellite communications impaired by tropospheric fading effects. In: *Military communications conference proceedings, 1999. MILCOM 1999*. IEEE, vol 1. IEEE, pp 603–607
- Kishore KV, Rajesh GS, Kumar V, Srinivasulu P, Kavya KCS, Kotamraju SK (2015) Design and simulation of 8-way unequal amplitude equal phase RF feeder network using conventional microstrip technology for 430 MHz tropospheric wind profiling radar. In: *2015 International conference on signal processing and communication engineering systems (SPACES)*. IEEE, pp 226–230
- Kubista E, Fontan FP, Castro MV, Buonomo S, Arbesser-Rastburg BR, Baptista JPVP (2000) Ka-band propagation measurements and

- statistics for land mobile satellite applications. *IEEE Trans Veh Technol* 49(3):973–983
- Kumar A, Hudiara IS, Jassal BS, Singh J (2004) Measurement of rain-induced zenith-path attenuation using 19.9 GHz radiometer at Amritsar (India). *IEEE Trans Antennas Propag* 52(3):702–708
- Lee JH, Kim YS, Kim JH, Choi YS, Pack JK (2001) Influence of rain attenuation models on the link availability of Ka-band non-GSO FSS system. In: *Antennas and propagation society international symposium, 2001. IEEE, vol 3. IEEE*, pp 104–108
- Lee JH, Choi YS, Lee HS, Pack JK (2003). Real-time estimation of rain attenuation on the satellite link. In: *Vehicular technology conference, 2003. VTC 2003-Spring. The 57th IEEE semiannual, vol 4. IEEE*, pp 2291–2294
- Lekkla R, Prapinmongkolkarn P (1998) Diurnal variations in rain attenuation on Ku band earth-space paths. *Int J Satell Commun* 16(5):219–236
- Lekkla R, McCormick KS, Rogers DV (1998) 12-GHz fade duration statistics on earth-space paths in South-East Asia. In: *Proceedings of URSI commission F open symposium on climatic parameters in radiowave propagation prediction (CLIMPARA 98), Ottawa, Ontario, Canada*, pp 167–170
- Lin DP, Chen HY (2002) An empirical formula for the prediction of rain attenuation in frequency range 0.6–100 GHz. *IEEE Trans Antennas Propag* 50(4):545–552
- Liou YA (2000) Radiometric observation of atmospheric influence on space-to-earth Ka-band propagation in Taiwan. In: *Proceedings of national science council, Republic of China. Part A, physical science and engineering, vol 24, no. 3*, pp 238–247
- Locatelli JD, Hobbs PV (1974) Fall speeds and masses of solid precipitation particles. *J Geophys Res* 79(15):2185–2197
- Madhuri AS, Immadi G, Narayana MV (2018) Estimation of cumulative distribution of scintillation effect on Ku band frequencies for one of the tropical regions using various models. *J Eng Sci Technol Rev* 11(1):151–155
- Maekawa Y, Fujiwara T, Shibagaki Y, Sato T, Yamamoto M, Hashiguchi H, Fukao S (2004) First year results on rain attenuation characteristics of satellite links at equatorial atmospheric radar. In: *Antennas and propagation society international symposium, 2004. IEEE, vol 2, IEEE*, pp 1660–1663
- Maitra A (2004) Rain attenuation modeling from measurements of rain drop size distribution in the Indian region. *IEEE Antennas Wirel Propag Lett* 3(1):180–181
- Maitra A, Chakravarty K (2005) Ku-band rain attenuation observations on an earth-space path in the Indian region. In: *28th URSI-GA*, pp 23–29
- Maitra A, Adhikari A, Bhattacharya A (2012) Some characteristics of earth-space path propagation phenomena at a tropical location. *92.60. jf; 92.60. Kc*
- Manabe T, Yoshida T (1995) Rain attenuation characteristics on radio links. In: *1995 URSI international symposium on signals, systems, and electronics, 1995. ISSSE'95, proceedings. IEEE*, pp 77–80
- Mandeep J, Islam M (2012) Evaluation of statistical tropospheric scintillation models using SUPERBIRD-C satellite for Malaysia. *Acta Geophys* 60(4):1180–1191
- Mandeep SJS, Hassan SIS, Tanaka K, Ain F, Igarashi F, Iida M (2007) Measurement of tropospheric scintillation from satellite beacon at Ku-band In South East Asia. *IJCSNS Int J Comput Sci Netw Secur* 7(2):251–256
- Mandeep JS, Hassan SIS, Tanaka K (2008) Rainfall effects on Ku-band satellite link design in rainy tropical climate. *J Geophys Res: Atmos* 113(D5)
- Manning RM (1984) A rain attenuation model for satellite link attenuation predictions incorporating the spatial inhomogeneity of rainrate. *Int J Satell Commun* 2(3):187–197
- Martellucci A (2002) Radiowave propagation modelling for SatCom services at Ku-band and above. COST action 255 final report
- Martellucci A, Luini L, Testoni A, Paraboni A, Riva C (2000) Assessment of the spatial correlation of water vapour, liquid water and rain amounts in Europe. North America and Japan using ECMWF, Re-analysis data, ESA report
- Matricciani E (1991) Rain attenuation predicted with a two-layer rain model. *Eur Trans Telecommun* 2(6):715–727
- Matricciani E (2008) Global formulation of the synthetic storm technique to calculate rain attenuation only from rain rate probability distributions. In: *Antennas and propagation society international symposium, 2008. AP-S 2008. IEEE. IEEE*, pp 1–4
- Matricciani E, Mauri M (1995) Italsat-Olympus 20-GHz orbital diversity experiment at Spino d'Adda. *IEEE Trans Antennas Propag* 43(1):105–108
- Matricciani E, Riva C, Castanet L (2006) Performance of the synthetic storm technique in a low elevation 5 slant path at 44.5 GHz in the French Pyrénées. In: *First European conference on antennas and propagation, 2006. EuCAP 2006. IEEE*, pp 1–6
- Mauri M, Paraboni A, Tarducci D (1986) Depolarization measurements at 11.6 GHz on earth-space paths using the SIRIO satellite. *IEEE Trans Antennas Propag* 34(4):582–585
- McCarthy DK, Allnutt JE, Salazar WE, Wanmi F (1994) Results of 11.6 GHz radiometric experiment in Cameroon: second year. *Electron Lett IEE* 30(17):1449–1450
- Mello L, Pontes MS (2012) Unified method for the prediction of rain attenuation in satellite and terrestrial links. *J Microw Optoelectron Electromagn Appl* 11(1):1–14
- Michelson DG, Liu W (2009) Simulation of rain fading and scintillation on Ka-band earth-LEO satellite links. In: *Canadian conference on electrical and computer engineering, 2009. CCECE'09, IEEE*, pp 635–640
- Misme P, Waldeufel P (1980) A model for attenuation by precipitation on a microwave earth-space link. *Radio Sci* 15(3):655–665
- Mitra SK, Vohl O, Ahr M, Pruppacher HR (1990) A wind tunnel and theoretical study of the melting behavior of atmospheric ice particles. IV: experiment and theory for snow flakes. *J Atmos Sci* 47(5):584–591
- Mohamad H, Hamaguchi K, Teh CH, Lee SW, Arif NAM, Wee CY (2006) Joint study of rain attenuation effect for fixed wireless access system. Project report NICT-MMU, Japan
- Mondal NC, Sarkar SK (2003) Rain height in relation to 0 °C isotherm height for satellite communication over the Indian Subcontinent. *Theoret Appl Climatol* 76(1–2):89–104
- Mondal NC, Timothy KI, Battacharya AB, Sakar SK (1997) Technical note The effect of rate of decay of rain path profile on microwave communication. *Int J Remote Sens* 18:3669–3675
- Mondal NC, Sarkar SK, Bhattacharya AB, Mali P (2001) Rain height in relation to 0 °C isotherm height over some Indian tropical locations and rain attenuation for an Indian south coastal station for microwave and millimeter wave communication systems. *Int J Infrared Millimeter Waves* 22(3):495–504
- Moupfouma F (1984) Improvement of a rain attenuation prediction method for terrestrial microwave links. *IEEE Trans Antennas Propag* 32(12):1368–1372
- Moupfouma F (1993) Point rainfall rate cumulative distribution function valid at various locations. *Electron Lett* 29(17):1503–1505
- Nackoney OG, Davidson D (1982) Results of 11.7-GHz CTS rain attenuation measurements at Waltham, Massachusetts. *Radio Sci* 17(06):1435–1442
- Naicker K, Mneney SH (2006) Propagation measurements and multipath channel modelling for line-of-sight links at 19.5 GHz. *Res J* 97:162–171
- Nelson B, Stutzman WL (1996) Fade slope on 10 to 30 GHz earth-space communication links—measurements and modelling. *IEE Proc Microw Antennas Propag* 143(4):353–357

- Oguchi T (1983) Electromagnetic wave propagation and scattering in rain and other hydrometeors. *Proc IEEE* 71(9):1029–1078
- Ojo JS, Omotosho TV (2013) Comparison of 1-min rain rate derived from TRMM satellite data and raingauge data for microwave applications in Nigeria. *J Atmos Solar Terr Phys* 102:17–25
- Ojo JS, Owolawi PA (2015) Application of synthetic storm technique for diurnal and seasonal variation of slant path Ka-band rain attenuation time series over a subtropical location in South Africa. *Int J Antennas Propag*
- Ojo JS, Ajewole MO, Sarkar SK (2008) Rain rate and rain attenuation prediction for satellite communication in Ku and Ka bands over Nigeria. *Prog Electromagn Res* 5:207–223
- Ojo JS, Ajewole MO, Emiliani LD (2009) One-minute rain-rate contour maps for microwave-communication-system planning in a tropical country: Nigeria. *IEEE Antennas Propag Mag* 51(5):82–89
- Ong JT, Timothy KI, Chong JH, Rao SVB (2003) Heavy rain effects on the propagation of free space optical links in Singapore
- Pan QW, Allnutt JE (2004) 12-GHz fade durations and intervals in the tropics. *IEEE Trans Antennas Propag* 52(3):693–701
- Pan QW, Allnutt JE, Haidara F (2000) Seasonal and diurnal rain effects on Ku-band satellite link designs in rainy tropical regions. *Electron Lett* 36(9):1
- Panagopoulos AD, Kanellopoulos JD (2003) On the rain attenuation dynamics: spatial-temporal analysis of rainfall rate and fade duration statistics. *Int J Satell Commun Netw* 21(6):595–611
- Panagopoulos AD, Arapoglou PDM, Cottis PG (2004) Satellite communications at Ku, Ka, and V bands: propagation impairments and mitigation techniques. *IEEE Commun Surv Tutorial* 6(3):2–14
- Panchal P, Joshi R (2016) Performance analysis and simulation of rain attenuation models at 12–40 GHz band for an earth space path over Indian cities. *Proc Comput Sci* 79:801–808
- Paraboni A, Capsoni C, Zaccarini F (1995) The horizontal structure of rain and its impact on the design of advanced satellite systems at centimetre and millimetre wavelengths. In: *Microwave and optoelectronics conference, 1995. Proceedings., 1995 SBMO/IEEE MTT-S international, vol 2*. IEEE, pp 519–525
- Park YH, Lee JH, Jambaljav N, Pack JK (2002) Empirical study on the rain drop-size model for rain attenuation calculations. In: *Proceedings of the URSI general assembly, vol 4*, pp 01213. <http://www.ursi.org/Proceedings/ProcGA05/pdf/F01P>
- Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen–Geiger climate classification. *Hydrol Earth Syst Sci* 11:1633–1644. <https://doi.org/10.5194/hess-11-1633-2007>
- Polaries Baptista JPV, Kubista E, Witternig N, Randeu WL (1989) Worst-month statistics for high outage probabilities. In: *Antennas and propagation, 1989. ICAP 89., sixth international conference on (Conf. publ. no. 301)*, IET, pp 10–13
- Polaries Bapista JPV, Kubista E, Witternig N, Randeu WL (1990) Worst-month statistics for high outage probabilities. ESA/ESTEC, INW/TU Graz, The Netherlands, Austria
- Polonio R, Riva C (1998) ITALSAT propagation experiment at 18.7, 39.6, and 49.5 GHz at Spino D’Adda: three years of CPA statistics. *IEEE Trans Antennas Propag* AP 46(5):631–635
- Pontes MS, Mello LS, Migliora CGS (1990) Ku-band slant-path radiometric measurements at three locations in Brazil. *Int J Satell Commun* 8(3):239–249
- Pontes MS, da Silva Mello L, de Souza RSL, Miranda ECB (2005) Review of rain attenuation studies in tropical and equatorial regions in Brazil. In: *2005 Fifth international conference on information, communications and signal processing*. IEEE, pp 1097–1101
- Rachan L, Prasit P (1998) Diurnal variations in rain attenuation on Ku band earth–space paths. *Int J Satel Commun* 16:219–236
- Rahim RA, Leong LC, Chan KS, Pang JF (2005) Data acquisition process in optical tomography: signal sample and hold circuit. In: *1st international conference on computers, communications, & signal processing with special track on biomedical engineering, 2005. CCSP 2005*. IEEE, pp 189–192
- Raina MK (1996) Atmospheric emission measurements of attenuation by microwave radiometer at 19.4 GHz. *IEEE Trans Antennas Propag* 44(2):188–191
- Raina M, Uppal G (1981) Rain attenuation measurements over New Delhi with a microwave radiometer at 11 GHz. *IEEE Trans Antennas Propag* 29(6):857–864
- Raina M, Uppal G (1984) Frequency dependence of rain attenuation measurements at microwave frequencies. *IEEE Trans Antennas Propag* 32(2):185–187
- Rakshit G, Adhikari A, Maitra A (2017) Modelling of rain decay parameter for attenuation estimation at a tropical location. *Adv Space Res* 59(12):2901–2908
- Ramachandran V, Kumar V (2004) Rain attenuation measurement on Ku-band satellite TV downlink in small island countries. *Electron Lett* 40(1):49–50
- Ramachandran V, Kumar V (2007) Modified rain attenuation model for tropical regions for Ku-Band signals. *Int J Satell Commun Netw* 25(1):53–67
- Rao TN, Rao DN, Mohan K, Raghavan S (2001) Classification of tropical precipitating systems and associated Z–R relationships. *J Geophys Res: Atmos* 106(D16):17699–17711
- Raynaud L, Chenierie I, Lemorton J (1999) Improved modelling of propagation and backscattering of millimetre waves in the melting layer
- Recommendation CCIR (1992) Method for the subjective assessment of the quality of television pictures. CCIR Broadcasting Service (Television), pp 166–189
- Reddy GV (2005) Atmospheric constraints in HF UHF satellite communication in the Indian Sub Continent
- Rice P, Holmberg N (1973) Cumulative time statistics of surface-point rainfall rates. *IEEE Trans Commun* 21(10):1131–1136
- Rodda MJ, Williamson AG (1997) Results of a two year radiometric measurement programme in New Zealand. *Electron Lett* 33(4):326–328
- Rogister A, Mertens D, Vanhoenacker-Janvier D, Martellucci A, Arbesser-Rastburg B (2003) RAPIDS: radio propagation integrated database system. In: *Meeting and joint workshop with COST272, ESTEC, The Netherlands*
- Russchenberg HWJ, Ligthart LP (1996) Backscattering by and propagation through the melting layer of precipitation: a new polarimetric model. *IEEE Trans Geosci Remote Sens* 34(1):3–14
- Salonen ET, Baptista JP (1997) A new global rainfall rate model. In: *Tenth international conference on antennas and propagation, (conf. publ. no. 436)*, vol 2. IET, pp 182–185
- Salonen E, Uppala S (1991) New prediction method of cloud attenuation. *Electron Lett* 27(12):1106–1108
- Sánchez-Lago I, Fontán FP, Mariño P, Fiebig UC (2007) Validation of the synthetic storm technique as part of a time-series generator for satellite links. *IEEE Antennas Wirel Propag Lett* 6:372–375
- Sarat Kumar K, Vijaya Bhaskara Rao S, Narayana Rao D (2008) Prediction of Ku band rain attenuation using experimental data and simulations for Hassan, India. *Int J Comput Sci Netw Secur IJC-SNS*, 8(4):P10–P15. ISSN: 1738:7906
- Sarkar SK, Prasad MVSAN, Dutta HN, Reddy BM, Rao DN (1989) Rain and extent of cells over the Indian subcontinent. In: *Sixth international conference on antennas and propagation, ICAP 89 (Conf Publ No 301)*, vol 2, pp 318–321
- Sarkar SK, Prasad MVSAN, Dutta HN, Reddy BM (1995) Rain attenuation for satellite paths over two tropical stations. In: *Ninth international conference on antennas and propagation, 1995., (conf. publ. no. 407)*, vol 2. IET, pp 94–98
- Semire FA, Mohd-Mokhtar R, Ismail W, Mohamad N, Mandeep JS (2015) Modeling of rain attenuation and site

- diversity predictions for tropical regions. *Annales Geophysicae* (09927689) 33(3):321–331
- Shrestha S, Choi DY (2017) Characterization of rain specific attenuation and frequency scaling method for satellite communication in South Korea. *Int J Antennas Propag* 2017
- Singliar R, Bitó J, Din J, Tharek AR (2006) Comparison of predicted attenuation of satellite rain attenuation distribution in Malaysia and Hungary. In: *Proceedings of the 16th international Czech–Slovakia scientific conference radioelektronika*, pp 246–249
- Stewart RE, Marwitz JD, Pace JC, Carbone RE (1984) Characteristics through the melting layer of stratiform clouds. *J Atmos Sci* 41(22):3227–3237
- Stutzman WL (1993) Prolog to the special section on propagation effects on satellite communication links. *Proc IEEE* 81:850–855
- Stutzman WL, Dishman WK (1982) A simple model for the estimation of rain-induced attenuation along earth–space paths at millimeter wavelengths. *Radio Sci* 17(6):1465–1476
- Sum CS, Din J, Tharek AR, Abidi MZ (2003) Studies on characteristics of rain fade at 23 GHz for terrestrial links. In: *Asia-Pacific conference on applied electromagnetics (APACE 2003)*, 7803-8129-7/03/2003. IEEE, Shah Alam, Malaysia, pp 76–79
- Suryana J, Utoro S, Tanaka K, Igarashi K, Jida M (2005a) Study of prediction models compared with the measurement results of rainfall rate and Ku-band rain attenuation at Indonesian tropical cities. In: *2005 Fifth international conference on information, communications and signal processing*. IEEE, pp 1580–1584
- Suryana J, Utoro S, Tanaka K, Igarashi K, Jida M (2005b) Two years characterization of concurrent Ku-band rain attenuation and tropospheric scintillation in Bandung, Indonesia using JCSAT3. In: *2005 Fifth international conference on information, communications and signal processing*. IEEE, pp 1585–1589
- Svjatogor L (1985) Telecommunication working group. Dresden, GDR, Report
- Thurai M, Deguchi E, Okamoto K, Salonen E (2005) Rain height variability in the tropics. *IEE Proc Microw Antennas Propag* 152(1):17–23
- Timothy KI, Ong JT, Choo EBL (2000) Descriptive fade slope statistics on INTELSAT Ku-band communication link. *Electron Lett* 36(16):1422–1424
- Timothy KI, Ong JT, Choo EB (2002) Raindrop size distribution using method of moments for terrestrial and satellite communication applications in Singapore. *IEEE Trans Antennas Propag* 50(10):1420–1424
- van de Kamp MMJL (2002a) Rain attenuation as a Markov process: how to make an event. In: *2nd international workshop of COST action*, vol 280, pp 26–28
- van de Kamp MMJL, Castanet L (2002b) Propagation modeling and fade mitigation for Ka-band satellite system. In: *1st international workshop COST-280*, PM3018
- Ventouras S, Wrench CL (1999) Diurnal variations of 20 GHz and 40 GHz slant path attenuation statistics in southern England
- Verma AK, Jha KK (1996) Rain drop size distribution model for Indian climate. *J Radio Space Phys* 25:15–21
- Vogel WJ (1982) Measurements of satellite beacon attenuation at 11.7, 19.04, and 28.56 GHz and radiometric site diversity at 13.6 GHz. *Radio Sci* 17(6):1511–1520
- Walther Å, Terje T (2003) A novel method for predicting site dependent specific rain attenuation of millimeter radio waves. *IEEE Trans Antennas Propag* 51:2987–3000
- Watson PA, Hu YF (1994) Prediction of attenuation on satellite-earth links for systems operating with low fade margins. *IEE Proc Microw Antennas Propag* 141(6):467–473
- Watson PA, Leitao MJ, Turney O, Sengupta N (1985) Development of a climatic map of rainfall attenuation for Europe. Postgraduate School of Electrical & Electronic Engineering, University of Bradford, Report No. 372 (Final report for ESA/ESTEC contract No. 5192/82/NL/GM)
- Wilson PS, Toumi R (2005) A fundamental probability distribution for heavy rainfall. *Geophys Res Lett* 32(14):L14812
- Yamada M, Miura Y (1998) Rain attenuation characteristics at 12 GHz on an earth–space path. In: *1998 International conference on microwave and millimeter wave technology proceedings, 1998*. ICMMT'98. IEEE, pp 1016–1019
- Yang H, He C, Zhu H, Song W (2000) Prediction of slant path rain attenuation based on artificial neural network. In: *The 2000 IEEE international symposium on circuits and systems, 2000*. Proceedings. ISCAS 2000 Geneva, vol 1. IEEE, pp 152–155
- Yang H, He C, Zhu H, Song W (2001) Earth–space rain attenuation model based on EPNet-evolved artificial neural network. *IEICE Trans Commun* 84(9):2540–2549
- Yee TS, Kooi PS, Leong MS, Li LW (2001) Tropical raindrop size distribution for the prediction of rain attenuation of microwaves in the 10–40 GHz band. *IEEE Trans Antennas Propag* 49(1):80–83
- Yussuff AI, Khamis NH (2012) Rain attenuation modelling and mitigation in the tropics: brief review. *Int J Electr Comput Eng* 2(6):748–757
- Yussuff AI, Khamis NHH (2013) Modified ITU-R rain attenuation prediction model for a tropical station. *J Ind Intell Inf* 1(3)
- Zhang W, Karhu SI, Salonen ET (1994) Predictions of radiowave attenuations due to a melting layer of precipitation. *IEEE Trans Antennas Propag* 42(4):492–501
- Zhenwei Z, Leke L, Yumei L (2003) A prediction model of rain attenuation along earth–space path. In: *2003 6th international symposium on antennas, propagation and EM theory, 2003*. Proceedings. IEEE, pp 516–519
- Zhenwei Z, Leke L, Yumei L (2004) Prediction models for rain effects on earth–space links. In: *Radio science conference, 2004*. Proceedings. 2004 Asia–Pacific. IEEE, pp K22–K25



Studying permafrost by integrating satellite and in situ data in the northern high-latitude regions

Nureldin A. A. Gido^{1,2} · Mohammad Bagherbandi^{1,2} · Lars E. Sjöberg^{1,2} · Robert Tenzer³

Received: 28 October 2017 / Accepted: 27 February 2019 / Published online: 7 March 2019
© The Author(s) 2019

Abstract

There is an exceptional opportunity of achieving simultaneous and complementary data from a multitude of geoscience and environmental near-earth orbiting artificial satellites to study phenomena related to the climate change. These satellite missions provide the information about the various phenomena, such as sea level change, ice melting, soil moisture variation, temperature changes and earth surface deformations. In this study, we focus on permafrost thawing and its associated gravity change (in terms of the groundwater storage), and organic material changes using the gravity recovery and climate experiment (GRACE) data and other satellite- and ground-based observations. The estimation of permafrost changes requires combining information from various sources, particularly using the gravity field change, surface temperature change, and glacial isostatic adjustment. The most significant factor for a careful monitoring of the permafrost thawing is the fact that this process could be responsible for releasing an additional enormous amount of greenhouse gases emitted to the atmosphere, most importantly to mention carbon dioxide (CO₂) and methane that are currently stored in the frozen ground. The results of a preliminary numerical analysis reveal a possible existence of a high correlation between the secular trends of greenhouse gases (CO₂), temperature and equivalent water thickness (in permafrost active layer) in the selected regions. Furthermore, according to our estimates based on processing the GRACE data, the groundwater storage attributed due to permafrost thawing increased at the annual rates of 3.4, 3.8, 4.4 and 4.0 cm, respectively, in Siberia, North Alaska and Canada (Yukon and Hudson Bay). Despite a rather preliminary character of our results, these findings indicate that the methodology developed and applied in this study should be further improved by incorporating the in situ permafrost measurements.

Keywords Climate change · Permafrost · Gravity · Grace · Greenhouse gas

Introduction

The permafrost is frozen soil, sediment or rock that remains under 0 °C for at least two consecutive years (Brown et al. 1998). There are different layers, of which frozen ground represents just one portion. The active layer is the surface layer of soil that is seasonally frozen in winter and thaws each summer, causing relatively large surface movements at the order of centimetres, or even decimetres over periods of

months (seasonal variations) and years (cf. Biskaborn et al. 2015). The talik is an unfrozen ground layer that lies below the active layer and above the permafrost. As seen in Fig. 1, the permafrost is found in large parts of the Northern Hemisphere, mostly in Siberia, Canada, Alaska, Greenland, and the northern parts of Sweden and Norway (Schaefer et al. 2012). Climate warming is more rapid in the northern high-latitude regions and causes thawing of the permafrost over a great expanse of area in the Arctic region (Liu et al. 2010). This has major impact on the arctic water cycle, as ground ice melts and flow pathways between layers and groundwater increase. The permafrost thaw is a very complex phenomena that can in some cases lead to the water storage change that could be detected in gravity variations. Soils in the Arctic permafrost region are estimated to hold about twice as much carbon as the atmosphere (Hugelius et al. 2011). As previously frozen permafrost organic carbon thaws and decomposes, there is a potential positive feedback between

✉ Nureldin A. A. Gido
Nureldin.Gido@hig.se; naag@kth.se

¹ Faculty of Engineering and Sustainable Development, University of Gävle, SE-80176 Gävle, Sweden

² Division of Geodesy and Satellite Positioning, Royal Institute of Technology (KTH), SE-10044 Stockholm, Sweden

³ Department of Land Surveying and Geo-Informatics, Hong Kong Polytechnic University, Kowloon, Hong Kong

Fig. 1 Permafrost distribution in the northern high latitudes. Image credit: Philippe Rekacewicz, 2005, UNEP/GRID-Arendal maps and graphics library based on International Permafrost Association (1998) circumpolar active-layer permafrost system (CAPS), version 1.0. https://www.wunderground.com/resources/climate/melting_permafrost.asp



the permafrost thaw and the global warming if the Arctic turns from a net sink to a net source of carbon. Hence, studying thawing permafrost is very important, not only from a perspective of localized geo-hazard such as erosion, damage to buildings and infrastructure, but also with respect to its possible global impact due to greenhouse gas emissions.

The permafrost regions are one of the most heavily impacted regions by the climate change. Studying the permafrost might allow us to investigate a long-term development of the earth's shape interior, gravity field, climate change and assessment of its impact. The permafrost thawing has been investigated using satellite gravimetry techniques (GRACE; Tapley et al. 2004) in many publications in addition to streamflow analyses (e.g. Sjöberg et al. 2013) that use thermal sensors in boreholes that measure the thickness of the active permafrost layer. For example, Velicogna et al. (2012) studied the permafrost thawing using GRACE data in the Lena river basin (cf. Muskett and Romanovsky 2009, 2011; Vey et al. 2013) and Shabanloui and Müller (2015) conducted a similar study for the whole Siberia (cf. Chao et al. 2011). They determined the water storage

changes attributed to the permafrost thawing using GRACE data and compared their results with various hydrological effects such as the soil moisture and the river run-off. The importance of the permafrost for the earth's ecosystem (in terms of the released carbon dioxide and nitrogen into the atmosphere due to thawing of permafrost active layer) was also recognized by Treat et al. (2014), Yang et al. (2013), Schaefer et al. (2011) and others.

In this study, we propose a unique approach to study the permafrost thawing using satellite data, e.g. GRACE and other satellite- and ground-based observations in order to determine surface changes/deformations in the northern high-latitude region. For this purpose, the mass variations due to the permafrost thawing are estimated in terms of the groundwater level change in the study region. By using the outcomes of this study, we try to find geophysical mechanisms being the most plausible, e.g. the thawing of permafrost and the glacial isostatic adjustment (GIA). The satellite-based data, e.g. GRACE, is collected from low-orbited dedicated missions, which provide homogeneous and global coverage. However, these data suffer from inherent data processing problems and

need to be addressed properly in order to achieve reliable and accurate results. Another task to be tackled is finding an optimal method for data noise filtering (especially for GRACE data (Sjöberg and Bagherbandi 2017, Sect. 8.7) and meteorological signals (rain, snow water, surface water, etc.). How to extract the permafrost signal from satellite data is an important challenge here. For example, studying temporal gravity changes and their relation with the permafrost using GRACE data requires the knowledge about the land uplift (GIA) (Peltier 2004; Ågren and Svensson 2007) and the continental hydrological signal (GLDAS, Global Land Data Assimilation System; Rodell et al. 2004). The potential of the GRACE for detecting mass changes related to different phenomena has been studied before and during the first years of its operation, for instance, by Wahr et al. (1998, 2004) and Swenson et al. (2003). It is important to mention here that the gravity data collected by GRACE require applying a smoothing procedure in order to reduce the effects of errors present in the short-wavelength component. Various methods have been proposed to filter the GRACE data (cf. Swenson and Wahr 2006; Wouters and Schrama 2007; Klees et al. 2008; and Kusche et al. 2009). Among them, we could mention the most commonly applied isotropic Gaussian (Wahr et al. 1998) or the non-isotropic (Han et al. 2005) filters. However, none of these methods accounts for a correlated noise in data.

In this study, we estimate water storage changes due to permafrost thawing (not dry permafrost, bare rocks, thawing) using satellite gravimetry methods in the study region. Therefore, studying the contribution of some disturbing signals such as land uplift (GIA), surface water and surface snowfall on the signal detected by GRACE is necessary. Moreover, we also address the determination of the permafrost-related near-surface mass changes (due to permafrost thawing) and their correlation with released greenhouse gases.

Monitoring near-surface mass change by GRACE

Today, the most successful data for studying temporal changes in the gravity field stem from the very long record of satellite laser ranging for very long wavelengths (Moore et al. 2005), and more recently also from the accurate gravity-dedicated satellite mission GRACE for shorter/more regional phenomena (cf. Lemoine et al. 2007). GRACE is appropriate for studying some geodynamic phenomena such as GIA, climate change and permafrost thawing, because these phenomena are related to a mass change and consequently propagated into gravity field change. Moreover, these changes occur over relatively large areas, thus having prevailing long-wavelength spatial pattern that could be detected by GRACE that could measure secular gravity changes related to near-surface mass changes.

A satellite disturbing potential model (T) typically consists of numerical values for the potential coefficients (C_{nm}). A time-dependent change in the earth’s gravity field due to mass transport causes changes in the harmonic coefficients. These changes can be described by the residual spherical harmonic coefficients ΔC_{nm} that are obtained after subtracting the mean value from monthly solutions. The absolute disturbing potential and its changes are then described by

$$\begin{pmatrix} T \\ \Delta T \end{pmatrix} = \frac{GM}{R} \sum_{n=0}^{n_{\max}} \left(\frac{R}{r}\right)^{n+1} \sum_{m=-n}^n \begin{pmatrix} C_{nm} \\ \Delta C_{nm} \end{pmatrix} Y_{nm}(\theta, \lambda), \quad (1a)$$

$$\Delta C_{nm} = C_{nm} - \bar{C}_{nm}, \quad (1b)$$

where \bar{C}_{nm} is the mean value of spherical harmonics over time, ΔT is either the change in T from one time to another or the difference between T at one time and the time average of T , GM is the geocentric gravitational constant (i.e. the product of Newton’s gravitational constant and the total mass of the earth including the atmosphere), R the Earth’s mean radius, n_{\max} is maximum degree of harmonic expansion, (r, θ, λ) are, respectively, the geocentric radius, colatitude and longitude of the computational point, Y_{nm} is the surface spherical harmonic of degree n and order m , and

$$\Delta C_{nm} = \{ \Delta C_{nm} \}_{\text{surface mass}} + \{ \Delta C_{nm} \}_{\text{solid mass}}. \quad (2)$$

As seen in Eq. (2), the residual spherical harmonic coefficients ΔC_{nm} consist of the direct gravitational contribution of the surface mass $\{ \Delta C_{nm} \}_{\text{surface mass}}$ and the additional contribution of the surface mass load $\{ \Delta C_{nm} \}_{\text{solid mass}}$ that deforms the underlying solid earth. Hence, we have (cf. Wahr et al. 1998)

$$\begin{bmatrix} \{ \Delta C_{nm} \}_{\text{surface mass}} \\ \{ \Delta C_{nm} \}_{\text{solid mass}} \end{bmatrix} = \begin{bmatrix} 3/[4\pi R\bar{\rho}(2n+1)] \\ 3k_n/[4\pi R\bar{\rho}(2n+1)] \end{bmatrix} \iint_{\sigma} \Delta\sigma(\theta, \lambda) Y_{nm}(\theta, \lambda) d\sigma, \quad (3)$$

where $\Delta\sigma(\theta, \lambda)$ is the surface density change (mass/area), and $\bar{\rho}$ is the earth’s mean density. The load Love numbers k_n are modelled based on some available earth model (e.g. Farrell 1972; Sun and Sjöberg 1999; Bevis et al. 2016).

By repeating GRACE satellite tracks for measuring the gravity field, the linear least-squares regression analysis can be applied to determine temporal changes (i.e. secular trend) of the disturbing potential (cf. Sjöberg and Bagherbandi 2017, Chapt. 2). The regression equations are given by:

$$a + bt_i + \sum_{j=1}^J (c_j \cos \omega_j t_i + d_j \sin \omega_j t_i) = \Delta T_i - \varepsilon_i; \quad i = 1, \dots, n, \quad (4)$$

where $\omega_j = 2\pi/T_j$, periods are denoted by T_j and J is the number of periodic signals included in the adjustment. Considering a linear regression (i.e. $a + bt_i = \Delta T_i - \varepsilon_i$) in time (t_i) with observation equations (Eq. 4), the secular change of the spherical harmonic coefficients (C_{nm}) is given by:

$$C_{nm}^{\cdot} = \frac{1}{4\pi} \iint_{\sigma} \mathbf{b} Y_{nm}(\theta, \lambda) d\sigma. \quad (5)$$

where \mathbf{b} is a matrix which includes the computed secular trends using Eq. (4).

Sjöberg and Bagherbandi (2017, Sect. 8.7) showed that estimated rates of a secular trend in the gravity field do not differ significantly when either including or not periodic terms in the analyses. Using C_{nm}^{\cdot} , the secular changes of the disturbing potential and gravity disturbance can be described up to some limited degree n_{\max} as follows (Sjöberg and Bagherbandi 2017, Chapters 3 and 8)

$$\dot{T} = \frac{GM}{R} \sum_{n=0}^{n_{\max}} \left(\frac{R}{r}\right)^{n+1} \sum_{m=-n}^n C_{nm}^{\cdot} Y_{nm}(\theta, \lambda), \quad (6)$$

$$\dot{\delta g} = \frac{GM}{R^2} \sum_{n=0}^{n_{\max}} \left(\frac{R}{r_P}\right)^{n+2} (n+1) \sum_{m=-n}^n C_{nm}^{\cdot} Y_{nm}(\theta, \lambda), \quad (7)$$

According to Sjöberg and Bagherbandi (2017), Chapter 8, the near-earth's surface mass change rate can be determined from

$$\dot{\kappa} \approx \frac{\bar{\rho}R}{3} \sum_{n=0}^{n_{\max}} \frac{2n+1}{1+k_n} \sum_{m=-n}^n C_{nm}^{\cdot} Y_{nm}(\theta, \lambda). \quad (8)$$

The monthly mass changes propagate into changes in the earth's gravity field. For computational purposes, these mass changes are often considered as being concentrated in a very thin layer of water thickness changes. Their vertical extent is measured in centimetres of the equivalent water thickness. To get a general picture of the lateral distribution of mass anomalies, one may also represent the secular trend of the disturbing potential in terms of the total water storage (TWS) changes. Using GRACE data, we estimate the secular trend of TWS change according to Wahr et al. (1998) by

$$\dot{\text{TWS}} = \frac{R\bar{\rho}}{3\rho_w} \sum_{n=0}^{n_{\max}} \frac{2n+1}{1+k_n} \sum_{m=-n}^n C_{nm}^{\cdot} Y_{nm}(\theta, \lambda), \quad (9)$$

where ρ_w is the water density.

GRACE data and processing procedure for permafrost studies

There are different products for GRACE, monthly, weekly and 10-day solutions provided by several analysis centres. The three main analysis centres are: Center for Space Research (CSR) at the University of Texas at Austin, Jet Propulsion Laboratory (JPL) in Pasadena and GeoForschungsZentrum (GFZ) in Potsdam. In addition, there are solutions from the University of Bonn (ITG), the Centre National d'Etudes Spatiales (CNES) in Toulouse and the Technical University Delft (DEOS Mass Transport model). For this study, we used the Release-05 GRACE level-2 monthly time-variable gravity field products from CSR over the period from August 2002 to May 2016 provided in terms of the fully normalized geopotential spherical harmonic coefficients to determine the TWS changes. These GRACE data were filtered and corrected for GIA, hydrological signal and glaciers melting in order to estimate the equivalent water height secular trend due to the permafrost thawing layer (EWHP). The filtering and correction procedures are briefly summarized in the next sections.

Noise filtering

The epoch-wise GRACE gravity field models require smoothing in order to reduce the effects of errors at short wavelengths. The GRACE data contain coloured noise that shows up in the provided spherical harmonic coefficients (Kusche et al. 2009). Basically, truncating the spherical harmonic series at long wavelengths, where the noise is not yet significant, yields the loss of an unacceptably high portion of the signal. The noise can usually be described by systematic/correlated stripping patterns. The reason for this is the mission geometry, as GRACE twin-satellites fly in the same near-polar orbital plane, and the inter-satellite ranging observable used in gravity modelling transforms into a distinct along-track sensitivity (Sjöberg and Bagherbandi 2017, Chap. 8). Deficiencies in de-aliasing models cause an anisotropic error that cannot be removed, for instance, by applying the Gaussian isotropic filter. Instead, the de-correlation method should be used in the post-processing of GRACE data, such as those discussed by Kusche (2007), Klees et al. (2008) and Kusche et al. (2009). Kusche's technique uses a priori synthetic model of the observation geometry for the de-correlation. In this study, we used Kusche's technique (DDK2 filter) to remove stripping patterns in the CSR GRACE data (see Joud et al. 2017).

GIA correction

In the study region, the earth's crust is rising continuously since the last glacial maximum due to de-loading of the

former ice sheet. The process of ongoing relaxation due to the viscoelastic structure of the earth mantle is known as glacial isostatic adjustment (GIA). The land uplift is well documented from various studies (see e.g. Ekman 1991 and Peltier and Argus 2015 for historical reviews). Although GIA is a well-known process for more than a century, it still attracts the attention of geoscientific community due to its extended number of applications in the understanding, modelling, measuring and correcting for different phenomena related to the present issues of interest to the global society.

GIA causes changes in the earth's gravity field. The estimation of mass changes, obtained from GRACE data, in terms of the equivalent water thickness will not be correct without removing the GIA contribution. For this purpose, the land uplift rate should be converted to equivalent land uplift water height (ELUWH) by

$$\text{ELUWH} = \frac{\bar{\rho}}{3\rho_w} \sum_{n=0}^{n_{\max}} \frac{2n+1}{1+k_n} \sum_{m=-n}^n h_{nm} Y_{nm}(\theta, \lambda), \quad (10)$$

where h is the land uplift rate.

Hydrological corrections

The hydrological monthly data from the global land data assimilation system (GLDAS) hydrological model over the period from January 2002 to May 2016 with 0.25×0.25 or 1.0×1.0 arc-degree spatial resolution can be extracted (Chen et al. 1996; Koren et al. 1999) via the Hydrology Data and Information Service Center (HDISC). GLDAS provides global information on land surface status (e.g. surface temperature, soil moisture, precipitation, river run-off, etc.) with three-hour or monthly temporal resolutions based on integrating satellite and ground-based observation data. Four land surface models are currently used by GLDAS, specifically Mosaic, Noah, CLM and VIC (see Rodell et al. 2004). In this study, we used the Noah model according to Chao et al. (2011), which has a total of four soil layers thickness (i.e. 0–10, 10–40, 40–100 and 100–200 cm). The precipitation and other hydrological parameters are very important because their effect on the water storage change is significant in the study area (see Fig. 5). The regression analysis has been used to calculate the secular trend of the hydrological data (similar to Eq. 4).

Ice-melting correction

For extracting the water storage change due to permafrost thawing, the effect of ice-mass loss of Greenland ice sheet triggered by regional warming should be removed from GRACE data. The rate of ice elevation change from 2003 to 2012 was computed by using altimeter surveys from NASA's ATM flights during 2003–2012 supplemented with high-resolution

ice, cloud and land elevation satellite (ICESat) data from 2003 to 2009 (cf. Khan et al. 2014). The result shows that large areas in the centre of the Greenland ice sheet (GrIS) have few or no measurements during the period between 2010 and 2012. Thus, the variability in the elevation change rate during this period is not fully covered in the elevation change estimates. To overcome this problem, the ENVISAT radar altimeter measurements have been used from April 2009 to March 2012. This dataset is obtained from the RA-2 SGDR product (see Khan et al. 2014), while the standard (SGDR) corrections for atmospheric propagation and tides were applied. Observables provided by a radar altimetry suffer from problems with the slope effect; therefore, in this study, we did not use data around the edge of the GrIS (from the ice margin and about 50–70 km inland). While other studies have made attempts to merge laser altimetry data with radar altimetry data, here we used the elevation change rates estimated from ENVISAT and laser altimetry separately and combined these rates. Ice-melting results using the above-mentioned data are presented in Fig. 2. Similar to Eq. (10), the correction due to equivalent ice-melting water height was computed.

The near-earth surface mass variations obtained from GRACE data are related to hydrological (water) variations. These variations are, however, attributed not only to the permafrost thawing, but also to the precipitation and run-off variations. Therefore, the secular trend of equivalent water height due to permafrost thawing (EWHP) was estimated by removing the signals triggered by the GIA and hydrological contributions from the obtained TWS (Eq. 9). This procedure is described by the following scheme:

$$\text{EWHP} = \text{TWS}_{\text{GRACE}} - \text{EWHP} - \text{WS}_{\text{snow}} - \text{WS}_{\text{canopy}} - \text{WS}_{\text{run-off}} - \text{WS}_{\text{SoilMoisture}} - \text{WS}_{\text{ice-melting}}, \quad (11)$$

where ELUWH is the land uplift water equivalent height (Eq. 10), WS_{snow} is the water storage due to the water equivalent of snow, $\text{WS}_{\text{canopy}}$ is the canopy water storage, $\text{WS}_{\text{run-off}}$ is the water storage variation due to run-off river, $\text{WS}_{\text{SoilMoisture}}$ denotes the water storage due to soil moisture variation (accumulated soil moisture internally from different layers) and $\text{WS}_{\text{ice-melting}}$ is water storage from the ice melting in Greenland (for this study).

Results

Using the static 30-day GRACE gravity field solutions over the period from 2002 to 2016, the residual spherical harmonic coefficients were determined according to Eq. (1b), and consequently the disturbing potential changes were computed according to Eq. (1a). The spectral rates of the gravity

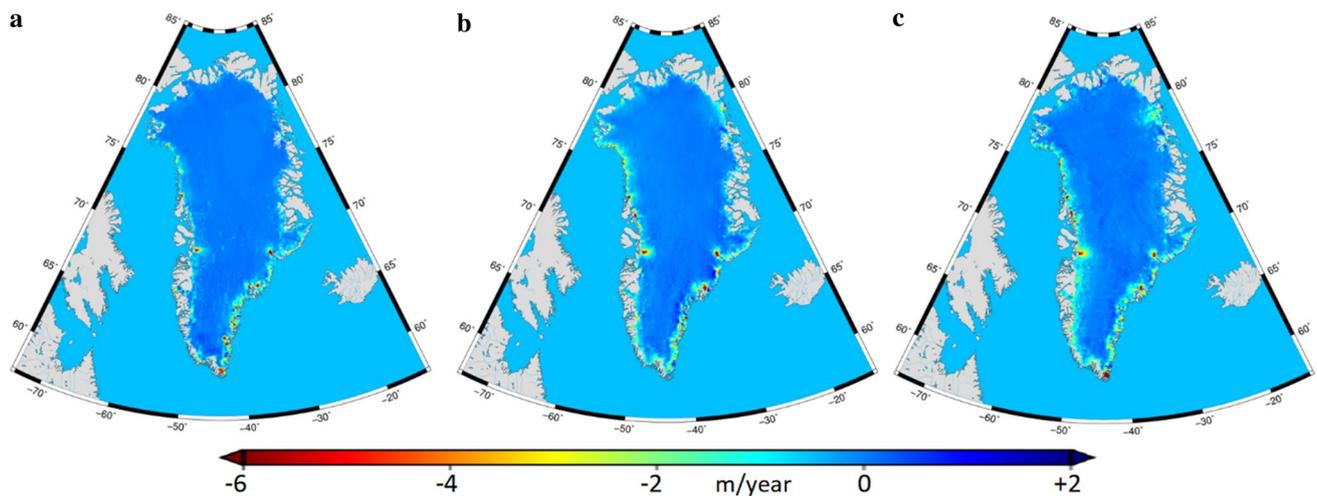


Fig. 2 Observed elevation change rates (m/year) in Greenland obtained from ICESat mission during periods: **a** from April 2003 to April 2006, **b** from April 2006 to April 2009 and **c** from April 2009 to April 2012

field change (in terms of C_{nm}) were then determined using a time-series analysis according to Eq. (5). After obtaining the values of \dot{C}_{nm} , the total water storage (TWS) was computed from Eq. (9). The resulting secular rates were then filtered and corrected for the contributions of GIA, hydrology (such as precipitation, soil moisture, river run-off and canopy using GLDAS data), and ice melting (in Greenland) in order to be used for calculating the equivalent water height trend due to permafrost thawing (E_{WHP}).

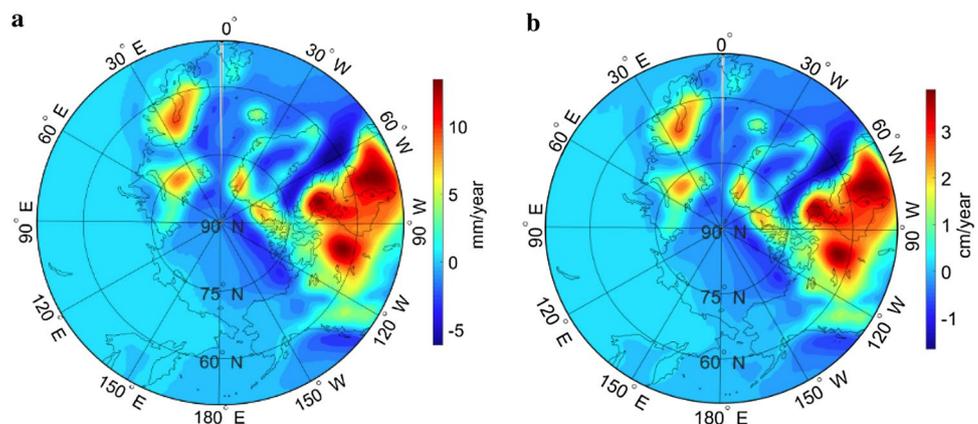
The obtained E_{WHP} from the GRACE data is verified using in situ data, e.g. land surface temperature from GLDAS (Rodell et al. 2004 and Swenson et al. 2003), borehole data and carbon dioxide changes (e.g. using NASA's EOS Aqua mission, see https://mirador.gsfc.nasa.gov/collections/AIRX3C2M_005.shtml). We expected to obtain a correlation between the rate of groundwater change (obtained from GRACE data) and in situ data. The distribution of permafrost monitoring stations (boreholes) in the study region is shown in Fig. 6b. More information about

these stations can be found through the global terrestrial network for permafrost through <http://gtnpdatabase.org/>, where the following important data are available: surface temperature, ground temperature, active layer thickness, surface soil moisture and air temperature.

To remove the effect of GIA from the secular rate of total water storage (TWS) obtained from GRACE, we used the ICE-6G (VM5a) model (Peltier and Argus 2015). This model shows the rate of radial displacements (Fig. 3a). Red and yellow areas in the figure indicate rising due to removal of the ice sheets. In ICE-6G (VM5a), the most recently available GPS observations have been employed. For this purpose, the land uplift rate was converted to equivalent land uplift water height (ELUWH) using Eq. (10). The result of the ELUWH is presented in Fig. 3b.

The resulting secular rates of TWS from the CSR GRACE data centre are shown in Fig. 4a. We further removed ELUWH (Fig. 4b) from TWS. Figure 4b shows the resulting values of TWS without the presence of GIA,

Fig. 3 Regional maps of: **a** the land uplift rate obtained from the ice model ICE-6G (VM5a) and **b** the equivalent land uplift water height change (ELUWH)



and especially it can be seen that the land uplift signal is removed in Laurentia and Fennoscandia.

Since GRACE detects all hydrological changes as well as the water level change due to the permafrost, we further modelled and removed the hydrological signals (in terms of the equivalent water height) from the secular rates of the TWS. In order to filter out the non-permafrost signals, we used GLDAS hydrological model. In Fig. 5, we illustrate the hydrological parameters adopted in our computation, particularly soil moisture, precipitation (snow), canopy and river run-off that were obtained from the GLDAS model using Eq. (4). It is worth mentioning that most of the hydrological signals, such as precipitation, are hidden implicitly in the soil moisture variations, i.e. one part of the precipitation is converted to the river run-off water masses, while the other part is evaporated. Therefore, we were able to filter out the effect of precipitation by subtracting the soil moisture rate from the TWS rate. The result after subtracting the hydrological signals is shown in Fig. 6a.

Table 1 shows the summary of different corrections to the total water storage computed from GRACE data, i.e. GIA, hydrological signal and glaciers melting. The results show that the gravitational effect of some corrections (such as river run-off and canopy) is in general at least one order of the magnitude smaller than the EWHP.

As emphasized by Lawrence et al. (2015), the large-scale permafrost thaw predicted by GRACE results in significant soil drying due to an increased drainage following permafrost thaw, even though permafrost domain water inputs are projected to rise. This phenomenon could clearly be recognized in our result shown in Fig. 6a. The results show positive trend for the water level after the permafrost layer thawing in selected stations (about 3–4 cm/year; see Table 2 for more details). This was confirmed in previous studies by Brown et al. (2000), Frauenfeld et al. (2004), Zhang et al. (2005) and Lemke et al. (2007). They reported a notable increase in Siberian permafrost active layer since 1950s. Muskett and Romanovsky (2009) demonstrated an increase in the groundwater storage in Lena, Ob' and Yenisei basins

in Siberia. Wu and Zhang (2010) reported a similar trend at the Tibetan Plateau. An increasing water level can also be caused by the topography/slope of surrounding areas of the stations, as more water masses are transported from higher to lower elevations after the permafrost layer thawing. For this reason, a much more relevant is the study of the lowland permafrost in unconsolidated sediments that are not ice-rich and aquifer can infiltrate additional water after permafrost thaw. It usually stores considerable amounts of water in form of ground ice. However, the permafrost thaw processes are quite complex even in lowland regions. Depressions form after the melt of ground ice/water and thermokarst lakes form (Phillips et al. 2003). These lakes usually undergo a cyclic growth and drainage. Ice wedges and related ice wedge polygons are also very important source of ground ice/water in these environments. However, in some locations the type of the permafrost contains significant amounts of ground ice of very different origins (French and Hugh 2013, Chapter 7). It can happen that the thawed soil/sediment is already saturated with water, and then an increase in water storage might not be possible in some areas, leading to a mass decrease, as is the case in the southern parts of Alaska and Greenland (e.g. in the areas with negative equivalent water height trend in Fig. 6a).

Comparison with greenhouse gases and temperature changes

The permafrost regions cover almost 80% of the study area, while locally reaching a maximum thickness up to about 1 km (Yakutia in central Siberia). Therefore, it is expected that the permafrost thawing plays a significant role in the climate change and mass transportation (cf. Steffen et al. 2012; Shabanloui and Müller 2015). Greenhouse gases emissions due to the permafrost thawing further change the ecosystem (Treat et al. 2014; and Yang et al. 2013). A large amount of carbon was stored in the permafrost by a process that took thousands of years. Mixing of soil layers due to repeated

Fig. 4 Regional maps of the equivalent water height secular trends (TWS) obtained from the CSR GRACE data over the period from August 2002 to May 2016: **a** before and **b** after removing the equivalent land uplift water height change

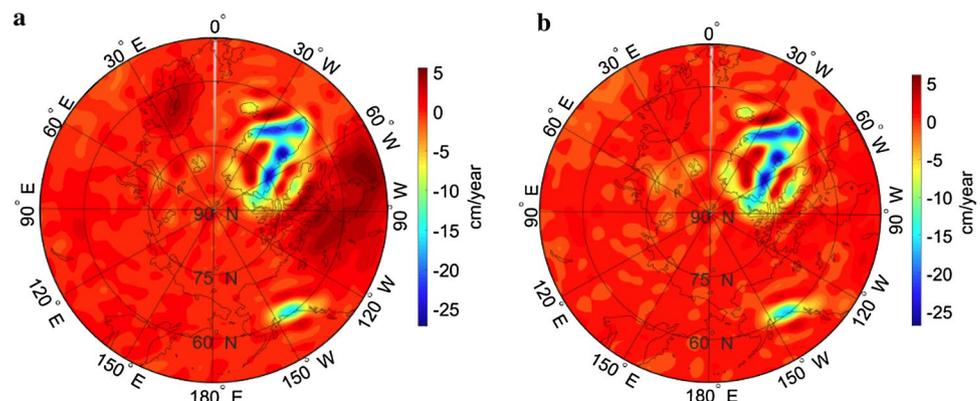
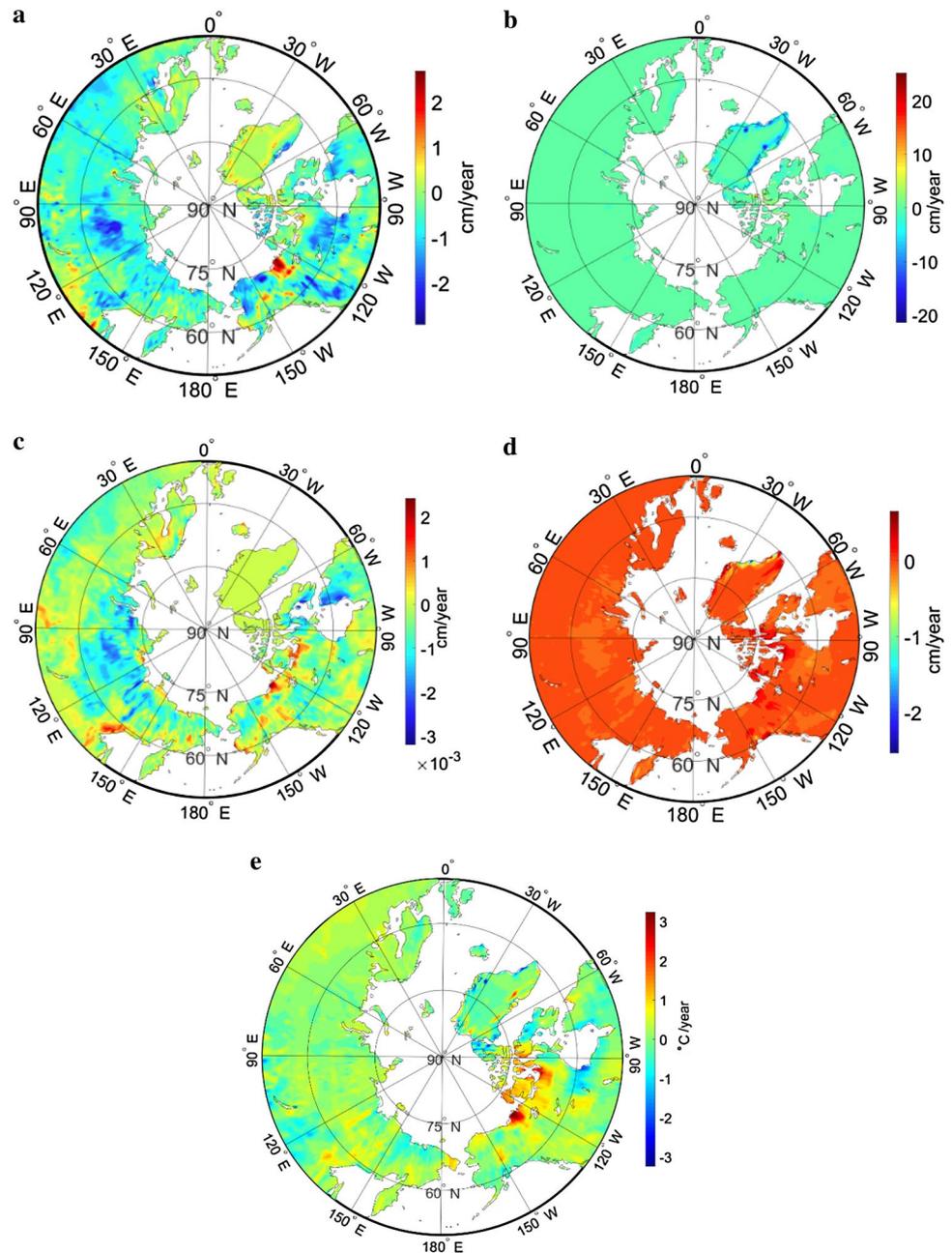


Fig. 5 Regional maps of the equivalent water height secular trends according to the GLDAS/NOAH land hydrological model due to hydrological contributions in terms of the water equivalent of **a** soil moisture ($WS_{\text{SoilMoisture}}^{\cdot}$), **b** snow (WS_{snow}^{\cdot}) and ice melting ($WS_{\text{ice-melting}}^{\cdot}$), **c** canopy ($WS_{\text{canopy}}^{\cdot}$), **d** river run-off ($WS_{\text{run-off}}^{\cdot}$) and **e** soil surface temperature



freeze–thaw processes (Cryoturbation) in the active layer accelerates the burial process by mixing the carbon-rich organic soil from the surface down to the permafrost horizon (Zech et al. 2008). Warming in high latitudes leads to a soil warming and permafrost thaw, which results in carbon decomposition by soil microbes (Schuur et al. 2009) and releasing the greenhouse gases as carbon dioxide and methane (Koven et al. 2011). Studies and observations indicate a notable widespread permafrost thawing in the Northern Hemisphere (Lemke et al. 2007). For example, Osterkamp (2007) reported an increasing trend in the permafrost temperature by 2–3 °C over the last two decades at depth of

20 m and also at depths up to 20 m increased 0–2 °C in Canada (cf. Smith et al. 2004 and Lemke et al. 2007). Similar results presented in Siberia by Lemke et al. (2007) show the temperature increases between 0.3 and 2.8 °C at depths up to 10 m.

For this study, we used the monthly gridded 2.5×2.0 arc-degree data about greenhouse gases emissions provided by NASA (https://mirador.gsfc.nasa.gov/collections/AIRX3_C2M_005.shtml). In particular, we used data of the AIRS mid-tropospheric carbon dioxide collected by the atmospheric infrared sounder (AIRS) and the advanced microwave sounding unit (AMSU) instruments on board of the Aqua

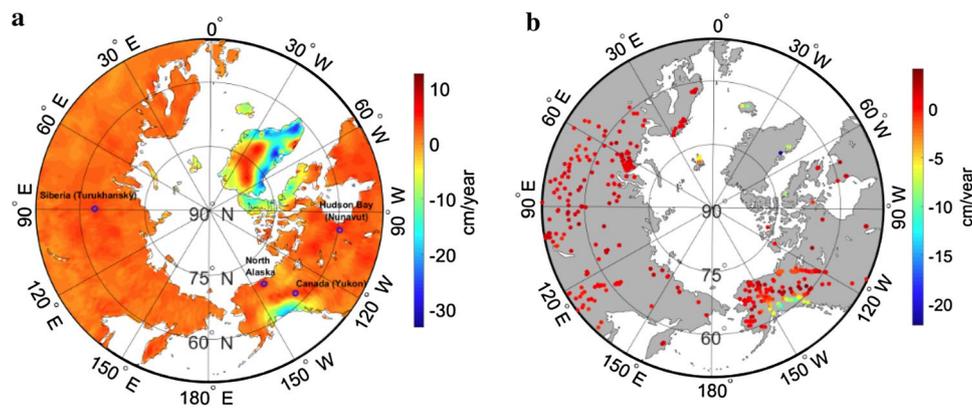


Fig. 6 **a** Regional map of the equivalent water height secular trend due to permafrost thawing after removing hydrological contribution obtained using GRACE and ICESat data and GLDAS/NOAH land hydrological model during 2002/08–2016/05 in the northern high-lat-

itude region. Blue small circles show locations of selected stations, **b** shows the distribution (boreholes stations) and secular trend of equivalent water height due to permafrost in the study region

Table 1 Statistics of total water storage (TWS), GIA, hydrological signal and glaciers melting. Unit: cm/year

	Max	Mean	Min	Standard deviation
Total water storage (TWS_{GRACE})	6.1	-0.3	-27.8	3.2
GIA correction (ELUWH)	4.0	0.1	-1.8	0.8
Snow (WS_{snow})	27.5	0.6	-22.6	5.3
Canopy (WS_{canopy}) $\times 10^{-4}$	25	-3.2	33	5.7
River run-off ($WS_{run-off}$)	0.7	0.0	-2.6	0.1
Soil Moisture ($WS_{SoilMoisture}$)	2.8	-0.4	-3.0	0.6
Ice melting ($WS_{ice-melting}$)	25.9	-0.2	-22.1	1.2

satellite in mole fraction units (PPM in volume). These data cover the period of 110 months from 1 January 2003 to 1 February 2012. AIRS is a facility instrument lunched into 705 km altitude and covered a period from 1 September 2002 to 29 February 2012.

The CO_2 time series is illustrated in Fig. 7 at four selected stations in Siberia (Turukhansky), North Alaska (Sagwon) and two more stations in Canada (Yukon and East of Hudson Bay), where reliable in situ data are available for the same time span as GRACE data. CO_2 variations at these four

stations are obviously not fully attributed to only the permafrost thawing, but are also affected by different phenomena. The analysis of the CO_2 data shows generally positive trend in the study region. The permafrost contains a staggering 850 gigatonnes of frozen carbon (i.e. dry ice) that is much more than the total of the carbon currently contained in the earth's atmosphere, much of which could be released in the form of methane (<https://nsidc.org/cryosphere/frozensound/methane.html>). Temperature, CO_2 and methane changes are capable of retaining heat in the atmosphere that may accelerate the possibility of catastrophic global warming (cf. Schaefer et al. 2011). It was confirmed by several studies that thawing soils due to permafrost active layer are now the causes pouring carbon dioxide into the air.

The secular trends in CO_2 and soil temperature (provided via NOAA-GLDAS) were extracted for the study region and compared with the EWHP obtained from GRACE (see Fig. 8). The slope of linear trend of the EWHP, soil surface temperature and CO_2 at selected stations is summarized in Table 2. The result shows that the secular trends at all stations are rising, with a relatively small increase in the soil surface temperature corresponding to relatively large secular trends in the EWHP and CO_2 .

Table 2 Secular trends of EWHP, average surface temperature and CO_2 at selected stations

Location	EWHP (cm/year)	Soil surface temperature ($^{\circ}C/year$)	Soil surface temperature (Celsius within 100 years)	CO_2 (PPM)
Siberia (Turukhansky)	+3.4	+0.013	+1.2	2.31
North Alaska (Sagwon)	+3.8	+0.081	+8.1	2.16
Canada (Yukon)	+4.4	-0.114	-11.4	2.21
Canada (Hudson Bay, Nunavut)	+4.0	+0.050	+5.0	2.20

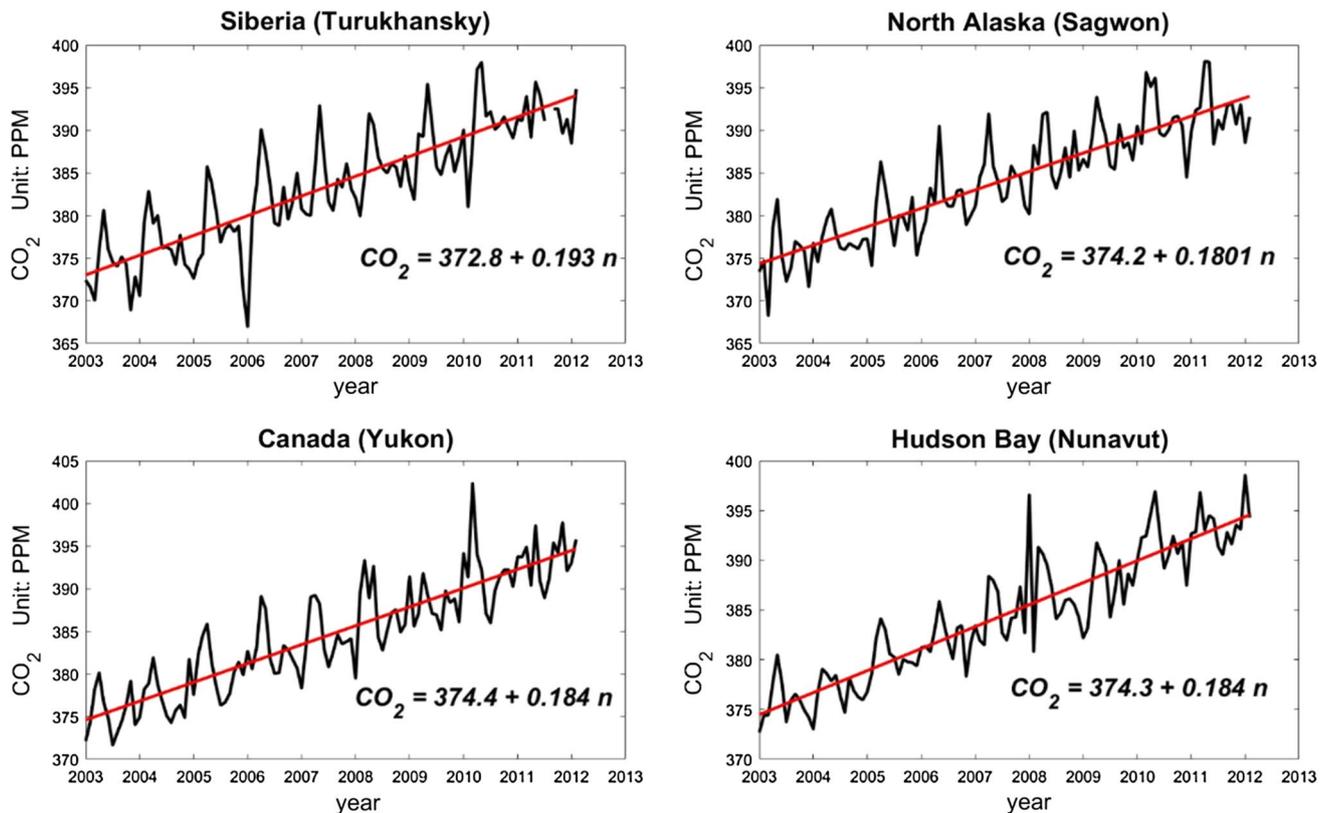


Fig. 7 Carbon dioxide (CO_2) trend at selected stations in the study region. Note that “ n ” denotes on the number of months from initial epoch

Discussion

The permafrost plays important role also in present-day processes related to the earth’s system, because there is a good data record suitable for studying the long-term characteristics of this phenomenon. In this study, we focused on a large-scale permafrost monitoring as well as an identification of permafrost hotspots using GRACE monthly solutions. However, these data required the application of additional corrections in order to achieve reliable and accurate products. We studied the permafrost thawing according to Eq. (11) by applying two numerical steps. Firstly, we removed the GIA effect from TWS obtained from GRACE data. This correction was computed according to the ICE-6G (VM5a) model (Peltier and Argus 2015). In this step, the role of GIA in the permafrost thawing determination is a vital issue to derive the EWHP secular trend. It is thus important to choose the most accurate GIA model; see an overview of different GIA models in Huang (2013). Our test results, not shown herein, revealed that the ICE-6G (VM5a) model has the best RMS fit with the GPS velocity rates in Fennoscandia and Laurentia. We further used the GLDAS data to investigate and remove the effect of various hydrological contributions in the EWHP modelling. By analogy with the role of GIA model in the EWHP determination, it is

also important to study and choose the best and most precise hydrological model that we did not focus on this matter in this paper. Other large-scale lithospheric motions, such as tide effects, will not affect our results due to their periodic character. Moreover, most of these periodic motions have already been removed from the GRACE data.

The EWHP map in Fig. 6a revealed the gravity signature of which pattern corresponds with a total ground water change due to the permafrost layer thawing. Negative values prevail over regions with glacial cover where the ice masses are decreasing (Greenland and Alaska), while other areas are dominated by positive values (i.e. increasing equivalent water storage). A possible method to overcome this issue (i.e. problem of removing the effect of ice melting from EWHP in Fig. 4b) is to use external ice data information such as ICESat data. However, even after removing the surface mass change obtained from ICESat data, a strong negative ice-melting effect in Greenland and Alaska still remains. This ice-melting signal is stronger than the permafrost active layer signal particularly in Greenland and Alaska. Possible reasons for ice melting involve global warming, seasonal and inter-annual periodic signals and warm winds from oceans towards continents. The ice-melting effect is remained even after correcting TWS using ICESat data in the EWHP

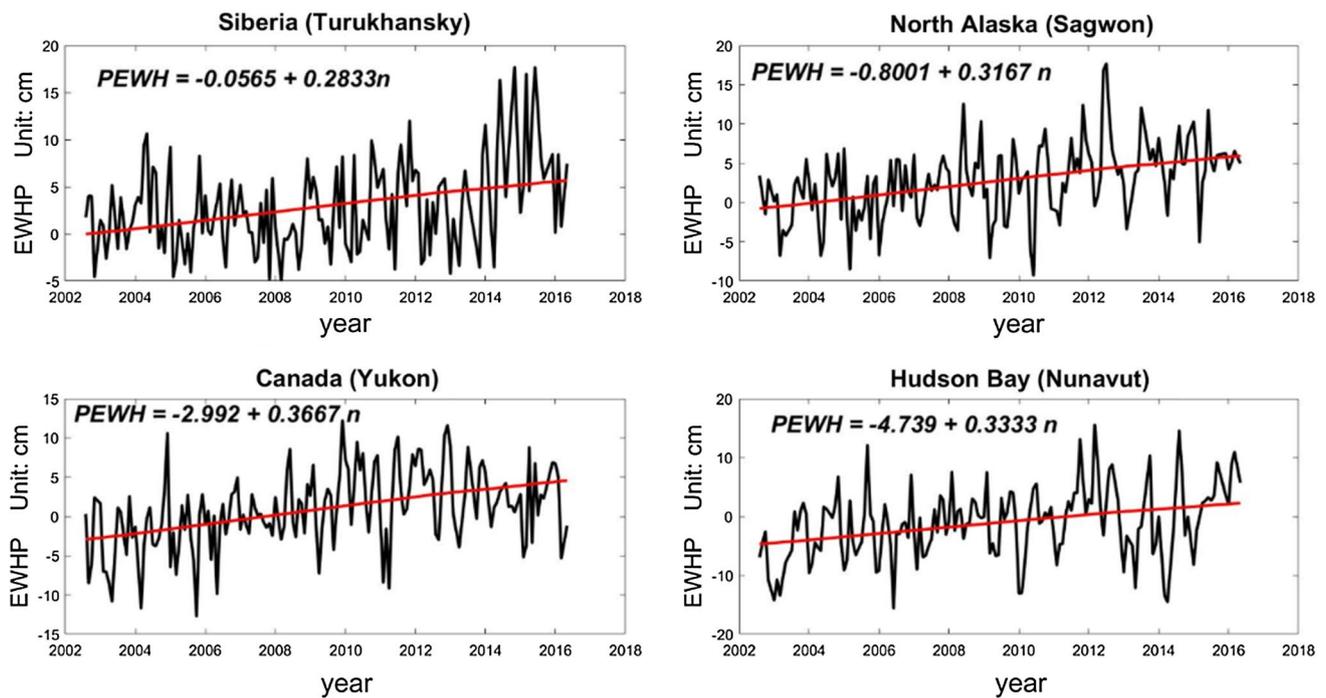


Fig. 8 Equivalent water height due to permafrost (EWHP) time series at selected stations in the study area. Note that “*n*” denotes the number of months since the initial epoch

(Fig. 6a), because the ICESat measures only a surface mass change instead of a mass change in deeper layers.

According to the calculated EWHP values in the northern high-latitude region, our results and those from the other studies are similar. For example see Velicogna et al. (2012) and Shabanloui and Müller (2015). We find that the EWHP varies from 2 to 4 cm/year. Shabanloui and Müller (2015) studied the permafrost thawing in Siberia based on GRACE solutions and GLDAS model. They reported that the time series as well as secular trends of GLDAS hydrological mass variations are similar to GRACE results, but in some periods of time differences are obvious. The permafrost thawing dynamics and geophysical processes are very complex. However, we should emphasize that the permafrost is a phenomenon that cannot be associated only with the presence of underground ice. Its degradation is not only the melting of the ice. We also observed differences between the GRACE and GLDAS time series that might be explained by permafrost thawing effects not modelled in GLDAS. Therefore, GLDAS does not include permafrost contributions and it shows the precipitation and surface water masses. Hence, the estimated positive secular trends for the GRACE minus GLDAS results during the period of 2002–2016 may show that the permafrost thawing is progressing in the target region. This is the reason why we removed soil moisture, canopy and river run-off (obtained using GLDAS model) from GRACE observations in order to model the permafrost

thawing more realistically. In addition, it should be emphasized that our obtained results using GRACE data are preliminary and can be improved if one combines it with in situ data (e.g. using the published data in the Global Terrestrial Network for Permafrost (GTN-P)). It is also worth mentioning that it would be possible to determine the uncertainty of EWHP, if we had access to all errors in Eq. (11). Only the accuracies of the harmonic coefficients are available for the GRACE data.

The surface temperature is a main component in the climate change and permafrost carbon feedback processes. The obtained soil surface temperature (from GLDAS) varies from -3 to 3 °C/year in the study region and it shows generally positive trend with a maximum impact in the northern part of Canada (see Fig. 5e). Few scattered areas include small parts in Greenland, and Canada has a negative temperature trend. Three locations out of the four selected stations show positive trend ranged from $+0.013$ to $+0.081$ °C/year, and one shows average negative trend of -0.114 °C/year (see Table 2). Therefore, as the air and soil temperature rose over time, the active layer part of the permafrost and the permafrost temperature will increase forming a talik layer, where the permafrost top part remains at 0 °C and the bottom of the active layer froze/melt seasonally.

The CO_2 analysis shows a positive trend all over the study region, ranging from 1.9 to 2.5 PPM/year. The trend for the selective stations ranges from 2.16 to 2.31 PPM/

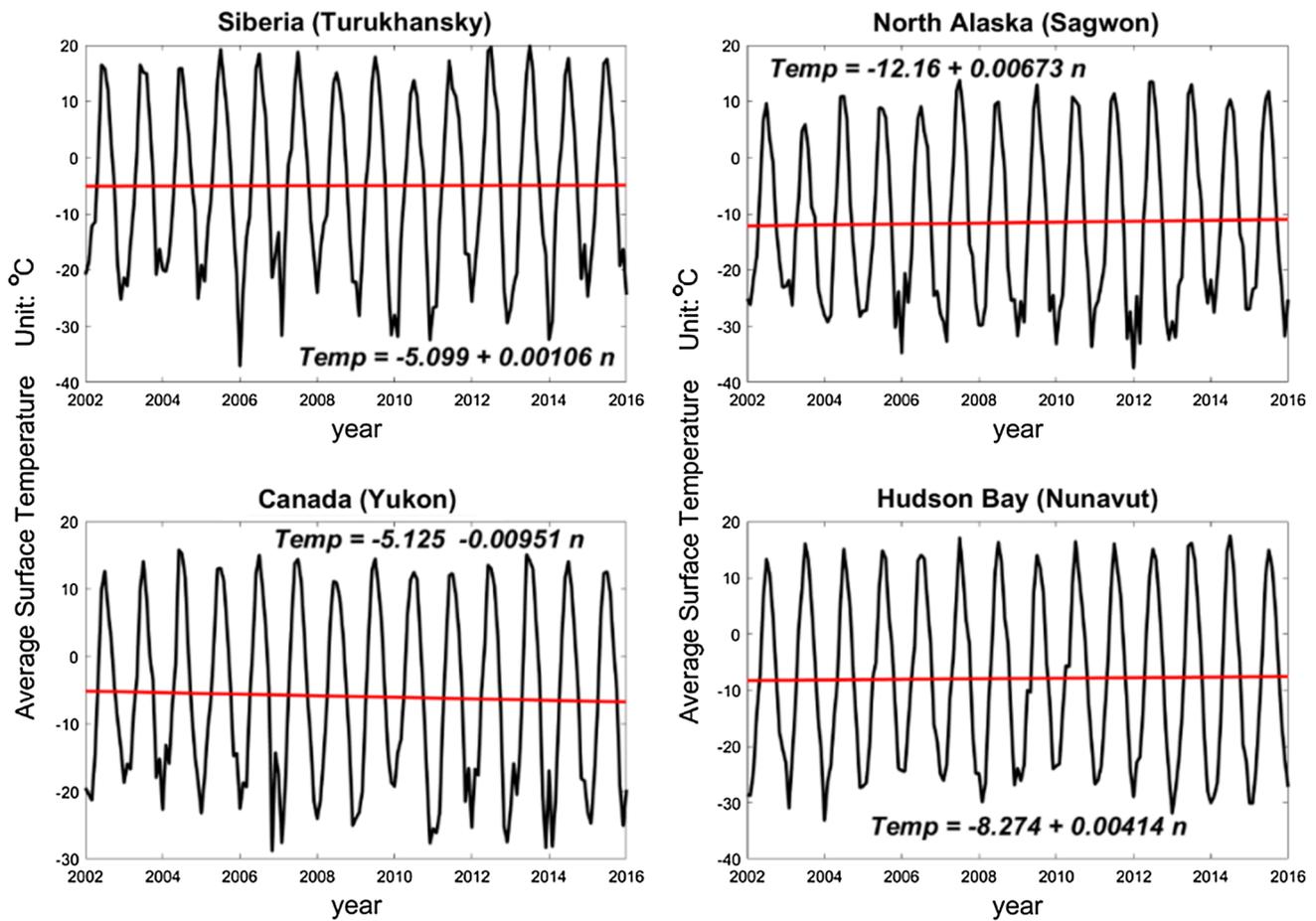


Fig. 9 Soil surface temperature time series at selected stations in the study region. Note that “*n*” denotes the number of months from initial epoch

year. Our results show that CO_2 increases in average of 2.2 PPM/year in the selected regions, and there is correlation between the secular trend of greenhouse gases and the equivalent water height (i.e. EWHF shown in Fig. 8) that is due to the thawing process and temperature increase. Consequently, our results can be used to determine future emissions and the impact of CO_2 /methane releases. The comparison of the permafrost equivalent water thickness with CO_2 secular trends in Fig. 7 shows that frozen soils (permafrost) are unleashing an increasing amount of CO_2 into the air as they thaw in summer or subsequently fail to refreeze as they once did, particularly in late fall and early winter. However, we should emphasize that the release of CO_2 in the seasonally frozen layer, i.e. the active layer is sporadically occurring but the biologic processes of release and consumption, is highly variable and not well understood nor measured at sufficient detail and scale for a meaningful conclusion at this time. Our results also indicate an increasing amount of CO_2 loss (Fig. 7) as well as an increasing surface temperature at selected stations

(Fig. 9) in late spring and summer as the climate warmest steadily. Obviously, the large areas of permafrost in Canada and Siberia are partially responsible for some of the carbon dioxide emissions into the atmosphere. Since the Carbon dioxide can originate from anywhere on the Earth, proving that the carbon is coming from the thawed permafrost need special measuring tool and looking at the isotopic signatures to see that the carbon is old or coming from recent carbon cycling. Schaefer et al. (2012) stated that permafrost carbon starts thawing when the active layer exceeds certain thickness in the warm year. Therefore, in our case we conclude that the obtained CO_2 could be partially or completely from the thawed permafrost.

Conclusions

We have used monthly GRACE data, GIA and GLDAS models to study the groundwater storage change and permafrost thawing in the northern high-latitude region. The used time series covered the period from August 2002

to May 2016. Subtracting the GIA effect and hydrological water cycle contribution from the total water storage obtained using GRACE data reveals changes in the groundwater storage. Our result shows that the GIA effect is significant for determining the water storage change. We postulate that these water storage changes can be related to the permafrost thawing. We obtained negative values of the equivalent water height in Greenland and Alaska regions and we removed these effects from the groundwater storage using ICESat data. Permafrost thawing might be detected with positive values of the equivalent water height, and it quantifies the increase in the ground water storage. We showed that the ground water storage is changing by up to 3.4, 3.8, 4.4 and 4.0 cm/year in Siberia, North of Alaska and Canada (Yukon and Hudson Bay), respectively, most probably due to permafrost thawing using GRACE, GLDAS and ice-melting data. We also illustrated those greenhouse gases, i.e. carbon dioxide (CO₂) and methane, increased up to 2.2 PPM/year in the study areas over the same period. Temperature also shows positive trend generally and in the study area as well as CO₂ and equivalent water height trend (due to permafrost thawing) which might confirm the permafrost carbon feedback process.

Acknowledgements Shfaqat A. Khan from DTU Space, National Space Institute, Technical University of Denmark, Department of Geodesy is cordially acknowledged for preparing ICESat data for this study. Also, the first author would like to thank Faramarz Nilfouroushan, University of Gävle, Sweden, for his great support.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ågren J, Svensson R (2007) Postglacial land uplift model and system definition for the new Swedish height system RH 2000. Reports in geodesy and geographical information systems, LMV-Rapport:4. Lantmäteriet, Gävle, Sweden
- Bevis M, Melini D, Spada G (2016) On computing the geoelastic response to a disk load. *Geophys J Int* 205(3):1804–1812. <https://doi.org/10.1093/gji/ggw115>
- Biskaborn BK, Lanckman JP, Lantuit H, Elger K, Streletskiy DA, Cable WL, Romanovsky VE (2015) The global terrestrial network for permafrost database: metadata statistics and prospective analysis on future permafrost temperature and active layer depth monitoring site distribution. *Earth Syst Sci Data Discuss* 8:279–315
- Brown J, Ferrians OJ, Heginbottom JA, Melnikov ES (1998) Circum-Arctic map of permafrost and ground-ice conditions. Boulder, CO: National Snow and Ice Data Center/World Data Center for Glaciology. Digital Media
- Brown J, Hinkel KM, Nelson FE (2000) The circumpolar active layer monitoring (CALM) program: research designs and initial results. *Polar Geogr* 24:165–258
- Chao BF, Wu YH, Zhang Z, Ogawa R (2011) Gravity variation in Siberia: GRACE observation and possible causes. *Terr Atmos Ocean Sci* 22:149–155. [https://doi.org/10.3319/tao.2010.07.26.03\(tibxs\)](https://doi.org/10.3319/tao.2010.07.26.03(tibxs))
- Chen F, Mitchell K, Schaake J, Xue Y, Pan H, Koren V, Duan Y, Ek M, Betts A (1996) Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *J Geophys Res* 101(D3):7251–7268
- Ekman M (1991) A concise history of postglacial land uplift research (from its beginning to 1950). *Terra Nova* 3:358–365
- Farrell WE (1972) Deformation of the Earth by surface loading. *Rev Geophys* 10:761–797
- Frauenfeld OW, Zhang T, Barry RG, Gilichinsky D (2004) Interdecadal changes in seasonal freeze and thaw depths in Russia. *J Geophys Res*. <https://doi.org/10.1029/2003jd004245>
- French HM (2013) The periglacial environment. John Wiley & Sons, Hoboken
- Han S-C, Shum CK, Jekeli C, Kuo C-Y, Wilson C, Seo K-W (2005) Non-isotropic filtering of GRACE temporal gravity for geophysical signal enhancement. *Geophys J Int* 163:18–25. <https://doi.org/10.1111/j.1365-246x.2005.02756>
- Huang Z (2013) The role of glacial isostatic adjustment (GIA) process on the determination of present-day sea-level rise. Report No. 504. Geodetic Science The Ohio State University Columbus, Ohio. 43210 October
- Hugelius G, Virtanen T, Kaverin D, Pastukhov A, Rivkin F, Marchenko S, Romanovsky V, Kuhry P (2011) High-resolution mapping of ecosystem carbon storage and potential effects of permafrost thaw in periglacial terrain, European Russian Arctic. *J Geophys Res Biogeosci* 116:G03024. <https://doi.org/10.1029/2010JG001606>
- Joud SM, Sjöberg LE, Bagherbandi M (2017) Use of GRACE data to detect the present land uplift rate in Fennoscandia. *Geophys J Int* 209(2):909–922. <https://doi.org/10.1093/gji/ggx063>
- Khan SA, Kjær KH, Bevis M, Bamber JL, Wahr J, Kjeldsen KK, Bjørk AA, Korsgaard NJ, Stearns LA, van den Broeke MR, Liu L, Larsen NK, Muresan IS (2014) Sustained mass loss of the Northeast Greenland ice sheet triggered by regional warming. *Nat Clim Chang* 4:292–299. <https://doi.org/10.1038/nclimate2161>
- Klees R, Revtova EA, Gunter BC, Ditmar P, Oudman E, Winsemius HC, Savenije HHG (2008) The design of an optimal filter for monthly GRACE gravity models. *Geophys J Int* 175:417–432. <https://doi.org/10.1111/j.1365-246x.2008.03922.x>
- Koren V, Schaake J, Mitchell K, Duan QY, Chen F, Baker JM (1999) A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *J Geophys Res* 104:19569–19585
- Koven DC, Ringeval B, Friedlingstein P, Ciais P, Cadule P, Khvorostyanov D, Krinner G, Tarnocai C (2011) Permafrost carbon-climate feedbacks accelerate global warming. *PNAS* 2011 108(36):14769–14774. <https://doi.org/10.1073/pnas.1103910108>
- Kusche J (2007) Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models. *J Geodesy* 81(11):733–749
- Kusche J, Schmidt R, Petrovic S, Rietbroek R (2009) Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model. *J Geodesy*. <https://doi.org/10.1007/s00190-009-0308-3>
- Lawrence DM, Koven CD, Swenson SC, Riley WJ, Slater AG (2015) Permafrost thaw and resulting soil moisture changes regulate projected high-latitude CO₂ and CH₄ emissions. *Environ Res Lett*. <https://doi.org/10.1088/1748-9326/10/9/094011>
- Lemke P, Ren J, Alley RB, Allison I, Carrasco J, et al. (2007) Observations: changes in snow, ice and frozen ground. In: Solomon S, Qin

- D, Manning M, Chen Z, Marquis M, et al. (eds) Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the Intergovernmental panel on climate change. Cambridge University Press, Cambridge, UK and New York, NY, USA
- Lemoine FG, Luthcke SB, Rowlands DD, Chinn DS, Klosko SM, Cox CM (2007) The use of mascons to resolve time-variable gravity GRACE. In: Tregoning P, Rizos C (eds) Dynamic planet: monitoring and understanding a dynamic planet with geodetic and oceanographic tools. Springer, Berlin, pp 231–236. ISBN 978-3-540-49349-5
- Liu L, Zhang T, Wahr J (2010) InSAR measurements of surface deformation over permafrost on the North Slope of Alaska. *J Geophys Res* 115:F03023. <https://doi.org/10.1029/2009jf001547>
- Moore P, Zhang Q, Alothman A (2005) Annual and semiannual variations of the Earth's gravitational field from satellite laser ranging and CHAMP. *J Geophys Res*. <https://doi.org/10.1029/2004jb003448>
- Muskett RR, Romanovsky VE (2009) Groundwater storage changes in Arctic permafrost watersheds from GRACE and in-situ measurements. *Environ Res Lett* 4(4):045009
- Muskett RR, Romanovsky VE (2011) Alaskan permafrost groundwater storage changes derived from grace and ground measurements. *Remote Sens* 3:378–397. <https://doi.org/10.3390/rs3020378>
- Osterkamp TE (2007) Characteristics of the recent warming of permafrost in Alaska. *J Geophys Res*. <https://doi.org/10.1029/2006jf0005788>
- Peltier WR (2004) Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE, invited paper. *Ann Rev Earth Planet Sci* 32:111–149
- Peltier WR, Argus DF (2015) Drummond, R. Space geodesy constrains ice-age terminal deglaciation: the global ICE-6G_C VM5a model. *J Geophys Res Solid Earth* 120:450–487. <https://doi.org/10.1002/2014jb011176>
- Phillips M, Springman SM, Arenson LU (eds) (2003) Proceedings of the eighth international conference on permafrost, Zurich, Switzerland, 21–25 July, vol 1 and 2. A.A. Balkema, Lisse, The Netherlands, 1319 p
- Rodell M, Famiglietti JS, Chen J, Seneviratne SI, Viterbo P, Holl S, Wilson CR (2004) Basin scale estimates of evapotranspiration using GRACE and other observations. *Geophys Res Lett*. <https://doi.org/10.1029/2004GL020873>
- Schaefer K, Zhang T, Bruhwiler L, Barrett AP (2011) Amount and timing of permafrost carbon release in response to climate warming. *Tellus B* 63:165–180. <https://doi.org/10.1111/j.1600-0889.2011.00527.x>
- Schaefer K, Lantuit H, Romanovsky V, Schuur EAG (2012) Policy implications of warming permafrost. United Nations Environment Programme Special Report, Nairobi, Kenya, p 50
- Schuur EAG, Vogel JG, Crummer KG, Lee H, Sickman JO, Osterkamp TE (2009) The effect of permafrost thaw on old carbon release and net carbon exchange from tundra. *Nature* 459:556–559
- Shabanloui A, Müller J (2015) Mass variations in the siberian permafrost region based on new GRACE results and auxiliary modeling. In: International association of geodesy symposia. Springer International Publishing Switzerland. https://doi.org/10.1007/1345_2015_186
- Sjöberg LE, Bagherbandi M (2017) Gravity inversion and integration: theory and applications in geodesy and geophysics. Springer, Berlin
- Sjöberg Y, Frampton A, Lyon S (2013) Using streamflow characteristics to explore permafrost thawing in northern Sweden. *Hydrogeol J* 21(1):121–131. <https://doi.org/10.1007/s10040-012-0932-5>
- Smith NV, Saatchi SS, Randerson T (2004) Trends in high latitude soil freeze and thaw cycles from 1988 to 2002. *J Geophys Res*. <https://doi.org/10.1029/2003jd004472>
- Steffen H, Müller J, Peterseim N (2012) Mass variations in the Siberian permafrost region from GRACE. In: Kenyon S, Pacino MC, Marti U (eds) Geodesy for planet earth, vol 136. Springer, Berlin, pp 597–603
- Sun W, Sjöberg LE (1999) Gravitational potential changes of a spherically symmetric earth model caused by a surface load. *Geophys J Int* 137:449–468
- Swenson S, Wahr J (2006) Post-processing removal of correlated errors in GRACE data. *Geophys Res Lett* 33:L08402. <https://doi.org/10.1029/2005gl025285>
- Swenson S, Wahr J, Milly PCD (2003) Estimated accuracies of regional water storage variations inferred from the gravity recovery and climate experiment (GRACE). *Water Resour Res* 39:1223. <https://doi.org/10.1029/2002wr001808>
- Tapley BD, Bettadpur S, Watkins M, Reigber C (2004) The gravity recovery and climate experiment: mission overview and early results. *Geophys Res Lett* 31:L09607. <https://doi.org/10.1029/2004GL019920>
- Treat CC, Wollheim WM, Varner RK, Grandy AS, Talbot J, Frolking S (2014) Temperature and peat type control CO₂ and CH₄ production in Alaskan permafrost peats. *Glob Chang Biol* 20(8):2674–2686. <https://doi.org/10.1111/gcb.12572>
- Velicogna I, Tong J, Zhang T, Kimball JS (2012) Increasing subsurface water storage in discontinuous permafrost areas of the Lena river basin, Eurasia, detected from GRACE. *Geophys Res Lett*. <https://doi.org/10.1029/2012gl016273>
- Vey S, Steffen H, Müller J, Boike J (2013) Inter-annual water mass variations from GRACE in central Siberia. *J Geodesy* 87(3):287–299
- Wahr J, Molenaar M, Bryan F (1998) Time variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE. *J Geophys Res Solid Earth* 103(B12):30205–30229
- Wahr J, Swenson S, Zlotnicki V, Velicogna I (2004) Time-variable gravity from GRACE: first results. *Geophys Res Lett* 31(11):L11501. <https://doi.org/10.1029/2004GL019779>
- Wouters B, Schrama EJO (2007) Improved accuracy of GRACE gravity solutions through empirical orthogonal function filtering of spherical harmonics. *Geophys Res Lett* 34:L23711. <https://doi.org/10.1029/2007gl032098>
- Wu Q, Zhang T (2010) Changes in active layer thickness over the Qinghai Tibetan Plateau from 1995 to 2007. *J Geophys Res* 115:D09107. <https://doi.org/10.1029/2009jd012974>
- Yang Z, Gao J, Zhao L, Xu X, Ouyang H (2013) Linking thaw depth with soil moisture and plant community composition: effects of permafrost degradation on alpine ecosystems on the Qinghai-Tibet plateau. *Plant Soil* 367(1–2):687–700. <https://doi.org/10.1007/s11104-012-1511-1>
- Zech M, Zech R, Zech W, Glaser B, Brodowski S, Amelung W (2008) Characterisation and palaeoclimate of a loess-like permafrost palaeosol sequence in NE Siberia. *Geoderma* 143:281–295
- Zhang T, Frauenfeld OW, Serreze MC, Etringer A, Oelke C et al (2005) Spatial and temporal variability in active layer thickness over the Russian Arctic drainage basin. *J Geophys Res* 110:1–9. <https://doi.org/10.1029/2004jd005642>



Correction to: Estimated casualties in possible future earthquakes south and west of the M7.8 Gorkha earthquake of 2015

Max Wyss¹ · Deepak Chamlagain²

Published online: 11 March 2019

© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Correction to: Acta Geophysica

<https://doi.org/10.1007/s11600-019-00265-7>

The original version of this article unfortunately contained a mistake: The affiliation of Deepak Chamlagain was incorrect. Correct it should be:

Department of Geology, Trichandra M. Campus, Tribhuvan University, Kathmandu, Nepal.

The original article can be found online at <https://doi.org/10.1007/s11600-019-00265-7>.

✉ Max Wyss
max@maxwyss.ch

¹ International Centre for Earth Simulation Foundation,
Geneva, Switzerland

² Department of Geology, Trichandra M. Campus, Tribhuvan
University, Kathmandu, Nepal



Correction to: VSP polarization angles determination: Wysin-1 processing case study

Mateusz Zaręba^{1,2} · Tomasz Danek¹

Published online: 5 March 2019

© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Correction to: Acta Geophysica (2018) 66:1047–1062
<https://doi.org/10.1007/s11600-018-0200-8>

In the paper, the authors used data acting on the base of public procurement ZDN/014/140/2016 and related permission. They acknowledged that these data were gathered on behalf of the Faculty of Geology, Geophysics and Environmental Protection. However, unfortunately, the details related to data ownership, involvement in survey preparation as well as data acquisition were not properly provided. This omission was not intentional. Given the importance of these information, herein authors would like to acknowledge tremendous work done by Department of Fossil Fuels, a part of Faculty of Geology, Geophysics and Environmental Protection, especially by Andrzej Pasternacki, Jacek Płoskonka, Janusz Lesiak, Tomasz Maćkowski and Michał Stefaniuk. Their and our works were done in the context of scientific project GASŁUPSEJSM, a part of Blue Gas I project no. BG1/GASŁUPSEJSM/13 financed by National Center of Research and Development (NCBiR), co-financed by Polskie Górnictwo Naftowe i Gazownictwo S.A. and Orlen Upstream. The authors hope that this supplementary information as well as original acknowledgements satisfies all parts involved in the project and allows better future collaboration.

Additionally, in the original paper second author affiliation, namely Tomasz Danek's, is wrong. It should be: The Faculty of Geology, Geophysics and Environmental Protection, AGH UST, Kraków, Poland (1).

The original article can be found online at <https://doi.org/10.1007/s11600-018-0200-8>.

✉ Tomasz Danek
tdanek@agh.edu.pl

¹ The Faculty of Geology, Geophysics and Environmental Protection, AGH UST, Kraków, Poland

² R&D Department of Seismic Signal Processing, Geofizyka Toruń S.A., Toruń, Poland