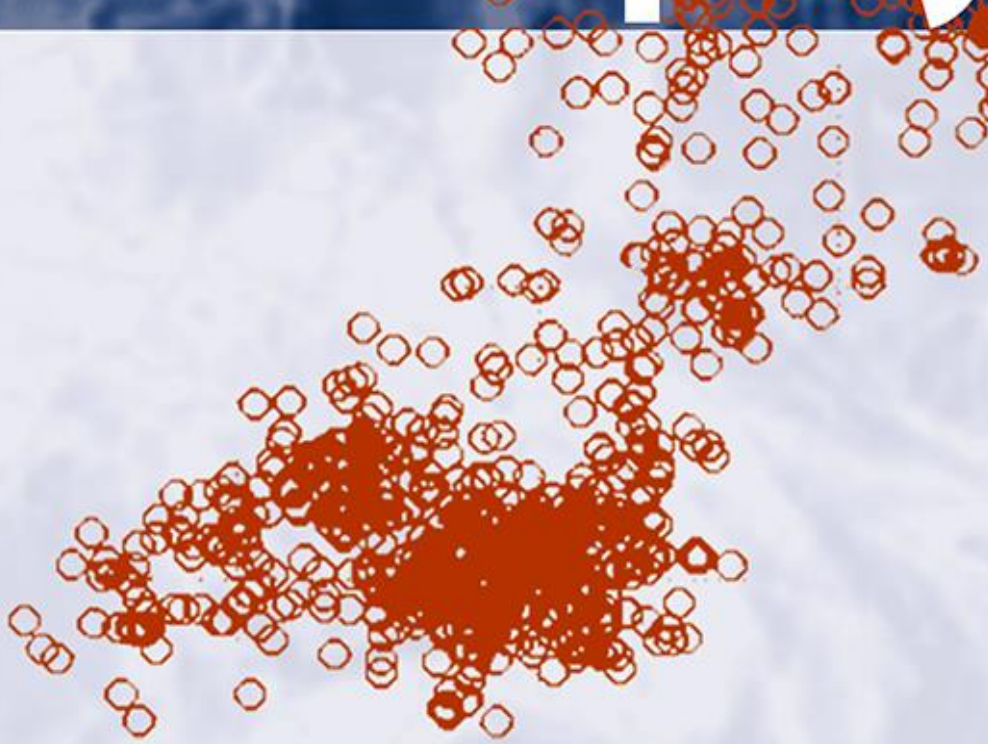
Volume 65 • Number 2 • April 2017

Acta Geophysica



PAN
POLISH ACADEMY OF SCIENCES

Institute of Geophysics
Polish Academy of Sciences

Springer

Acta Geophys. (2017) 65:269–274
DOI 10.1007/s11600-017-0025-x

RESEARCH ARTICLE

Propagation of Love waves in a void medium over a sandy half space under gravity

Pulak Patra[1] · **Asit Kumar Gupta**[2] · **Santimoy Kundu**[3]

Received: 21 February 2017 / Accepted: 1 March 2017 / Published online: 8 March 2017

Abstract The present study investigates the propagation of Love wave in a void layer resting over a sandy half space under the effect of gravitational force. The equations of motion have been gathered separately for different layers, and the boundary conditions have been introduced for two different layers at their interface. The mathematical analysis of the problem has been dealt with the help of Whittaker's function by expanding it asymptotically up to linear terms. The study reveals that in such a situation there exist two different wave fronts for the two above-mentioned layers: one is for the effects of gravity and sandy parameters, whereas other is for the effect of void parameter.

Keywords Love wave · Biot's parameter · Sandy parameter · Whittaker's function · Phase velocity

✉ Pulak Patra
pulakmath11@gmail.com

Asit Kumar Gupta
asitism@yahoo.co.in

Santimoy Kundu
kundu_santi@yahoo.co.in

[1] Department of Mathematics, SDET-Brainware Group of Institutions, Barasat, India

[2] Department of Physics, Asansol Engineering College, Asansol, India

[3] Department of Applied Mathematics, Indian School of Mines, Dhanbad, India

Introduction

It is a well-known fact that the earth's interior contains different hypothetical layers like crust, mantle and core; the seismic waves travel through these different layers with different situations. Theoretically for the earth, neglecting the effect of gravity force, body forces in the equation of motion might be used to represent the process that generates earthquakes. During an earthquake, the seismic waves travel through the layer and make devastations for the earth surface. In the year 1911, Edward Hough Love first observed a different type of surface wave during earthquake and named it the 'Love Wave'; they were actually first introduced in his renown book 'Some Problems of Geodynamics' (Love 1911). After Love's remarkable observation, many authors have started investigation into the generation and propagation of Love waves under various conditions. In an elastic medium, Cowin and Nunziato (1983) performed some remarkable work. For the propagation of Love wave in a poroelastic medium, Cowin (1985) extended the work into a linear elastic material. A quite good amount of information about the seismic waves is available in the well-known books of Ewing et al. (1957), Achenbach (1973), Anderson (1962), and Chapman (2004). The mechanical constructive equation satisfying the physical properties had been studied by Cowin and Nunziato (1983), and then, Dey and Gupta (1987) also figured out a conclusion that there may be two longitudinal wave fronts in such a void medium. With a void medium, many authors performed their investigations and drawn some interesting conclusions such as those by Biot (1955, 1962), Cowin (1985), Dey et al. (2004), Kundu et al. (2014).

In the case of a sandy medium, the present authors have considered for this study; Dey and Chandra (1983) have made some remarkable conclusion on the same layer at their research investigation, but before that Biot (1965)

introduced the effect of gravity at the boundary conditions as a body force component. Now it is quite clear that the layers of the soil are supposed to be more sandy than elastic. Hence, for considering the Love wave in soil models, it is more realistic to study them in a sandy medium.

In the present study, the authors investigate the propagation of Love wave in a void medium lying over a gravitating dry sandy half space. The authors are considering two different layers; one of them is the porous medium or void layer, and the other is a half space with sandy materials under the action of gravitational force. The porous or void layer is filled with different pores of different sizes, actually being considered as the void parameter and in case of half space for sandiness of the layer it considered as sandy parameter. The study reveals that in the assumed two media two different wave fronts are available; one of them is affected by the gravity and sandy parameter, and the other wave front is affected by the void parameter, and interestingly, both the wave fronts are dispersive in nature.

The geometry of the problem, as discussed earlier, consists of an elastic layer of thickness H with void pores over a vertically positioned dry sandy half space under the action of gravitational field. The origin is chosen in the interface of the void layer and the sandy gravitating half space. The z-axis is downward positive, as shown in Fig. 1.

Equation of motion and solution for the upper layer

For propagation of Love wave, we assume $u = 0$, $w = 0$ and $v = v(x, z, t)$. Then, the equation of motion for the upper layer, under no body forces, can be considered as:

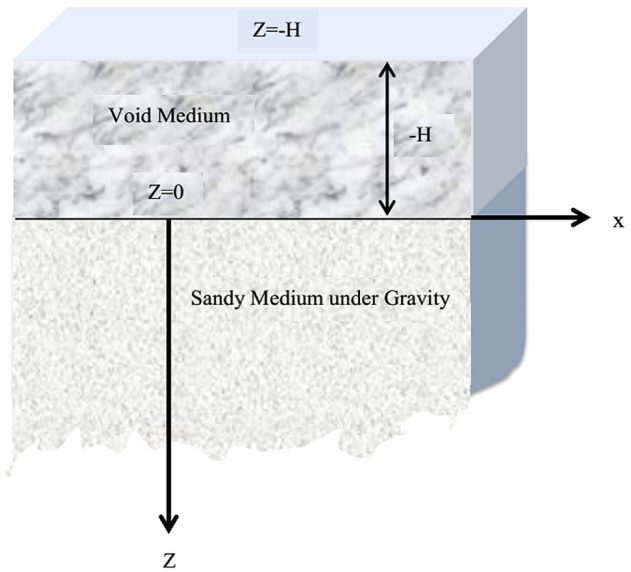


Fig. 1 Geometry of the problem

$$\rho\frac{\partial^2 v}{\partial t^2} = \mu\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial z^2}\right) + \beta\left(\frac{\partial \phi}{\partial x} + \frac{\partial \phi}{\partial z}\right) \tag{1}$$

$$\rho\bar{k}\frac{\partial^2 \phi}{\partial t^2} = \alpha\left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial z^2}\right) - \omega\frac{\partial \phi}{\partial t} - \xi\phi \tag{2}$$

where ρ is the density of the upper layer; μ, β are the Lame's constants, $\alpha, \bar{k}, \xi$ are the constants for a particular material, and ω is the damping term.

For wave propagation in the positive direction of the X-axis with velocity c, the solutions of Eqs. 1 and 2 may be taken as:

$$v = \bar{v}(z)e^{ik(x-ct)}$$

$$\phi = \bar{\phi}(z)e^{ik(x-ct)}$$

Then, from Eq. 1, we get

$$\bar{v}''(z) - N^2\bar{v}(z) + B\left[ik\bar{\phi}(z) + \bar{\phi}'(z)\right] = 0 \tag{3}$$

where $N = k\left(1 - \frac{c^2}{A^2}\right)^{\frac{1}{2}}, B = \frac{\beta}{\mu}$ and $A = \sqrt{\frac{\mu}{\rho}}$ and from Eq. 2 we get,

$$\bar{\phi}''(z) - M^2\bar{\phi}(z) = 0 \tag{4}$$

where $M = \left[(\alpha k^2 - i\omega kc + \xi - \rho\bar{k}k^2c^2)/\alpha\right]^{\frac{1}{2}}$

Now, ignoring the damping term ω, which is very small for sinusoidal wave, the value of M may be taken as:

$$M = \left[1 - \frac{c^2}{(\alpha/\rho\bar{k})} + \frac{1}{k^2(\alpha/\xi)}\right]^{\frac{1}{2}} \tag{5}$$

Then, the solution of Eq. 4 can be written as:

$$\phi = \left(R_3e^{Mz} + R_4e^{-Mz}\right)e^{ik(x-ct)} \tag{6}$$

or can be represented as:

$$\phi = \left(R_3e^{Mz} + R_4e^{-Mz}\right)e^{i(kx-\omega t)} \text{ as } c = \omega/k \tag{7}$$

Using relations 7 in Eq. 3, the solution becomes

$$v = \left[R_1e^{Nz} + R_2e^{-Nz} - R_3\frac{B(ik+M)}{M^2-N^2}e^{Mz} - R_4\frac{B(ik-M)}{M^2-N^2}e^{-Mz}\right]e^{i(kx-\omega t)} \tag{8}$$

Equation of motion and solution for the half space

The dynamical equation of motion in the half space given by Biot is:

$$\frac{\partial S_{11}}{\partial x} + \frac{\partial S_{22}}{\partial y} + \frac{\partial S_{33}}{\partial z} + \rho_1 g\omega_{32} + \rho_1 gz\frac{\partial \omega_{32}}{\partial z} - \rho_1 gz\frac{\partial \omega_{21}}{\partial x} = \rho_1\frac{\partial^2 v_1}{\partial t^2} \tag{9}$$

where ρ_1 is the density of the half space, g is the gravitational force, S_{ij} are the incremental stress components measured with reference to the rotated axis,

$$e_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{10}$$

$$\text{and } \omega_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i}\right) \tag{11}$$

where ω_{ij} are the components of the rotational vector and $u_i = (u_1, v_1, w_1)$. The stress–strain relations under the hydrostatic initial stresses are the same as in the classical cases and are given by $S_{ij} = \lambda\eta\delta_{ij} + 2\mu_1\eta e_{ij}$, where λ and μ_1 are elastic constants (Lame constant), δ_{ij} is the Kronecker delta, and η is the sandy parameter. For the propagation of Love wave $u_1 = 0$, $w_1 = 0$ and $v_1 = f(x, z, t)$, then from Eq. 9, using relations 10 and 11, we can write,

$$\left(\mu_1\eta - \frac{1}{2}\rho_1 gz\right)\frac{\partial^2 v_1}{\partial x^2} + \left(\mu_1\eta - \frac{1}{2}\rho_1 gz\right)\frac{\partial^2 v_1}{\partial z^2} - \frac{1}{2}\rho_1 g\frac{\partial v_1}{\partial z} = \rho_1\frac{\partial^2 v_1}{\partial t^2} \tag{12}$$

For plane harmonic wave propagation in the *X*-direction, one can adopt,

$$v_1 = f(z)e^{i(kx-\omega t)} \tag{13}$$

Then from Eq. 12, by using relation 13, we get

$$f''(z) - \frac{g}{(2a-gz)}f'(z) + k^2\left[\frac{2c^2}{2a-gz} - 1\right]f(z) = 0 \tag{14}$$

where $a = \frac{\mu_1\eta}{\rho_1}$ and $c = \omega/k$, c being the velocity of the Love wave. Now, substituting $f(z) = \frac{\varepsilon(z)}{\left(\frac{2a-gz}{2}\right)^{\frac{1}{2}}}$ in Eq. 14, we get

$$\varepsilon''(z) + \left[k^2\left(\frac{2c^2}{2a-gz} - 1\right) + \frac{1}{4}\left(\frac{g}{2a-gz}\right)^2\right]\varepsilon(z) = 0 \tag{15}$$

Again substituting $\delta = \frac{2k}{g}(2a - gz)$ into Eq. 15, the equation can be reduced to:

$$\varepsilon''(\delta) + \left[-\frac{1}{4} + \frac{m}{\delta} + \frac{1}{4\delta^2}\right]\varepsilon(\delta) = 0 \tag{16}$$

where $m = \frac{kc^2}{g}$

Equation 16 is well known as the Whittaker equation, whose solution can be taken as:

$$\varepsilon(\delta) = R_6 W_{m,0}(\delta) + R_5 W_{-m,0}(-\delta) \tag{17}$$

where $W_{m,0}(\delta)$ and $W_{-m,0}(-\delta)$ are known as Whittaker functions.

And, as the solution should vanish at $z = \infty$, i.e., for $\delta = -\infty$,we may take the solution as:

$$\varepsilon(\delta) = R_5 W_{-m,0}(-\delta) \tag{18}$$

Hence, the solution of Eq. 12 can be written as:

$$v_1 = R_5\left(\frac{2a-gz}{2}\right)^{-\frac{1}{2}} W_{-m,0}\left[-\frac{2k}{g}(2a-gz)\right]e^{i(kx-\omega t)} \tag{19}$$

Boundary conditions and a particular case

The following boundary conditions must be satisfied by the propagation:

$$\begin{aligned}
&(1)\left(\tau_{yz}\right)_1 = \left(\tau_{yz}\right)_2 && \text{at } z = 0\\
&\text{or, } \mu\frac{\partial v}{\partial z} = \eta\mu_1\frac{\partial v_1}{\partial z} && \text{at } z = 0\\
&(2)\ \left(\hat{n}\cdot\vec{\nabla}\phi\right) = 0 && \text{at } z = 0\\
&(3)\ (\phi) = 0 && \text{at } z = 0\\
&(4)\ v = v_1 && \text{at } z = 0
\end{aligned} \tag{20}$$

$$\begin{aligned}
&(5)\ \left(\hat{n}\cdot\vec{\nabla}\phi\right) = 0 && \text{at } z = -H\\
&(6)\ \left(\tau_{yz}\right)_1 = 0 &&\\
&\text{or, } \mu\frac{\partial v}{\partial z} = 0 && \text{at } z = -H
\end{aligned} \tag{21}$$

Now, by using the boundary conditions 20 and 21 on 7, 8 and 19, we get (expanding the Whittaker function asymptotically up to linear term)

$$R_1\mu N - R_2\mu N - R_3\mu\frac{BM(ik+M)}{M^2-N^2} + R_4\mu\frac{BM(ik-M)}{M^2-N^2} - R_5 a^{-\frac{1}{2}}\cdot\left(e^{\frac{2ak}{g}}\right)\cdot\left(-\frac{4ak}{g}\right)^m\eta\mu_1\cdot\left[\left(\frac{3g}{4ak} - \frac{mg}{2ak} - 1\right)\left(\frac{g}{4a}\left(m+\frac{1}{2}\right)^2 + k\right) - \frac{g}{2a}\right] = 0 \tag{22}$$

$$R_3 - R_4 = 0 \tag{23}$$

$$R_3 + R_4 = 0 \tag{24}$$

$$R_1 + R_2 - R_3\frac{B(ik+M)}{M^2-N^2} - R_4\frac{B(ik-M)}{M^2-N^2} - R_5 a^{-\frac{1}{2}}\left(e^{\frac{2ak}{g}}\right)\left(-\frac{4ak}{g}\right)^m\left[1 + \frac{g\left(m+\frac{1}{2}\right)^2}{4ak}\right] = 0 \tag{25}$$

$$R_3 e^{-MH} - R_4 e^{MH} = 0 \tag{26}$$

$$R_1 N e^{-NH} - R_2 N e^{NH} - R_3\frac{BM(ik+M)}{M^2-N^2}e^{-MH} + R_4\frac{BM(ik-M)}{M^2-N^2}e^{MH} = 0 \tag{27}$$

For nonzero solution of R_1, R_2, R_3, R_4 and R_5—we have

$$\begin{vmatrix} N\mu & -N\mu & -\mu\frac{BM(ik+M)}{M^2-N^2} & \mu\frac{BM(ik-M)}{M^2-N^2} & a^{-\frac{1}{2}}\cdot\left(e^{\frac{2ak}{g}}\right)\cdot\left(-\frac{4ak}{g}\right)^m\eta\mu_1 \\ 0 & 0 & 1 & 1 & \cdot\left[\left(\frac{3g}{4ak}-\frac{mg}{2ak}-1\right)\left(\frac{g}{4a}\left(m+\frac{1}{2}\right)^2+k\right)-\frac{g}{2a}\right] \\ 1 & 1 & -\frac{B(ik+M)}{M^2-N^2} & -\frac{B(ik-M)}{M^2-N^2} & 0 \\ 0 & 0 & e^{-MH} & -e^{MH} & a^{-\frac{1}{2}}\left(e^{\frac{2ak}{g}}\right)\left(-\frac{4ak}{g}\right)^m\left[1+\frac{g\left(m+\frac{1}{2}\right)^2}{4ak}\right] \\ Ne^{-NH} & -Ne^{NH} & -\frac{BM(ik+M)}{M^2-N^2}e^{-MH} & \frac{BM(ik-M)}{M^2-N^2}e^{MH} & 0 \end{vmatrix} = 0$$

Simplifying the above determinant, we geteither

$$\tan\left[kH\sqrt{\frac{c^2}{A^2}-1}\right] = -\frac{\eta\mu_1}{\mu}\frac{1}{\sqrt{\frac{c^2}{A^2}-1}} \left[\frac{3G}{4}-\frac{c^2}{2a}-1-\frac{2G}{4G+\left(\frac{c^2}{a}+\frac{G}{2}\right)^2}\right] \quad (28)$$

where $G=\frac{\rho_1 g}{\eta\mu_1 k}$ is the Biot's gravity parameter for the sandy half space.

or,

$$2+(kH)^2\left[1-\frac{c^2}{c_3^2}+\frac{1}{(kR)^2}\right]=0 \quad (29)$$

where $c_3=\sqrt{\frac{\alpha}{\rho k}}$ is the velocity of the wave due to change in void fraction; k is the wave number; $R=\sqrt{\frac{\alpha}{\xi}}$ is the displacement parameter; and$c$ is the velocity of the Love wave.

The study shows that the Love wave propagates in a void medium over a sandy half space under gravity, with two wave fronts: one is given in Eq. 28, which arises due to gravity and sandy parameters, while the other wave front is given in Eq. 29 which depends on void parameters.

When $G\to 0, \eta\to 1, \alpha\to 0$, then Eq. 29 is completely vanished and Eq. 28 reduces to

$$\tan\left[kH\sqrt{\frac{c^2}{A^2}-1}\right]=\frac{\mu_1}{\mu}\frac{\sqrt{1-\frac{c^2}{a}}}{\sqrt{\frac{c^2}{A^2}-1}}$$

This is the classical dispersion equation of Love wave by Love (1911); hence, there is validation of the solution of the problem discussed, which can be compared with the study of Ghorai et al. (2010).

Results and discussion

It is quite clear from the above dispersion equations that there exist two different wave fronts in different situations. The two wave fronts are graphically represented in Figs. 2 and 3; the data used are given in Tables 1 and 2.

In Table 1, there is comparison between two sets of generated curves (viz. curves 1 and 2 and curves 3 and 4) in the presence and absence of gravity with constant values of sandy parameter, rigidity and density ratios. In Table 2, there is a set of four curves for different values of A^2/c_3^2 with respect to a constant value of kH. The two generated waves are dispersive in nature as kH increases in Fig. 2 and Rk increases in Fig. 3. The figures are representing that the velocity of the Love waves depends on the change in volume of void pores. In Fig. 2, it is observed that the phase velocity of Love wave in the absence of gravity is greater than that in the presence of gravity, regardless whether the medium is elastic or sandy. The authors also reveal that the phase velocity of Love wave in a sandy medium is greater than that in the elastic medium in the presence or absence of gravity. In Fig. 3, the phase velocity of Love wave decreases as A^2/c_3^2 increases for a fixed value of kH. The second wave front depends on the void parameter of the layer only.

Practically, the real earth consists of void and also sandy media inside it, in the form of crust or mantle, and so it has an effect on the propagation of the surface or body waves. Not only the medium but also the gravity has a huge impact on the propagation of various surface and body waves during earthquakes. From the study, the authors revealed the fact that the existence of such kind of layers inside the earth implies that there exist two wave fronts in the real

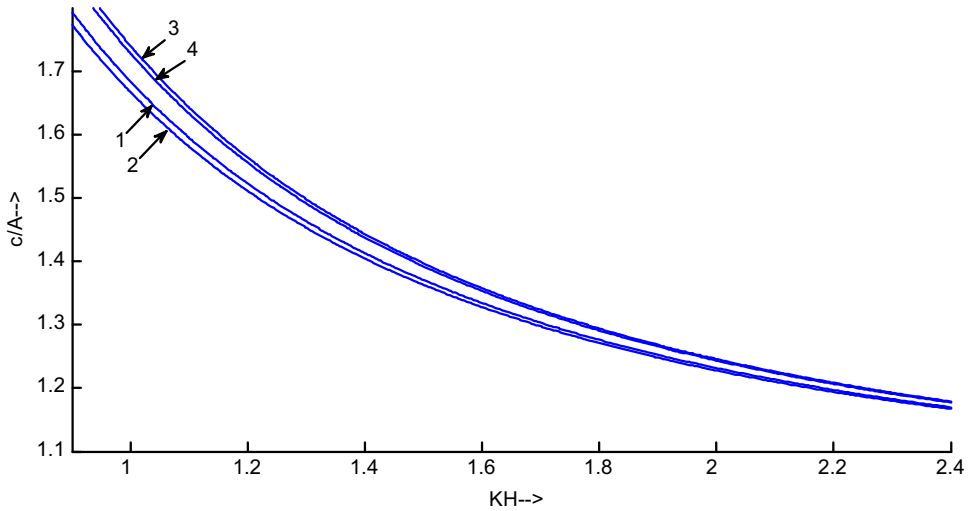


Fig. 2 Phase velocity of the Love wave in void medium lying over a sandy half space in the presence and absence of gravity

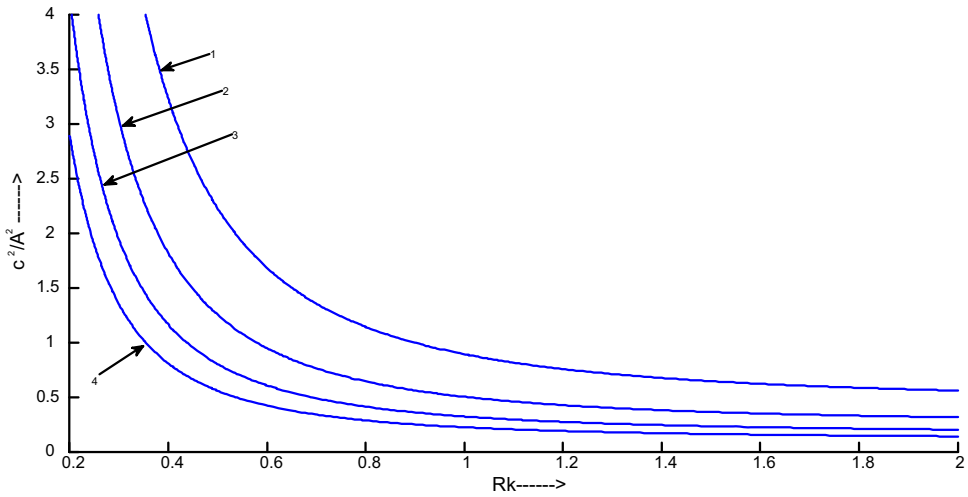


Fig. 3 Love wave dispersion curve in void medium over a gravitating sandy half space for different values of $\frac{A^2}{c_3^2}$

Table 1 Comparison of the wave velocities in the presence and absence of gravity with constant sandy parameter, rigidity and density ratios

Curve 1	Curve 2	Curve 3	Curve 4
$\eta = 1$	$\eta = 1$	$\eta = 2$	$\eta = 2$
$\mu : \mu_1 = 0.3$	$\mu : \mu_1 = 0.3$	$\mu : \mu_1 = 0.3$	$\mu : \mu_1 = 0.3$
$\rho : \rho_1 = 0.6$	$\rho : \rho_1 = 0.6$	$\rho : \rho_1 = 0.6$	$\rho : \rho_1 = 0.6$
$G = 0$	$G = 0.4$	$G = 0$	$G = 0.4$

Table 2 Comparison of the wave velocities for different values of A^2/c_3^2 with constant value of wave number

Curve 1	Curve 2	Curve 3	Curve 4
$kH = 20$	$kH = 20$	$kH = 20$	$kH = 20$
$\frac{A^2}{c_3^2} = 1.5$	$\frac{A^2}{c_3^2} = 2$	$\frac{A^2}{c_3^2} = 2.5$	$\frac{A^2}{c_3^2} = 3$

earth. As it is a very well-known fact that inside the real earth, as depth increases, the rigidity and density increase too. So for the present study, the authors are considering the rigidity and density ratio as of the real earth data, respectively, namely $\mu : \mu_1 = 0.3$ and $\rho : \rho_1 = 0.6$.

Conclusion

Propagation of Love waves in a void medium lying over a dry sandy half space under gravity has been investigated analytically. Thus, the dispersion equations were obtained, and it was observed that there is a significant effect of sandy, gravity and void parameters simultaneously on the propagation of Love waves. The study concludes with the facts that:

1. There exit two wave fronts in such an assumed medium.

2. One wave solely depends on gravity and sandy parameters.
3. The other wave front depends on the void parameter.
4. The phase velocity of Love wave in a sandy medium is greater than that in the elastic medium in the presence or absence of gravity.
5. One wave front decreases as kH increases, and the other wave front decreases as Rk increases.

Acknowledgements The authors are sincerely thankful to the SDET-Brainware Group of Institutions, Kolkata, Asansol Engineering College, Asansol and Indian School of Mines, Dhanbad, for their all kind of support.

References

Achenbach JD (1973) Wave propagation in elastic solids. North Holland, New York

Anderson DL (1962) Surface wave dispersion in layered anisotropic media. Ph.D. Thesis, California Institute of Technology, Pasadena

Biot MA (1955) Theory of elasticity and consolidation for a porous anisotropic solid. J Appl Phys 26(2):182–185. doi:10.1063/1.1721956

Biot MA (1962) Mechanics of deformation and acoustic propagation in porous media. J Appl Phys 33(4):1482–1498. doi:10.1063/1.1728759

Biot MA (1965) Mechanics of incremental deformations. Wiley, New York

Chapman CH (2004) Fundamentals of seismic wave propagation. Cambridge University Press, Cambridge

Cowin SC (1985) The viscoelastic behavior of linear elastic materials with voids. J Elast 15(2):185–191. doi:10.1007/BF00041992

Cowin SC, Nunziato JW (1983) Linear elastic materials with voids. J Elast 13(2):125–147. doi:10.1007/BF00041230

Dey S, Chandra A (1983) Surface waves in a dry sandy medium under gravity. Acta Geophys Pol 31(4):395–404

Dey S, Gupta S (1987) Longitudinal and shear waves in an elastic medium with void pores. Proc Indian Natl Sci Acad A 53(4):554–563

Dey S, Gupta S, Gupta AK (2004) Propagation of Love waves in an elastic layer with void pores. Sadhana 29(4):355–363. doi:10.1007/BF02703687

Ewing WM, Jardetzky WS, Press F (1957) Elastic waves in layered media. McGraw-Hill Book Co, New York

Ghorai AP, Samal SK, Mahanti NC (2010) Love waves in a fluid-saturated porous layer under a rigid boundary and lying over an elastic half-space under gravity. Appl Math Model 34(7):1873–1883. doi:10.1016/j.apm.2009.10.004

Kundu S, Manna S, Gupta S (2014) Love wave dispersion in pre-stressed homogenous medium over a porous half-space with irregular boundary surface. Int J Solids Struct 51(21–22):3689–3697. doi:10.1016/j.ijsolstr.2014.07.002

Love AEH (1911) Some problems of geodynamics. Cambridge University Press, London

Acta Geophys. (2017) 65:275–286
DOI 10.1007/s11600-017-0026-9

RESEARCH ARTICLE

On the possibility of producing definitive magnetic observatory data within less than one year

Igor Mandić[1,2] · **Monika Korte**[2]

Received: 1 March 2017 / Accepted: 3 March 2017 / Published online: 20 March 2017

Abstract Geomagnetic observatory data are fundamental in geomagnetic field studies and are widely used in other applications. Often they are combined with satellite and ground survey data. Unfortunately, the observatory definitive data are only available with a time lag ranging from several months up to more than a year. The reason for this lag is the annual production of the final calibration values, i.e. baselines that are used to correct preliminary data from continuously recording magnetometers. In this paper, we will show that the preparation of definitive geomagnetic data is possible within a calendar year and presents an original method for prompt and automatic estimation of the observatory baselines. The new baselines, obtained in a mostly automatic manner, are compared with the baselines reported on INTERMAGNET DVDs for the 2009–2011 period. The high quality of the baselines obtained by the proposed method indicates its suitability for data processing in fully automatic observatories when automated absolute instruments will be deployed at remote sites.

Keywords Geomagnetic observations · Definitive data · Quasi-definitive data · Baseline fitting · Data processing

✉ Igor Mandić
mandici@gfz.hr

Monika Korte
monika@gfz-potsdam.de

[1] Department of Geophysics, Faculty of Science, University of Zagreb, Zagreb, Croatia

[2] Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany

Introduction

Magnetic observatories are designed to perform continuous and accurate measurements of the strength and direction of the Earth's magnetic field over many years, or even centuries. Nowadays, the time resolution of these measurements is one minute or less. The observatory data are essential to understand how the field is changing on a wide range of scales from seconds to centuries, and to get an insight into dynamic processes inside and outside the Earth that generate the observed geomagnetic field. Ground observatories serve as base stations and are an important complement for various types of geomagnetic surveys (e.g. Mandea and Korte 2011). They provide data from a different observation altitude and pure time series in contrast to the data obtained for example from airborne or marine surveys and satellite missions, containing both temporal and spatial variations. Observatory data help to better constrain survey data and fill the gap between the present and future satellite missions.

Continuously recording vector magnetometers do not provide absolute field component values, but might be influenced by temperature variations, instable pillars or instrumental drifts. Manual absolute observations are still mandatory at modern magnetic observatories to calibrate the continuous recordings and are usually done once a week. These observations are used in conjunction with the data from the vector magnetometer (variometer) and an independent scalar magnetometer to derive observed base values. These values later serve for calculation of adopted baselines which are used to obtain absolute variometer data free from drift or environmental variations. Adopted values are derived by fitting, interpolation or hand scaling from the observed base values. Today, final baseline fitting is performed on an annual basis and data calibrated with these

final baselines are labelled as definitive (D) data. The final D data should be also free from spikes, jumps and other degradations in continuous recordings. For details of preparation of the final observatory data, see Jankowski and Sucksdorff (1996).

Since the establishment of INTERMAGNET (International Real-time Magnetic Observatory Network, http://www.intermagnet.org) in 1991, INTERMAGNET Magnetic Observatories (IMOs) reported three types of magnetic data until recently: D data (described above), reported (R) data, and adjusted (A) data. R data are raw data of the geomagnetic field variation measured with the variometer and should be available within 3 days from acquisition. These data are only suitable for studies of the rapid field variations (and may contain spikes, gaps and discontinuities). Within 7 days from transmission, observatories are allowed to modify R data to produce A data by removing spikes, filling gaps or to modify baselines. However, A data do not have sufficient absolute accuracy and are not free from drift. The D data required in particular for studies of secular variation and the geodynamo process in the Earth's core are reported only after the end of a calendar year with a time lag from several months up to more than a year. Unlike observatory data, fully calibrated data from satellites are available within a few days from acquisition.

Based on previous satellite missions (CHAMP, Ørsted, SAC-C), several studies demonstrated (Lesur et al. 2006; Macmillan and Olsen 2013; Olsen et al. 2006) the importance of the absolute observatory data for the ongoing Swarm mission, a constellation of three satellites (Friis-Christensen et al. 2006). The data from the Swarm satellites enable separation of the internal and external sources better than ever before. Processing, selection and validation of satellite data are usually based on data from ground observatories and their products, like geomagnetic activity indices. Absolute hourly mean values from world-wide observatories also play an important role in hour-by-hour spherical harmonic analysis used for deriving sophisticated magnetospheric models.

To increase the usefulness of observatory data and to fulfil the needs of the community for global modelling and other user groups, INTERMAGNET has defined the new type of "quasi-definitive" data (Clarke et al. 2013; Peltier and Chulliat 2010; Matzka 2013). Quasi-definitive (QD) data are corrected with temporary baselines, should be close to expected definitive values and are delivered much sooner (1–3 months) than the final D data. The variation part of QD data should have the same (or similar) quality as D data, i.e. without spikes or noise. QD data bridge the gap between preliminary and D data and pave the way to more efficient combination of ground observatory data and satellite, marine and airborne survey data.

Temporary quasi-definitive baselines (QDBLs) are estimated from all absolute measurement results performed in the current year. Some observatories report their QD data in near-real time using baselines obtained by extrapolation. In these cases, the annual baseline amplitude should not exceed 5 nT and this also requires prompt processing of the variation data by observatory staff. Nowadays, de-spiked and magnetically clean variation data and other observatory products like geomagnetic activity indices can be obtained relatively promptly, within several days from recording. Of course, this depends on the available observatory staff, amount and complexity of the encountered problems in the data and efficiency of processing protocols. The same is true for the absolute control of data, only on longer time scales. Human intervention is essential for the absolute control of observatory data. This includes manual absolute observations, retyping and processing of the observational results, quality checks and validation of results, decisions during baseline calculation, etc., both for temporary QDBLs and DBLs. For the production of QD and D data, those timescales correspond to several months and more than a year, respectively.

Today, most observatories use parametric (e.g. polynomial) or semi-parametric (e.g. smoothing splines) fits in the global domain, i.e. on annual basis to derive DBLs. Obtaining baselines in such manner has some disadvantages: (1) A full year of observed values has to be collected before calculating DBLs, (2) In the case of parametric fits, the convergence of temporary QDBLs toward the DBL is slow, i.e. the overlap between cumulative QDBLs and the DBL might not be good. The same is true for semi-parametric fits in the case when the distribution of observed values is non-uniform or the smoothing parameter of the spline functions is changed as the observational dataset is updated. This means that we need to have good baseline stability (i.e. high quality observatory infrastructure) to produce high-quality QD data. In general, global fits are good for estimating general trends in the observational dataset. On the other hand, if we have rapid trending of samples during a short period of time (for example, due to temporary problems with the mechanical stability of the variometer sensor), global fits may not be the appropriate tool to fit these short-term variations without affecting the baseline quality in other parts of the domain. Increasing the polynomial degree or flexibility of smoothing splines, to track samples when the baseline varies more rapidly, can cause "overfitting" or artificial variations in other parts of the domain, particularly in parts where observations are sparse (an example is given in Sect. "Special cases: data gaps and disagreeing intensity in scalar baselines").

In this paper, we first show that D data can be produced within the timespan less than a year. We show that it is enough to collect a reasonable number of observed values

during some period that enable us to do an accurate local fit and to determine DBL for this period. With absolute sampling (usually) once per week, we can collect enough samples within a few months to calculate one piece of DBL within a year. By constructing piecewise continuous fits, with smooth transitions on the edges of these pieces (windows), we are effectively constructing DBL by pieces. The advantage of the piecewise fitting is that observations in the future will not affect the shape of a baseline in the past, which is clearly justified from a physical point of view. Observed base values reflect the instrument response in the current environmental conditions and they are in fact independent from those in the future. We only need a reasonable number of good quality observations for some period of a time to obtain correct estimation of the baseline trend for days without observations. Correct approximation of samples in the periods when the baseline varies more rapidly and avoiding overfitting when we have good baseline stability are further advantages of the piecewise fitting.

We propose a new method for producing DBLs within a year. The developed routine can determine the "best fit" as candidate for QDBL and DBL without human intervention, although we emphasize that QDBL or DBL should be always reviewed by experienced observatory staff. We use the published INTERMAGNET observatory data from 2009 to 2011 to show that the proposed method gives high-quality results in a nearly automated manner. It thus can find its applicability in continuous baseline fitting when automated absolute instruments (Gonsette et al. 2013; Korte et al. 2013) become supplement or standard observatory equipment.

Methods

Proposed method for automatic production of temporary baselines

Motivated by the work and results obtained by Peltier and Chulliat (2010), we decided to design a routine based on cubic smoothing splines (De Boor 1978). They tested their method for producing QDBLs on nine IPGP (Institut de Physique du Globe de Paris) observatories (AAE, BOX, CLF, KOU, LZH, MBO, PHU, PPT, TAM) for the year 2008 and showed that the production of accurate QD data is possible even in the cases where baselines vary more rapidly through a year. The differences between definitive and temporary baselines, calculated after the end of each month, were within a fraction of nT. However, the obtained results might be too promising and unrealistic. The authors determined an optimal smoothing parameter for the dataset on an annual basis. Then, they simulated the production of QDBLs in retroactive manner with a priori known optimal smoothing parameters (see Peltier and Chulliat 2010, for details). In reality, in most cases the optimal smoothing parameter cannot be known in advance and in many cases simple smoothing splines will not give satisfactory results. This will be demonstrated in the next subsection.

To obtain high-quality baselines in an automatic manner, we use cubic smoothing splines in conjunction with additional constrains. First, we calculate a baseline by pieces (windows). We empirically found a length of eight observational days ($L = 8$) to be a good choice in most cases (one observational day can have several observation sets). Depending on the observation frequency, the time span of a window may vary from a week up to several months depending on the observatory. Additionally, samples from two ($d = 2$) observation days before and after the window are used for calculation of a baseline within the window. This alleviates edge effects between neighbouring baseline segments. The baseline within a window is a cubic smoothing spline function $f(t)$ that minimizes the expression

$$p\sum_{i=1}^{n} w_i(y_i - f_i)^2 \quad + \quad (1-p)\int_a^b \{f''(t)\}^2 dt, \tag{1}$$

where $f_i = f(t_i)$ are values of the baseline $f(t)$ at the observation times t_i, $f''(t)$ stands for the second derivative. In the integral expression, a and b are the limits of the extended window containing $L + 2d$ observations. Observations are denoted with y_i and w_i are corresponding weights. The smoothing parameter $p \in [0, 1]$ controls the smoothness of $f(t)$. For higher values of p, the baseline will be closer to the observations. In the case $p = 0$, the baseline will be a straight line and for $p = 1$ the interpolating baseline will pass exactly through the data points. To obtain weights first, we estimate a preliminary baseline $\tilde{f}(t)$ where all observations have weights equal to 1. Then, using the residuals $y_i - \tilde{f}_i$, weights are estimated according to the most frequent value (Steiner 1988). This way relatively flexible baseline segments (high values of p) are robust to bigger outliers or low-quality observations. Determination of the optimal smoothing parameter is performed automatically according to relation (De Boor 2003)

$$p = 1/\left(1 + h^3/c\right), \tag{2}$$

where h is the average sampling, i.e. average period between observation days. The current version of the program assumes the same coefficient c for three vector components while the scalar c may be the same or different. Coefficient c is varied in the range from 10^{-3} up to 10^2 and for each c (vector and scalar) baselines are calculated.

To determine the suitable value for c and check the quality of the determined baselines, the so-called "Delta F" values are calculated. Delta F presents the difference

between the total field intensity obtained from vector and independent scalar magnetometers. Here, the scalar total field is corrected for the field difference between the sensor site and the absolute pier (for details see St-Louis 2012). The best c corresponds to the minimum value of Delta F. For this purpose, the routine requires the variometer and scalar daily mean values in addition to the observations itself. In case no scalar magnetometer data are available, it is not possible to perform the Delta F check and the default value of $c = 0.5$ is used.

In some cases of IMOs, the annual long-term baseline drift is comparable or order of magnitude larger than random fluctuations of samples. On the other hand, some observatories have very frequent absolute observations (several sets, almost every day) in some time intervals. In this case, observations can show random fluctuations, which are not expected in the nature of the underlying baseline. Note also that with decrease of h (frequent observations), p increases. According to Eq. (1), this means that the baseline will try to follow the same random fluctuations by remaining close to the observations. Thus, in general we prefer "smoothness" over "staying close to the data". By choosing the default $c < 1$, we restrain the spline function to be smooth enough to capture the trend and not to follow random scattering or erroneous observations.

Once we have obtained piecewise baselines according to Eq. (1), with several constrains as described above, the resulting baseline pieces will have similar (but not equal) values on the edges to neighbouring windows. Thus, a final smoothing by a spline function with the default $c = 0.005$ for vector and $c = 0.0005$ for scalar is applied to piecewise baselines to obtain smooth transition on the edges of the windows for which the baselines are calculated. In this way, we obtain smooth, continuous baselines over the whole domain. Optionally, the final vector and scalar c can be set manually by the user. Except in the last window, the shape of a baseline obtained in this manner will not be affected as we add new observations.

Realignment or relocation of a magnetometer sensor (or some other reasons) may result in discontinuities, i.e. jumps in the variometer or scalar baseline. Generally, the number of discontinuities may be different for the vector and scalar instrument. These discontinuities are usually known at the observatories and are marked with "d" in the INTERMAGNET baseline files. The marker "d" and originally reported baseline values are used to identify jumps and their magnitudes. The routine requires a reference list with times and values of discontinuities. In our test cases, this list is created from the information provided in the INTERMAGNET baseline files. Observations are then corrected to a single reference level to obtain accurate estimations of the new adopted values.

Special cases: data gaps and disagreeing intensity in scalar baselines

The BOX example from Peltier and Chulliat (2010) was chosen to demonstrate how non-uniformity of sampling and assumption about an a priori "known" smoothing parameter can introduce artificial baseline variation that is not supported by samples. In Fig. 1, only the horizontal (H) and vertical (Z) baseline components are displayed, together with Delta F. Original baselines and Delta F (reported on the INTERMAGNET DVD) are obtained taking into account all observations. Now, let us imagine that the observations were not possible during spring due to some operational problems (e.g. the observatory location was unreachable for some reason or problems with the declination–inclination magnetometer occurred). Results obtained under the assumption that some observations in spring are missing with the IPGP and the newly proposed method are presented by blue and black lines, respectively. Vertical grey lines designate the automatically defined windows in the proposed method (Sect. "Proposed method for automatic production of temporary baselines "). The IPGP and original baselines overlap perfectly everywhere except in the third window. Clearly, we can see that the IPGP method gives a more erroneous estimation of the baseline than the new method, if observations in spring are omitted. The initially chosen smoothing parameter in the

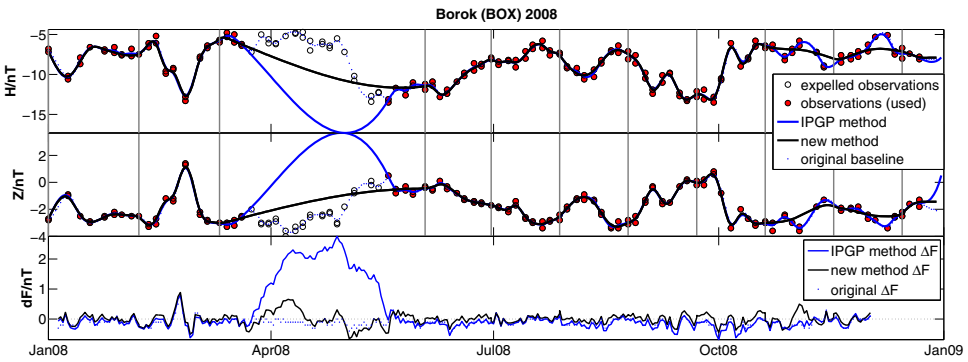


Fig. 1 Baselines calculated with the IPGP and new method (default settings) under the assumption that observations from spring (*white circles*) are missing. Original baselines and Delta F (*blue dotted line*) were obtained by taking all observations (*red and white circles*)

IPGP method will not be a good choice anymore, since the smoothing splines in the BOX case are quite flexible and we do not have any samples in spring to restrain the spline wiggling. The new baselines give more reasonable values due to several constraints that are described in Sect. "Proposed method for automatic production of temporary baselines ".

On the other hand, Fig. 1 shows somewhat smoother baselines obtained with the new method during November and December. Too smooth values of the new baselines are caused by missing information about the total field difference between the scalar sensor and absolute (i.e. reference) site. Prior to 2009, observatories did not report this information, named as "scalar baseline". Also, there is no information about Delta *F* in December, i.e. data from the scalar magnetometer are missing. Thus, the proposed baselines in the last two windows were obtained without any information about the scalar baseline and Delta *F* in December.

In many cases, the scalar baseline has the biggest impact on the Delta *F* residual and sometimes the baseline obtained from scalar samples cannot properly reduce the total field recordings from scalar magnetometer to the reference site. Although scalar magnetometers in general are absolute instruments, a potentially time-varying baseline results come from the fact that vector, scalar and absolute observations are carried out at different sites. All measurements in the end are referred to the same reference site. The difference between values at the site of the scalar instrument and the reference site determines the scalar baseline. As demonstrated in Fig. 2 ("*S*" diagram) in the case of the reported data, we have an obvious degradation in Delta *F* in this example. (IMOs on Figs. 2, 3 and 4a, b are left as anonymous). The major reason is the presence of a constant offset (approximately −0.5 nT) in the reported scalar baseline.

We presume that the presence of the offset in scalar samples on Fig. 2 is a consequence of erroneous or outdated information about the field difference between the scalar and reference site in the protocol for processing absolute observations. Generally speaking, the representativeness (offset, variability and scattering) of the scalar samples also depends on the magnetic homogeneity of the observatory site. We need to keep in mind that recordings and observations in the observatory are performed at three different sites. The gradiometer difference and its variation over a time, between the reference and scalar site, should be as small as possible. The same stands for the difference between the variometer and scalar site. Also, scalar and vector recordings during absolute observations must be free from spikes and the differences between natural short-term variations at all observatory sites should be negligible. These are basic prerequisites for obtaining quality base values and a proper reduction of the variometer and scalar data to the reference site.

In situations when we have disagreeing intensity in the scalar baseline (i.e. offset in Delta *F*) or increased scattering of *S* samples, the new method estimates the *S* baseline as a de-trended value of Delta *F* obtained in the

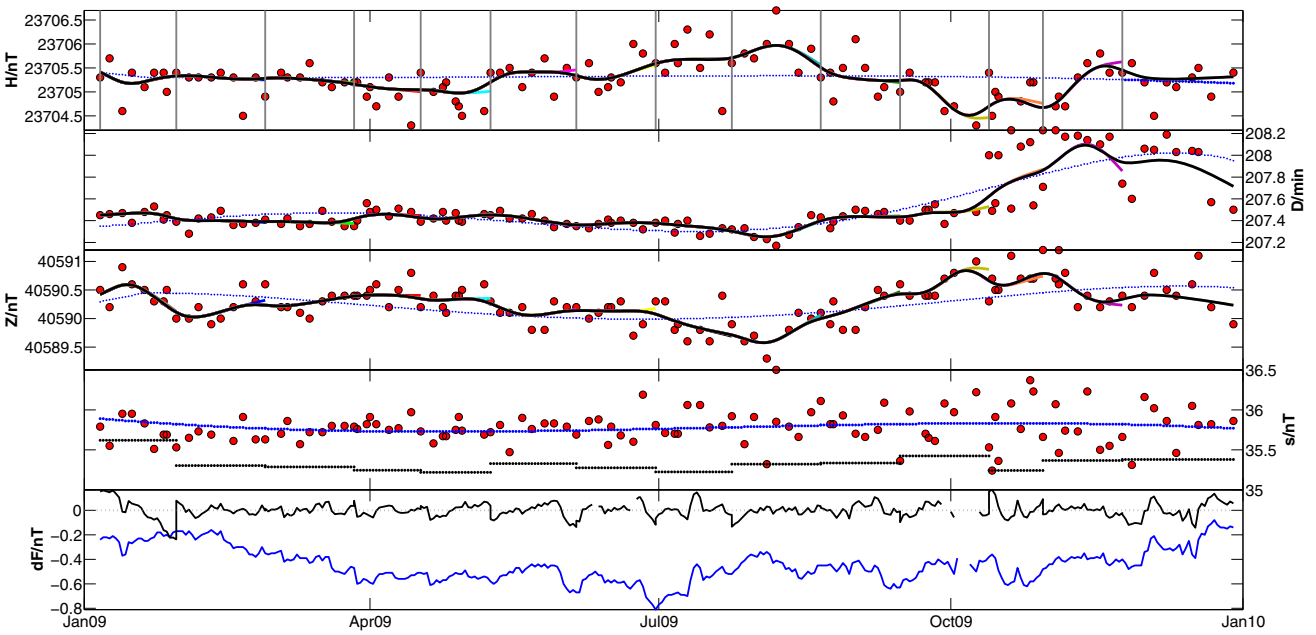


Fig. 2 Vector baselines obtained with the new method (*black line*) and original baselines (*blue dotted line*). Scalar baseline obtained with the new method is displayed by *black dotted lines* and original by *blue dotted line*. Observations are denoted with *red circles*. Original Delta *F* (*blue line*) corresponds to original baselines (*blue dotted lines*)

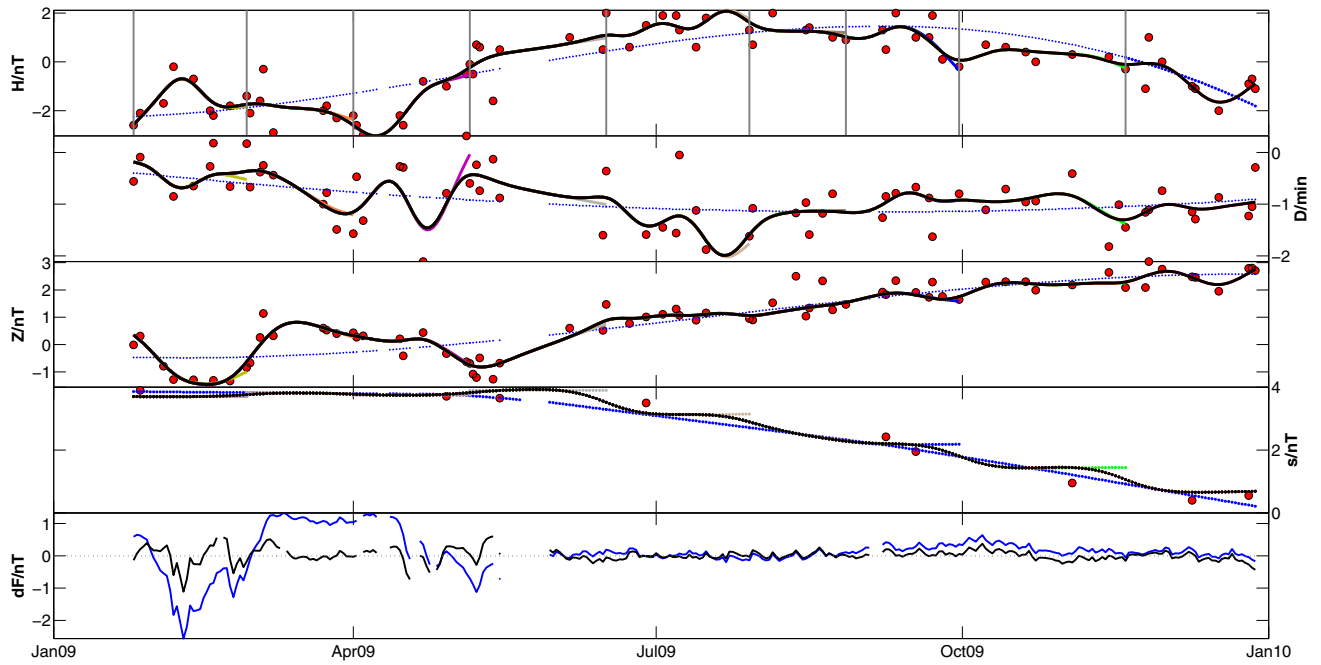


Fig. 3 Vector baselines obtained with the new method (*black line*) and original baselines (*blue dotted line*). Scalar baseline obtained with the new method (*black dotted line*) and original baseline (*blue dotted line*). The new temporary scalar baselines (within windows) are displayed by *coloured constant lines*. Observations are denoted with *red dots*. Original Delta *F* (*blue line*) corresponds to original baselines

iterative manner described in Sect. "Proposed method for automatic production of temporary baselines" under the initial assumption that *S* baseline is zero. The baseline obtained in this way is shown by black dotted lines in Fig. 2 ("*S*" diagram). Each piecewise constant segment of the annual baseline is estimated for each of the windows shown in the "*H*" diagram (Fig. 2, grey lines).

In case we do not have any information on the scalar baseline or less than 3 samples, necessary for spline calculation within the window, we assume that the scalar baseline is a constant. This assumption is also valid if Delta *F* calculated using the spline scalar baseline, obtained from scattered *S* samples, is bigger than Delta *F* obtained with a constant scalar baseline.

Furthermore, if we obtain discontinuities bigger than 0.5 nT between piecewise constant scalar baselines, then these constant baselines are smoothed by a final spline with default $c = 0.0005$ (see Sect. "Proposed method for automatic production of temporary baselines") to avoid sudden jumps in *F* from the independent scalar instrument. An example of a small jump (<0.5 nT) can be seen in Fig. 2 in Delta *F* and *S* (on the edge of the first and second window). In Fig. 3 ("*S*" diagram), we can see preliminary piecewise constant baselines (coloured dotted lines). The final adopted baseline obtained by smoothing the constant baseline segments is represented by a black dotted line.

Looking at the fit of the vector baselines in Figs. 2 and 3, we see that new baselines capture the trend of the samples much better than the original baselines. This is especially pronounced in the *Z* component in Fig. 3. In the first half of the year, the originally adopted values could not completely compensate a drift in the *Z* baseline. As a consequence, the Delta *F* residuals are increased in this period. Note that in this period, the original and new values of the scalar baseline are practically identical. In Fig. 2, we see a long-term drift through the whole year in addition to a constant offset in the original Delta *F*. The reported baselines in this case also underestimate the baseline trends indicated by samples in *H* and *Z*.

Manual adjustment and quasi-definitive baselines

A single computational method may not be appropriate to handle all different dataset types and in some cases might not give satisfactory results. This can occur especially in cases where we have a relatively small number or different numbers of samples in each component, non-uniform sampling, sudden discontinuities and when every dataset has its own characteristics (like trends, amplitudes and sampling quality). All these are potential features of observational datasets obtained from standard measurements in geomagnetic observatories. Testing our new routine for all IMOs showed that in some cases improved results are obtained with manual adjustments of the automated procedure following visual inspection of the data and calculated baselines.

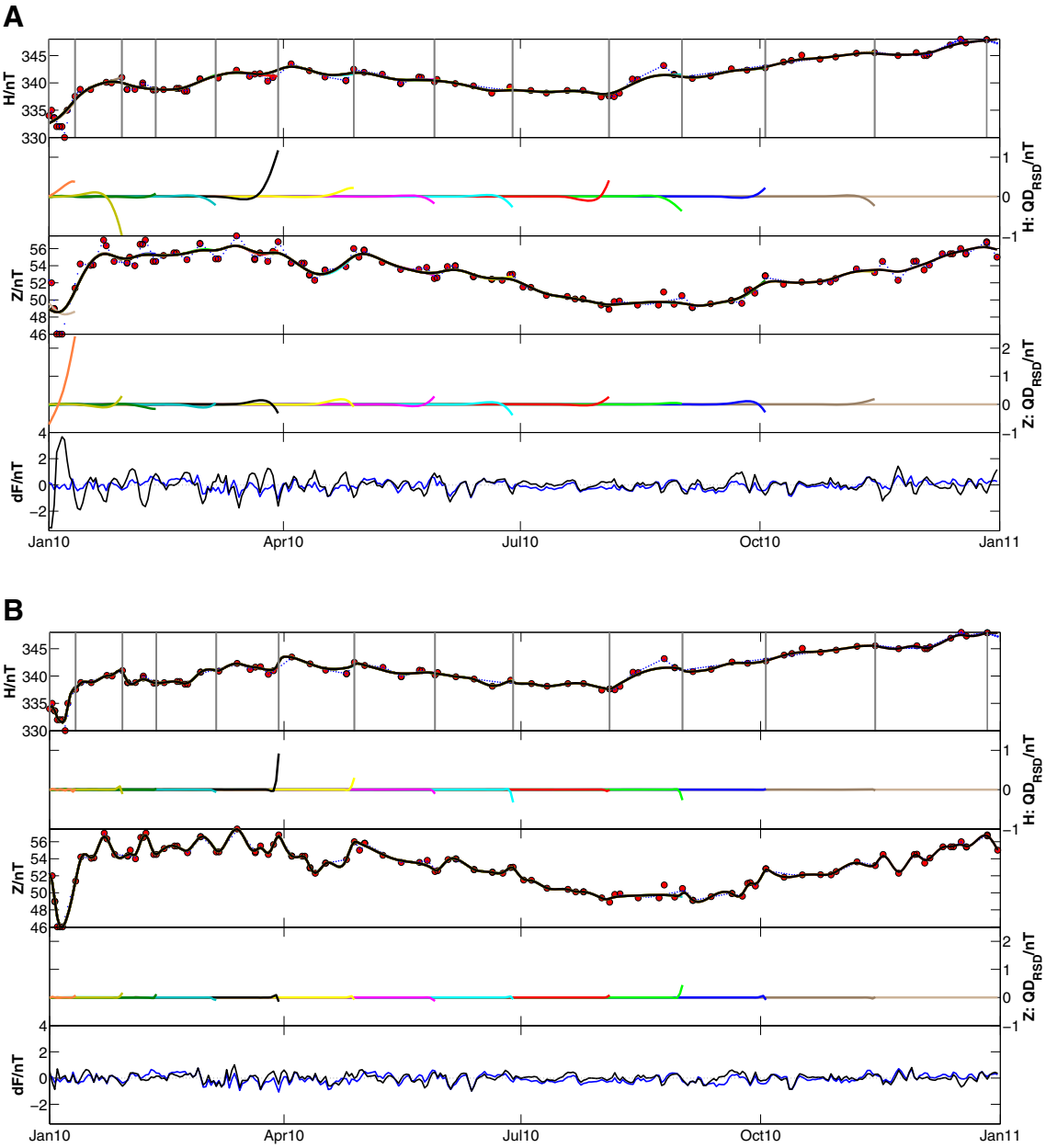


Fig. 4 **a** *H* and *Z* baselines obtained with the new method (*black lines*) and original baselines (*blue dotted lines*). Residuals between QDBLs and DBLs are shown with *coloured lines*. Observations are denoted with *red dots*. Original Delta *F* (*blue line*) corresponds to original baselines. **b** Same as **a** only the new *H* and *Z* baselines are obtained with manual adjustment. Results are presented with the same ordinate scaling as used in **a**

We manually varied *L* and *d* parameters (see Sect. "Proposed method for automatic production of temporary baselines") or the smoothing parameter of the final spline which is used to join sub-splines within windows. This procedure is demonstrated in Fig. 4a, b. Results obtained with default settings are presented in Fig. 4a. In this case, the new baselines are not flexible enough to fit rapid baseline variability at the start of the year. This clearly leads to degradation in Delta *F*.

If we manually increase coefficient *c* in relation (2) of the final spline that smooths the baselines within individual windows, a more flexible baseline will be obtained. This is illustrated in Fig. 4b. With this minor intervention (*L* and *d* were not changed), significantly better results were obtained.

In Figs. 2, 3 and 4a, b, the new temporary QDBLs (coloured lines) are simultaneously plotted under DBL. However, due to practically perfect overlapping between cumulative QDBLs and the final baseline, only small segments of QDBLs are visible in Figs. 2 and 3. A fraction of QDBLs, barely noticeable, can be seen on the right

edges of some windows. For example in Fig. 2 near the edge of eleventh window part of the vector, QDBLs (orange line) are noticeable, which are slightly different from final DBLs calculated from all observations in the year. Similarly, in Fig. 3 this can be seen in declination (D) near the edge of the third window. The residuals between QDBL and DBL (QD_{RSD}) are plotted separately in Fig. 4a, b for better visibility. Here, we see that small differences between QDBL and DBL exist only near the right edge of a window. If we denote the number of windows with N, then we have perfect overlapping QDBLs with DBL within all windows from 1 to $N-1$, i.e. QD_{RSD} are exactly zero. This means that DBLs are calculated within all windows, from first up to $N-1$. Only near the right edge of N-th window, small differences exist and till we define the $N+1$ window the baseline is labelled as QD. Generally, for all observatories used in this study these small differences are within fractions of an nT.

Verification of the proposed method

Data selection and preparation

To compare our baseline calibration results with the ones reported by IMOs, the data in the period 2009–2011 were used. This period was chosen because calibration data from the independent scalar magnetometer were not available before 2009. Also during the preparation of this paper, the last available INTERMAGNET DVD was published for 2011. In this period, we have used only data from observatories that report Delta F results. This way it is possible to estimate the adoption quality according to the Delta F check. The overall number of observatories that reported Delta F in this 3 year period is 255.

Standardly, the baseline adoption for a year (YY) includes observations from December of YY−1 and January YY + 1. Since baseline results are reported in standardized yearly IBF V2.0 files (http://www.intermagnet.org/data-donnee/formats/ibfv200-eng.php), new baselines are calculated using only observations within YY. Then, the original and new baseline results are compared for a time period between the first and last observational day of year.

Statistical analysis of results

In Sect. "Methods", we have presented only a few examples of the 255 analysed cases. To show all results in summarized form, some statistics are shown in the following.

To obtain an idea about the goodness of fit with the new and reported DBLs, we have simply used squared correlation coefficients (R^2) between observed and adopted values. Table 1 shows results of examples presented in Figs. 2, 3, and 4a, b. From the numerical values in Table 1 and from the plotted results, we can see that R^2s are quite good indicators of the goodness of fit.

Table 1 Squared correlation coefficients (R^2) between observed and adopted values presented in Figs. 2, 3, and 4a, b

Case/comp.	DBL type	R_1^2	R_2^2	R_3^2	R_4^2
Figure 2	Reported DBL	0.036	0.643	0.261	0.001
	New DBL	0.450	0.724	0.653	0.003
Figure 3	Reported DBL	0.674	0.198	0.787	0.978
	New DBL	0.869	0.526	0.943	0.972
Figure 4a, b	Reported DBL	0.990	–	0.990	–
	New DBL	0.949	–	0.899	–
	New manual DBL	0.986	–	0.990	–

Four coefficients (R_1^2,… R_4^2) correspond to three vector and one scalar component. R_2^2 and R_4^2 are not given for Fig. 4a, b because examples of these plots are not shown there

Then, the maximal absolute residuals (max|RSD|) and the average absolute residuals (avg|RSD|) between the new and reported baselines were determined on yearly basis. This was done for each of the 255 cases, for vector and scalar baselines and Delta F residuals. We have rejected 5 cases where max|RSD| and avg|RSD| were outside two standard deviations from the average results. In all rejected cases, the new baselines gave better results in Delta F and according to squared correlation coefficients between observed and adopted values as a measure of the fit. The remaining residuals are sorted into 0.5 nT bins and presented in the form of histograms.

A similar analysis was done for the new, cumulative temporary baselines (i.e. QDBLs) within a year. Here, the residuals (QD_{RSD}) were determined only with respect to the new DBL. In the QD case, only the maximal absolute residuals within windows are determined ($\max|RSD^i_{QD}|$) where $i = 1, \ldots, N$ and N is the number of windows. Then, the maximal ($\max(\max|RSD^i_{QD}|)$) and average ($\mathrm{avg}(\max|RSD^i_{QD}|)$) values from all windows within a year were determined. By considering only $\max|RSD^i_{QD}|$, we have excluded periods of perfect overlapping among the new QDBL and DBL (Fig. 4a, b, residual diagrams). Otherwise, we would obtain unreasonable small residuals (less than 0.01 nT), i.e. much smaller than the achievable absolute accuracy of the baseline calibration procedure. Residuals between the new QDBLs and reported DBLs were not calculated separately. These residuals correspond approximately to residuals between the new DBLs and reported DBLs. In other words, $\mathrm{avg}|RSD^i_{QD}|$ are much smaller than max|RSD| and avg|RSD|.

Results and discussion

In Fig. 5 (upper panels), the maximum and average absolute Delta F values are presented. New baselines, automatically calculated, with default settings are displayed with black histograms. Results of originally reported Delta F are presented with white histograms. We clearly see that new baselines in general give slightly better results according to the Delta F check. However, here we must emphasize that in some individual cases human intervention is crucial to improve the baseline fit. Minor manual adjustments were done in 46 cases for this purpose, based on visual inspection. After re-calculation of baselines, small improvements can be seen in Delta F (Fig. 5, grey histograms).

Histograms of differences between the new DBLs, obtained in automatic manner, and reported DBLs show that max|RSD| are within 1 nT and avg|RSD| are within 0.2 nT in approximately 80% of the cases. The biggest residuals are obtained in the Y (or D) component and the smallest are in the S component. These results are expected because usually the baseline variability is biggest in Y (or D) while the S baseline is the most stable. Stability of the S baseline is also evident from Fig. 6 where around 30% of S baselines are approximated by constant or piecewise constant fits, i.e. R^2s are $\leq$0.1. For the vector components, we clearly find a higher number of cases with R^2 between 0.5 and 0.9 obtained with the new baselines than with the originally reported baselines.

A somewhat higher number of observatories with R^2 around 1 in the case of originally reported baselines is mostly due to observatories that determine their DBLs by linear interpolation or piecewise linear fits. In this case, human intervention is essential. Here, all outliers and low-quality observations are excluded from the observational datasets. Also, the data processor makes decisions during piecewise fitting, for example which observations should be used to calculate one piece of a baseline, where to join the baseline pieces, which may be defined differently for different components, etc. These cases also show that to obtain high-quality data in some cases the intervention from a skilful and experience data processor is essential.

The same test was performed with the new baselines that are calculated using default c parameters for the splines within windows. This means fitting was not performed iteratively to minimize Delta F, i.e. to find an optimal c. The overall average differences between baselines with default c and automatically determined c are quite small. For all components, more than 90% of average/maximal residuals are <0.1 nT/1 nT while 0.3 nT/1 nT are rarely excited. However, in some cases significant improvements can be seen in the fit quality for baselines that are determined by minimization of Delta F.

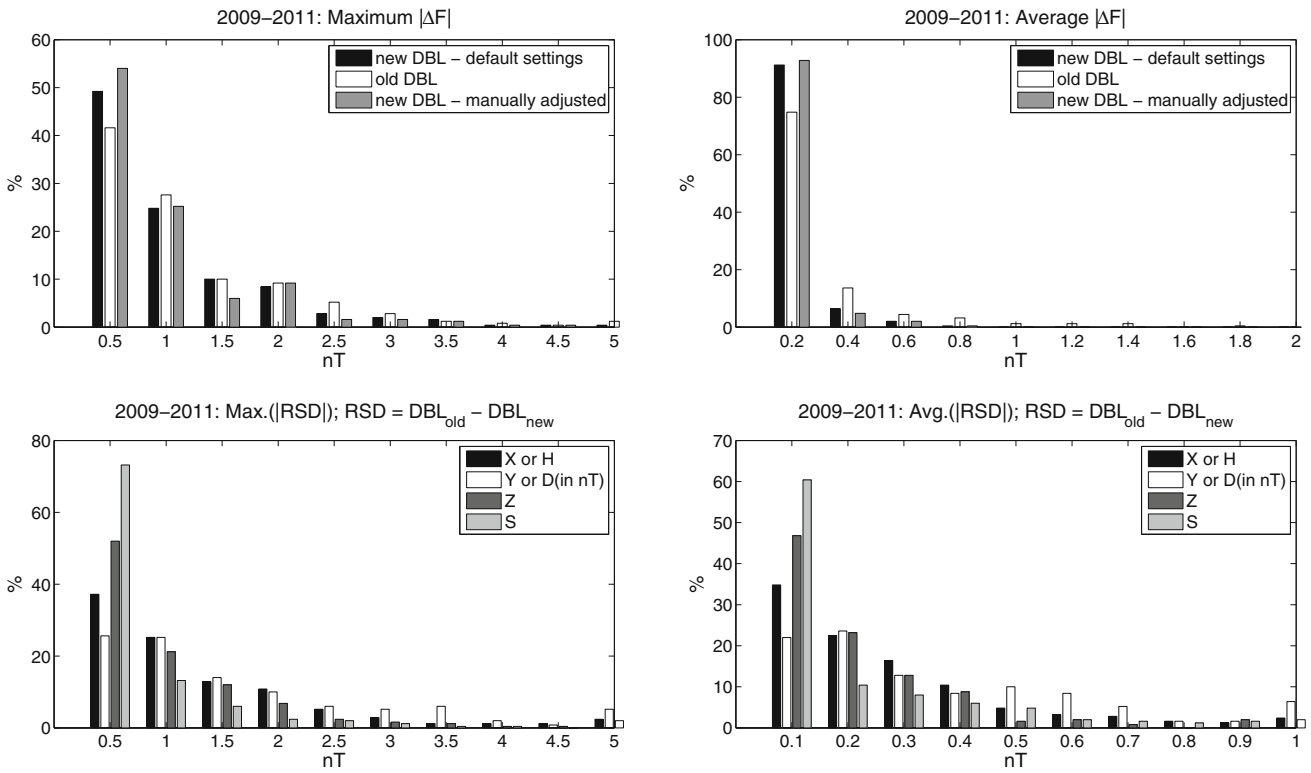


Fig. 5 *Upper panels* histograms of the absolute maximum and average absolute Delta F. *Lower panels* residuals between the new DBL (without manual adjustment) and reported DBL for three vector and the scalar component

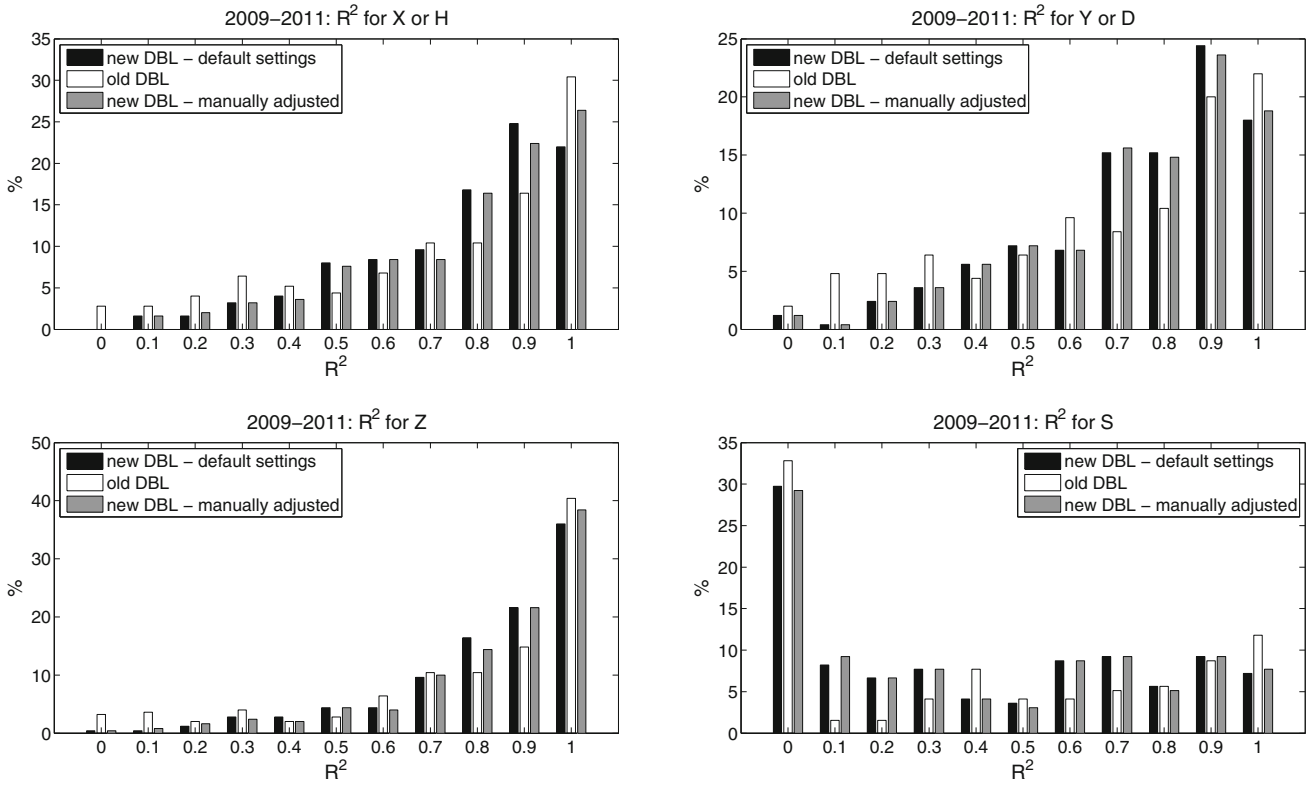


Fig. 6 Squared correlation coefficients (R^2) between adopted and observed values for three vector and the scalar component. Three cases are presented. R^2 obtained for newly automatically calculated baselines (*black*), reported baselines (*white*) and newly calculated after manual adjustment for some observatories (*grey*)

The new method thus has several advantages compared to traditional ones. First, the new baselines are indeed more accurate, since in most cases they track the observations better than the original baselines and consequently give a smaller Delta *F* (Figs. 2, 3). Secondly, residuals between the new DBL and QDBLs are practically negligible (Fig. 7). Small differences exist only near the right edge of the last window for which baseline is calculated. In average, more than 90% of these residuals are <0.3 nT (Fig. 7right) while yearly maximum residuals in all 250 cases rarely exceed 1 nT (Fig. 7left). Note that this discrepancy occurs only for a small time period, as demonstrated in Fig. 4a, b. Immediately after calculation of a new baseline using samples from the consecutive window (i.e. baseline update), this discrepancy is corrected. Last, but most important, the new DBLs are determined within a year. The time delay of DBL production may vary from several weeks up to several months and depends only on the frequency of absolute observations.

Our new temporary baselines also clearly fulfil the requirements to define them as QDBLs with respect to the reported DBLs. According to INTERMAGNET standards, the (average) differences between QDBLs and DBLs should be less than 5 nT nd for many observatories; these differences are within 1 nT. From Fig. 5, we can see that average avg|RSD| are within 1 nT in all components and that about 80% are within 0.2 nT. Here, we made the assumption that new temporary baselines have perfect overlapping with the new DBLs. This assumption is valid because differences between the new DBLs and its temporary segments are much smaller than differences between the new and original DBLs. Note that Fig. 5 shows maximum and average residuals while Fig. 7 shows annual maxima and averages of the maximal residual within all windows.

Conclusions

Traditionally definitive geomagnetic observatory data are published annually with a delay from several months up to more than a year. The reasons are either that it may be difficult to produce high-quality recordings (visually checked, de-spiked, corrected for jumps, etc.) soon after acquisition due to the lack of observatory staff and processing protocols or simply because the calculation of definitive baselines and calibration of recorded data on yearly basis is currently standard procedure.

In this study, we have investigated the possibility to produce definitive observatory data more continuously with

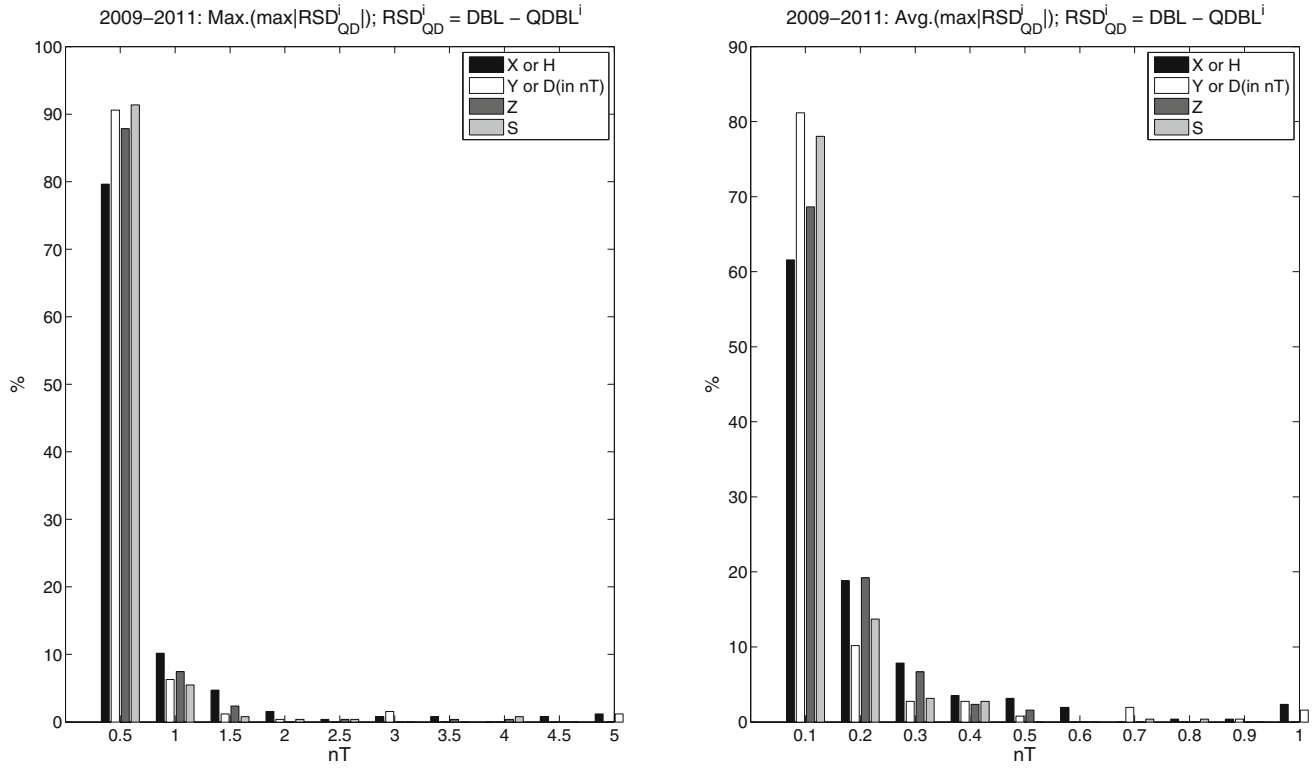


Fig. 7 *Left* annual maxima of the maximal residuals between the new DBL and its temporary baseline segments (QDBL), i.e. each maximal residual correspond to one baseline window. *Right* annual averages of the same residuals presented on the left subplot

smaller time lags. We have proposed a new method for reliable baseline adoption in a nearly fully automatic manner. Definitive baselines can be estimated by pieces from a sufficient number of observations in a way that the future results will not affect the shape of baselines from the past. This way it is also much easier to fit flexible baselines in case of rapid variation of the absolute observations within some period without overfitting surrounding observations that are stable. By re-processing the published baseline observations of the INTERMAGNET observatories from 2009 to 2011, we have shown that DBLs calculated with the new method in most cases give the same or better results (as indicated by Delta F and R^2) than DBLs determined by experienced observatory staff. However, in some cases manual intervention is necessary to obtain good results with the new method. We plan to develop a graphical user interface (GUI) that will allow others to use the method with options for manual intervention to improve results. For example, before fitting the user will be able to remove big outliers, define knots (i.e. window edges) differently in each component and modify the automatically determined smoothing parameter within the window and for each component separately. (For availability of GUI, please contact the corresponding author).

Most results in this study are obtained relying on the Delta F check. Continuous and accurate scalar observations are essential to obtain correct scalar F at the reference site to obtain reliable Delta F. Due to increased scattering or absolute inaccuracy in many cases, the proposed method estimates scalar baselines as a piecewise constant functions that do not fit observations. Examples similar to one shown in Fig. 2 (results obtained with the proposed method) can be also found on INTERMAGNET DVDs 2009–2011. Thus, the best observatory practice would be to set scalar baseline to zero, i.e. to perform continuous scalar measurement at the reference site (except during manual DI-flux observations). Alternatively, the gradiometer difference between the scalar and reference site, which is required for processing of the DI-flux observations and obtaining accurate scalar observations, should be regularly updated after each absolute DI-flux observation.

We have demonstrated that most of INTERMAGNET observatories should be able to distribute definitive data relatively soon after acquiring recent absolute measurements. Availability of the best available data, i.e. D data, within less than a year will be highly appreciated from the side of the observatory data users. With weekly absolute observations, most observatories should be able to produce their definitive data more continuously with a delay between 2 and 3 months. Increasing the number of observations, or shortening the window size, could significantly shorten this delay. In addition, with the proposed technique

maximal discrepancies between D and QD data (which can be available within a few days/weeks) rarely exceed 0.3 nT.

The fact that in many cases the new method gives good results without human intervention paves the way towards automating the whole data processing chain, and towards fully automated observatories when automated absolute instruments become supplement equipment or the only absolute instrument at remote (unmanned) observatories (Korte et al. 2009; Mandić et al. 2016). These instruments will provide information about the absolute values of the field and variometer base values much more frequently than today. The increase in the number of automatically collected observations (let say every 30 min, Gonsette et al. 2013) will require a routine for baseline fitting and calibration of variometer data continuously in near-real time. However, the automatic instruments show a larger scatter and we have to expect increased number of outliers. Thus, an automated processing protocol should be robust to outliers. In future work, we will show that our new method produces robust and reliable results with data from an automated absolute instrument producing many absolute observations per day, so that final definitive data could be available within a few days from acquisition.

Acknowledgements We would like to thank two anonymous reviewers for constructive suggestions that helped us to improve the quality of the paper. The results presented in this paper rely on the data collected at magnetic observatories. We thank the national institutes that support them and INTERMAGNET for promoting high standards of magnetic observatory practice (http://www.intermagnet.org). This work was supported by the DAAD (No. 91529536-50015537) and ERASMUS (No. ERA-2013-166) exchange programme.

References

Clarke E, Baillie O, Reay SJ, Turbitt CW (2013) A method for the near real-time production of quasi-definitive magnetic observatory data. Earth Planets Space 65(11):1363–1374. doi:10.5047/eps.2013.10.001

De Boor C (1978) A practical guide to splines. Springer, New York

De Boor C (2003) Spline toolbox for use with MATLAB, The MathWorks, Inc

Friis-Christensen E, Lühr H, Hulot G (2006) Swarm: a constellation to study the Earth's magnetic field. Earth Planets Space 58(4):351–358. doi:10.1186/BF03351933

Gonsette A, Rasson J, Marin JL (2013) AUTODIF: automatic absolute DI measurements. In: Proceedings 15th IAGA Workshop on Geomagnetic Observatory Instruments, Data Acquisition, and Processing, Boletin ROA, No. 03/13, 16–19

Jankowski J, Sucksdorff C (1996) Guide for magnetic measurements and observatory practice. International Association of Geomagnetism and Aeronomy, Boulder

Korte M, Mandea M, Linthe HJ, Hemshorn A, Kotzé P, Ricaldi E (2009) New geomagnetic field observations in the South Atlantic Anomaly region. Ann Geophys Italy 52(1):65–81

Korte M, Brunke HP, Bronkalla HP, Pulz E (2013), Status of the geomagnetic automated system GAUSS. In: Proceedings 15th IAGA Workshop on Geomagnetic Observatory Instruments, Data Acquisition, and Processing, Boletin ROA, No. 03/13, 20–23

Lesur V, Macmillan S, Thomson AWP (2006) Deriving main field and secular variation models from synthetic Swarm satellite and observatory data. Earth Planets Space 58(4):409–416. doi:10.1186/BF03351937

Macmillan S, Olsen N (2013) Observatory data and the Swarm mission. Earth Planets Space 65(15):1355–1362. doi:10.5047/eps.2013.07.011

Mandea M, Korte M (2011) Geomagnetic Observations and Models, IAGA Special Sopron Book Series, Springer, Dordrecht, DOI: 10.1007/978-90-481-9858-0

Mandić I, Vujić E, Heilig B, Pelajić I, Herak D (2016) Recent efforts toward the establishment of the Lonjsko Polje geomagnetic observatory. Acta Geophys 64(5):1311–1339. doi:10.1515/acgeo-2016-0051

Matzka J (2013) Preparation of quasi definitive (QD) data for the observatories Narsarsuaq, Qeqertarsuaq and Tristan Da Cunha. In: Proceedings 15th IAGA Workshop on Geomagnetic Observatory Instruments, Data Acquisition, and Processing, Boletin ROA, No. 03/13, 50–53

Olsen N, Haagmans R, Sabaka TJ, Kuvshinov A, Maus S, Purucker ME, Rother M, Lesur V, Mandea M (2006) The swarm end-to-end mission simulator study: a demonstration of separating the various contributions to Earth's magnetic field using synthetic data. Earth Planets Space 58(4):359–370. doi:10.1186/BF03351934

Peltier A, Chulliat A (2010) On the feasibility of promptly producing quasi-definitive magnetic observatory data. Earth Planets Space 62(2):e5–e8. doi:10.5047/eps.2010.02.002

Steiner F (1988) Most frequent value procedures. Geophys Trans 34(2–3):139–260

St-Louis B (2012) Intermagnet technical reference manual (Version 4.6)

Acta Geophys. (2017) 65:287–298
DOI 10.1007/s11600-017-0034-9

RESEARCH ARTICLE

Identification of seismic doublets occurred on Rudna mine, Poland

Łukasz Rudziński[1] · Janusz Mirek[1] · Grzegorz Lizurek[1]

Received: 13 February 2017 / Accepted: 27 March 2017 / Published online: 6 April 2017

Abstract Seismic events with similar focal mechanisms and similarity of Green's functions exhibit common waveform shapes recorded on the same seismic station. The similarity of seismograms can also be observed in the case of man-induced seismicity on the continuous excavated area. In this work, we focused on strong and potential mining-induced seismic doublets, our study was using concepts of signal analysis, using an in-mine network with quite low dynamic range. Since in such case records for strong events are usually affected by clipping, signal analysis were performed with both raw signal cross-correlation (CC) and binary signal cross-correlation (BCC). We found that for events which fulfilled the established criteria, waveform similarities are significantly higher than for other possible doublets. Using BCC and CC approaches, we noticed crucial influences of double couple nodal planes' as well as P and T axis orientations on waveform similarities although the focal mechanisms were characterized by very high non-double couple components. Finally analysis confirmed that human activity represented by mining is able to produce strong ($M > 3$) repeating mining tremors.

✉ Łukasz Rudziński
rudzin@igf.edu.pl

Janusz Mirek
jmirek@igf.edu.pl

Grzegorz Lizurek
lizurek@igf.edu.pl

[1] Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland

Keywords Mining induced seismology · Seismic doublets · Seismic signal

Introduction

The seismic signal recorded on a seismological network contains a lot of information needed to understand source physics. Of greatest importance are seismic phases used to solve for hypocenter location, focal mechanism or velocity model. Even a simple view of seismograms can provide some advanced analysis. One of the most interesting signal based studies are those dealing with repeated earthquakes with common shapes of recordings and clustered on particular areas. Doublets or more wider multiplets of events are often observed on active tectonic zones (e.g., Kagan and Jackson 1999), low-seismicity and volcanic areas (e.g., Gregersen et al. 2007; Quintanar et al. 2004) or human induced seismicity (e.g., Gibowicz 2006; Wehling-Benatelli et al. 2013) and provide a possibility to improve seismological analysis. Some conditions have to be fulfilled for multiplets in a sense of waveform similarities (Geller and Mueller 1980): two events should have almost the same location, with separation less than one quarter of the shortest wavelength ($\lambda/4$—criterion) also both source mechanisms and source time functions (STF) should be very similar. Such conditions might have a direct physical interpretation and the most popular is that similar earthquakes occurred on the same part of fault by re-rupturing of the same asperity at a different time. This simple concept provides a lot of possibilities to improve seismological studies, from the new events' location procedures up to new conceptual source models (e.g., Poupinet et al. 1984; Got et al. 1994; Waldhauser and Ellsworth 2000; Anooshehpoor and Brune 2001; Schaff and Waldhauser

2005; Baisch et al. 2008; Rudziński and Dębski 2011; Spottiswoode and Milev 1998; Godano et al. 2015).

In contrast to natural earthquakes, the mining induced events never occur without human activity which produces fast changes of stress distribution within rocks close to excavation panels (Gibowicz and Lasocki 2001). Even though dynamic response of the rock to industrial activity can change the velocity model, typical pattern of similar earthquakes might be observed in different induced environments (Evans et al. 2005). Due to clustering of the induced seismicity, waveform similarities might be treated as a good indicator of processes responsible for rock failure. It is well known that mining-induced seismicity tends to cluster in space and time (Gibowicz and Lasocki 2001). Clustering can be also observed when one is dealing with moment tensors (MT) for a number of close-spaced induced tremors in hard coal mining (Cesca et al. 2013; Sen et al. 2013). Analysis of similar tremors (Wehling-Benatelli et al. 2013) also showed that it might be a useful tool to indicate different types of failures caused by coal seam exploitation.

Multiplets and doublets of seismic events are also not exceptional in Polish copper mines. Based on catalogues studies, Gibowicz (2006) developed essential criteria characterizing multiplets in this particular area. Hypocenters of two events are separated by no more than 200 m, the difference in origin time is not longer than 20 days, and the difference in moment magnitude is not larger than 0.15. The study found that in the Rudna mine—one of the analyzed copper mines, 11 doublets, 4 triplets and 1 quadruplet were recorded in 2004 on different mining panels (Gibowicz 2006). In 2010, Rudna mine started excavation on the new mining panel no. XX/1 (Fig. 1). The panel was not affected by previous work. During excavation eight strong induced events with $M > 3$ occurred and gave a possibility to provide an analysis of possible doublets among strong mining tremors. Directly following Gibowicz's criteria, we found and studied possible seismic multiplets using waveforms analysis of records from in-mine network. Even though the network is dense enough to locate sources with epicentral errors less than 100 m, the low dynamic range influences seismograms by clipping. This is a factor which must be considered during signal analysis. In this paper, we try to deal with this problem and finally we summarize results obtained in this case study. Our investigation redefines the criteria for multiplets for clustered events occurring in Polish copper mines.

Site and data description

The Rudna mine is a deep copper mine belongs to the Legnica Głogów Copper District located on southwest Poland. Seismicity in the mine is quite strong with more

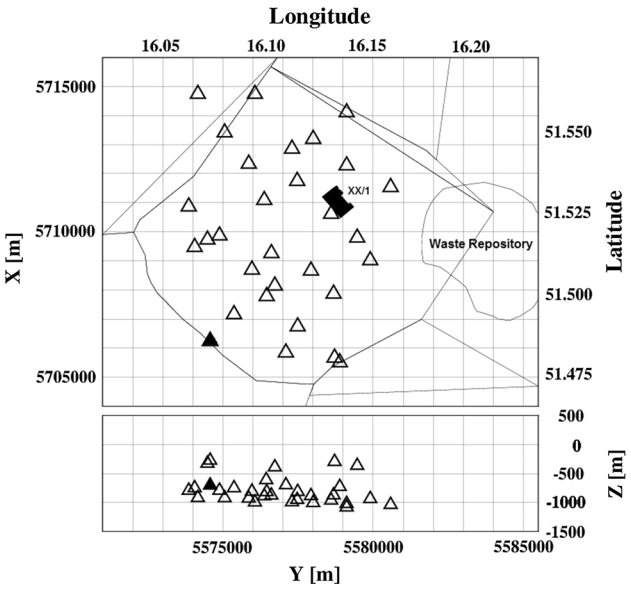


Fig. 1 The map of the Rudna mine seismic network (*triangles*) and the mining panel XX/1. Waveforms overlaying is shown for *black* fulfilled station (more details in *text*)

than 1000 events with $M > 1.5$ recorded every year. An ore level is located at the depth from 900 to 1200 m and a copper is, at the same time, excavated from more than five mining panels. The XX/1 was excavated between 2010 and 2012. The productive level of panel XX/1 of Rudna mine is located at the depth of about 1100 m and is overlain by some 60–92 m thick dolomite layer followed upwards by rigid anhydrite strata of around 160 m thickness. Below the deposit is a layer of hard Rotliegendes sandstones of about 300 m thickness. Ore is extracted by means of a room-and-pillar exploitation system. From 2010 to 2012, the XX/1 mining panel in the Rudna copper mine was struck with several strong seismic events (Orlecka-Sikora et al. 2014). Based on Gibowicz (2006) criteria, we decided to select eight events (Table 1; Fig. 2) for further analysis of possible doublets existence. Our work was facilitated by the mining catalogue obtained from records of the underground seismic network operated in the Rudna mine. The initial events location provided by mining geophysical surveys were corrected using the location procedure worked with probabilistic inverse approaches, the algorithm already tested in the Rudna mine (Rudziński and Dębski 2011). The final hypocentres location with errors for particular coordinates are included in Table 1. While epicentre coordinates are estimate with high precision, accuracy in depth is much lower. Anyway we believe that all events have a similar origin (mining induced) and occurred on the similar depth.

The seismicity in the Rudna mine is continuously monitored by a local seismic network (Fig. 1). The network belongs to the mine and consists of 32 short-period

Table 1 The list of strong seismic events selected as possible doublets

Event	X	ΔX	Y	ΔY	Z (m)	ΔZ	M_w	R
Ev1:**26.06.2010**	5711101	50	5578981	55	−930	240	3.5	206
Ev2:**13.11.2010**	5711112	50	5578965	50	−975	220	3.3	181
Ev3:**18.12.2010**	5711151	50	5578798	50	−827	235	3.4	167
Ev4:**20.01.2011**	5711071	55	5578987	50	−897	260	3.1	192
Ev5:**20.01.2011**	5711197	55	5578811	50	−827	235	3.5	206
Ev6:**08.06.2011**	5710981	45	5578844	45	−941	260	3.5	160
Ev7:**05.10.2011**	5710911	55	5578682	55	−949	290	3.3	188
Ev8:**11.03.2012**	5710866	60	5578701	55	−882	300	3.4	208

X and Y are local coordinates in Cartesian system, Z is a source depth in meters, ΔX, ΔY, ΔZ are location errors in meters. M_w is a moment magnitude while R is a foci radius

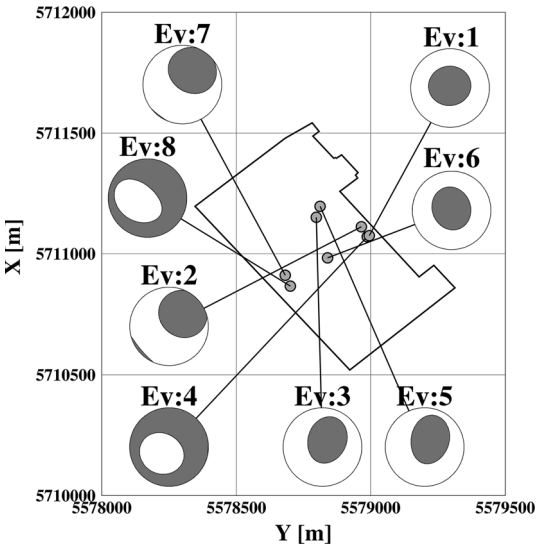


Fig. 2 The map of proposed seismic doublets and its full MT solutions

vertical seismometers Willmore MK-III. Even though the number of stations and 500 Hz sampling rate give an opportunity to locate hypocenters within an accuracy of about 50 m in epicentral distance and 150–200 m in depth (Rudziński and Dębski 2011), less than 70 dB dynamic ranges can cause clipping of seismic signals. This condition is especially obstructive when a very strong event is considered.

For selected events from Table 1, two source parameters, namely moment magnitude M_w and source radius, have been determined using spectra of P and S waves using the relationship between spectral and Brune's source parameters. All estimations were performed using the formalism of Andrews (1986), Brune (1970, 1971) and Wyss and Brune (1968), described by Gibowicz and Kijko (1994) and Niewiadomski (1997). According to those formalisms, moment magnitude is obtained with seismic moment which is directly related to the low frequency level of the far field displacement spectrum:

$$M_0 = \frac{4\pi\rho c^3 R\Omega_0}{F_c R_c S_c} \quad (1)$$

ρ—density of the source area, c—P or S wave velocity in source, R—source-receiver distance, Ω_0—spectral level, R_c—free surface correction, S_c—site correction, F_c—P or S wave coefficient.

Source radius r_0 was estimated using Madariaga's (1976) circular fault model; in such formalism (Brune 1970), the source radius is inversely proportional to the corner frequency f_c of P or S wave:

$$r_0 = \frac{K_c \beta_0}{2\pi f_c} \quad (2)$$

where K_c is a constant depending on source model and β_0 is S-wave velocity in the source region. The relevant parts of seismograms of P and S waves were selected manually, and then transformed by Fast Fourier Transformation (FFT). The resulting amplitude spectra were corrected for attenuation effects with $Q = 400$ and 200, respectively, for P and S waves. For further calculations we set velocities of P and S waves in source at $V_p = 5900$ m/s and $V_s = 3400$ m/s. The station number used for analysis differ from 15 to 25 in case of P-wave and from 4 to 10 stations in case of S-wave according to the unclipped recordings.

The source radii R of the studied events were between 160 and 208 m (Table 1), which corresponds to the corner frequency from 1.5 up to 3.5 Hz. The source size estimates are not exceptional when compared with the average event source radius of the events from Rudna mine, which was mainly about 200 m (Lizurek and Wiejacz 2011).

In next step for selected events a focal mechanisms were calculated using the full moment tensor (MT) inversion in time domain. Calculations of the full MT from the records of the underground network of the mine were performed with the use of FOCI software (Kwiatek et al. 2016). The input parameters are the amplitude and polarity information on the first P-wave displacement pulses. According to Fitch et al. (1980), De Natale et al. (1987) and Aki and

Richards (2002), the recorded displacement for the P-wave phase is:

$$\mathbf{U}^{\mathrm{P}}(\mathbf{x},t) = \frac{\boldsymbol{\gamma}\cdot\dot{\mathbf{M}}\left(t-\frac{r}{\alpha}\right)\cdot\boldsymbol{\gamma}}{4\pi\rho\alpha^3 r}\,\mathbf{l} \tag{3}$$

where ρ is the average medium density, r is the source-receiver distance, α is the average velocity of the P wave, $\mathbf{M}$ is the seismic moment tensor, $\mathbf{l}$ is the P wave direction at the receiver, and $\boldsymbol{\gamma}$ is the P-wave direction at the source.

The source time function (STF) was based on the Haskell's source model (Haskell 1953):

$$\dot{s} = \begin{matrix} 1/T, & 0<t<T \\ 0, & \text{elswhere} \end{matrix} \tag{4}$$

where T is the rupture time.

The MT is obtained by solution of a set of N equations of type (3). The deviatoric, pure shear and full MTs (Table 2) were calculated using the L2 norm as a measure of the misfit (Wiejacz 1992; Awad and Kwiatek 2005). The recorded first arrivals may be of different types: direct P waves or refracted waves from the overlying anhydrite layer or the thick strata of sandstone underlying the ore deposit. This is caused by the geological situation. The layers are almost flat, the dip of the layers being about 4° NE. The ore bearing strata are a dolomite layer of 60–90 m thickness. Above the ore bearing strata, there exists the evaporate series mainly composed of the anhydrite of 160 m thickness, and below the ore bearing strata there is a sandstone layer of about 300 m thickness.

Doublets analysis

In Fig. 2, all doublets are shown with its full MT solutions while Figs. 3, 4, 5 and 6 shows the corresponding fault plane solutions, i.e., the double-couple (DC) contribution (see Table 2). Based on the results of the moment tensor inversion, four doublets are proposed (Figs. 2, 3, 4, 5 and 6). Figures 3, 4, 5 and 6 show the following pairs, Ev2–Ev4, Ev3–Ev5 and Ev7–Ev8 exhibit significant similarities in nodal planes distributions and Ev1–Ev6 with similar MT solutions. The full MTs (Fig. 2; Table 2) for Ev1–Ev6, is characterized by almost the same decomposition and the smallest DC component among four events pairs. This observation is not an extraordinary feature, since the non-double-couple (non-DC) events are quite often observed in induced seismicity (Gibowicz 2009, Rudziński et al. 2016). Comparison between Figs. 6 and 2 show that although the MTs of events 1 and 6 are very similar, the corresponding DC solutions are very different.

Both the hypocenter location (inter-event distances) and other source parameters (Tables 1, 2 and 3) support our choice of possible doublets. Final inspections suggest the following four doublets:

- Ev1 and Ev6,
- Ev2 and Ev4,
- Ev3 and Ev5,
- Ev7 and Ev8.

These pairs fulfilled criteria developed by Gibowicz (2006).

Since the highest dominant frequency for P wave is 3.5 Hz and based on local average velocity model the P wave velocity is 5.5 km/s, the shortest distance satisfying the λ/4 criterion should be less than about 390 m. Note that all inter-event distances fit into these limits (Table 3). The last parameter characterizing seismic doublets in the Rudna mine is inter-event time. Since this parameter is very difficult to set up, in the previous studies has been selected arbitrarily as 20 days. In this work, we assumed that it is significantly less important than the distance criterion. Our selected doublets are characterized with following time differences: Ev2–Ev4: 28 days, Ev3–Ev5: 33 days, Ev7–Ev8: 127 days and Ev1–Ev6: 347 days. Even the shortest time distance exceeded 20 days criterion.

Table 2 The full MT solutions, scalar seismic moment and corresponding moment magnitude M_w for the selected seismic events

Event	Focal mechanism solution full moment tensor						
	ISO (%)	CLVD(%)	DC (%)	Nodal plane A (strike/dip/slip)	Nodal plane B (strike/dip/slip)	Scalar seismic moment (Nm)	M_w
Ev:1	28	55.3	16.7	84/50/90	265/40/91	$1.81 \cdot 10^{14}$	3.5
Ev:2	24	61.5	14.5	127/85/−80	246/11/−151	$9.2 \cdot 10^{13}$	3.2
Ev:3	30.9	32.3	36.8	185/53/68	39/43/116	$1.59 \cdot 10^{14}$	3.5
Ev:4	−30.4	−46.0	23.6	304/67/−82	105/24/−107	$5.50 \cdot 10^{13}$	3.1
Ev:5	26.6	29.4	43.9	180/56/68	36/40/119	$1.8 \cdot 10^{14}$	3.5
Ev:6	24	47.5	28.5	164/47/84	353/43/96	$2.13 \cdot 10^{14}$	3.5
Ev:7	7.7	54.8	37.5	124/87/93	259/4/45	$9.35 \cdot 10^{13}$	3.3
Ev:8	−18.4	−20.8	60.8	314/68/−77	104/26/−118	$1.30 \cdot 10^{14}$	3.4

Parts of the full MT: *ISO* isotropic, CLVD, DC

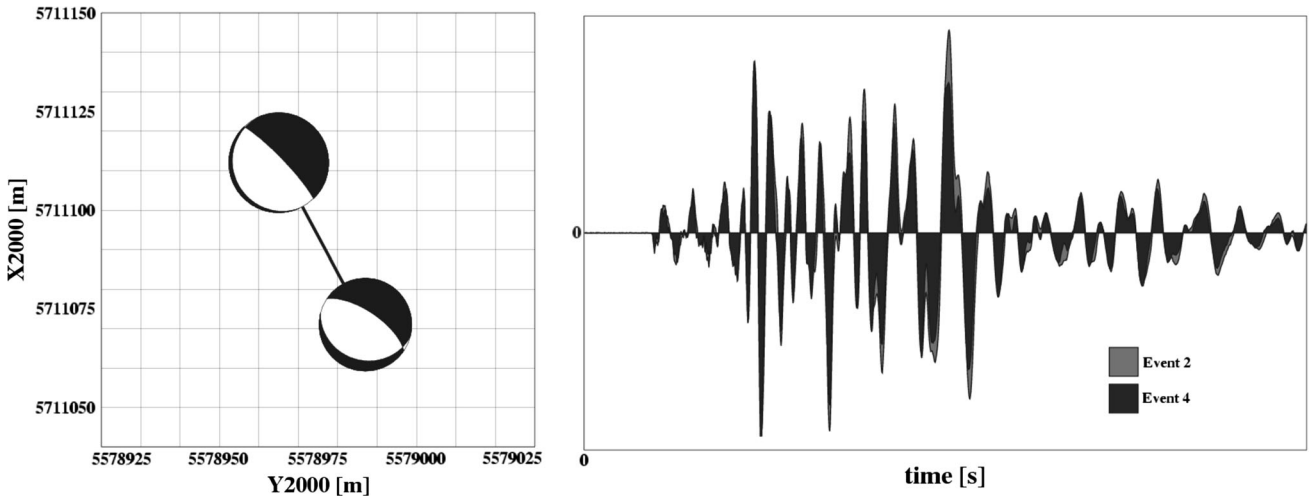


Fig. 3 Doublet Ev2–Ev4. DC solutions (*left*) and seismograms overlapping recorded on the same station

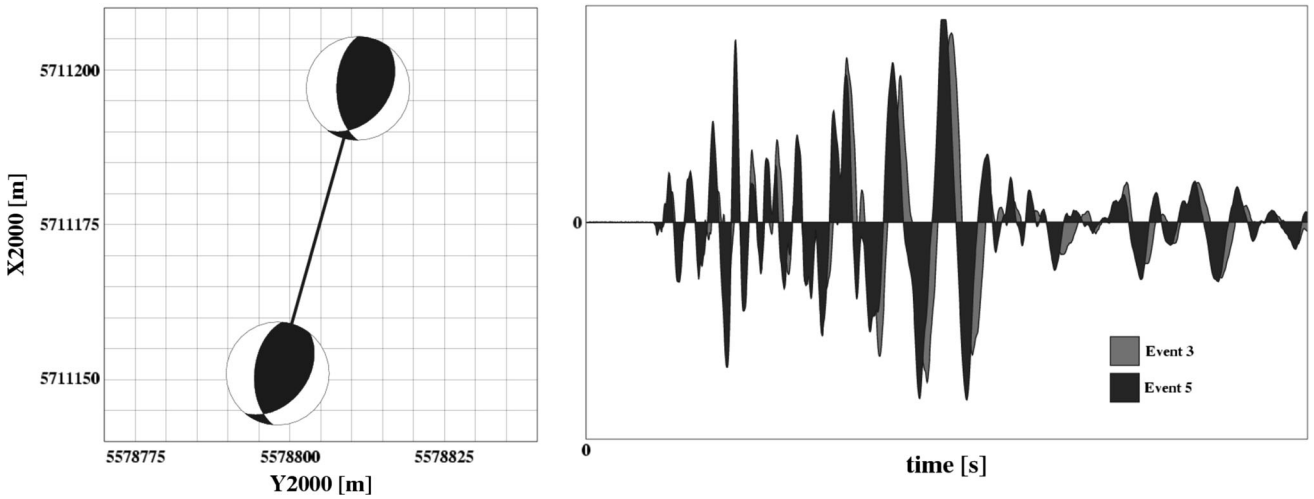


Fig. 4 Doublet Ev3–Ev5. DC solutions (*left*) and seismograms overlapping recorded on the same station

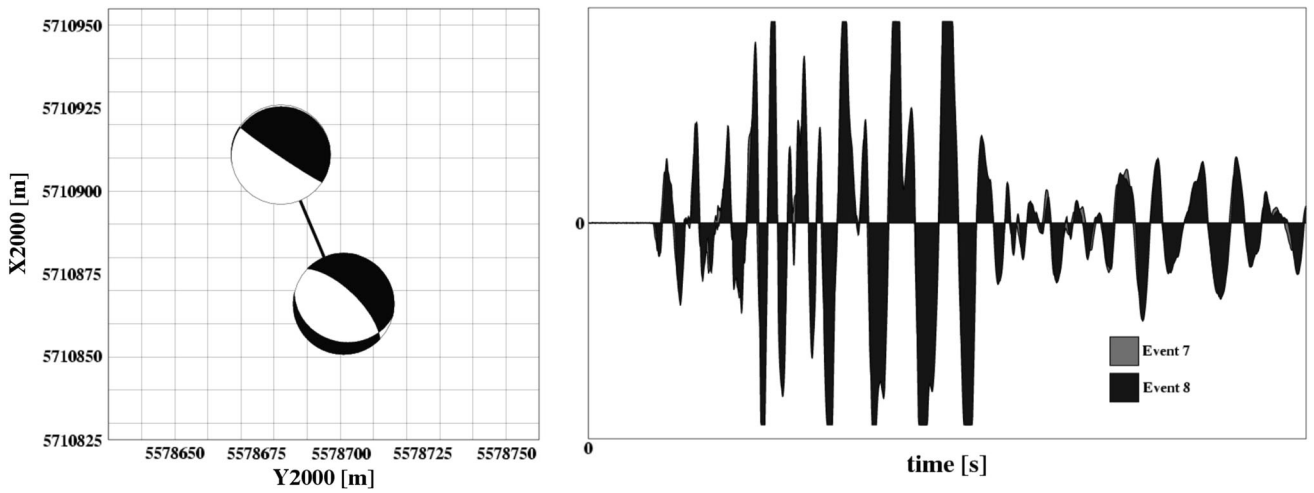


Fig. 5 Doublet Ev7–Ev8. DC solutions (*left*) and seismograms overlapping recorded on the same station

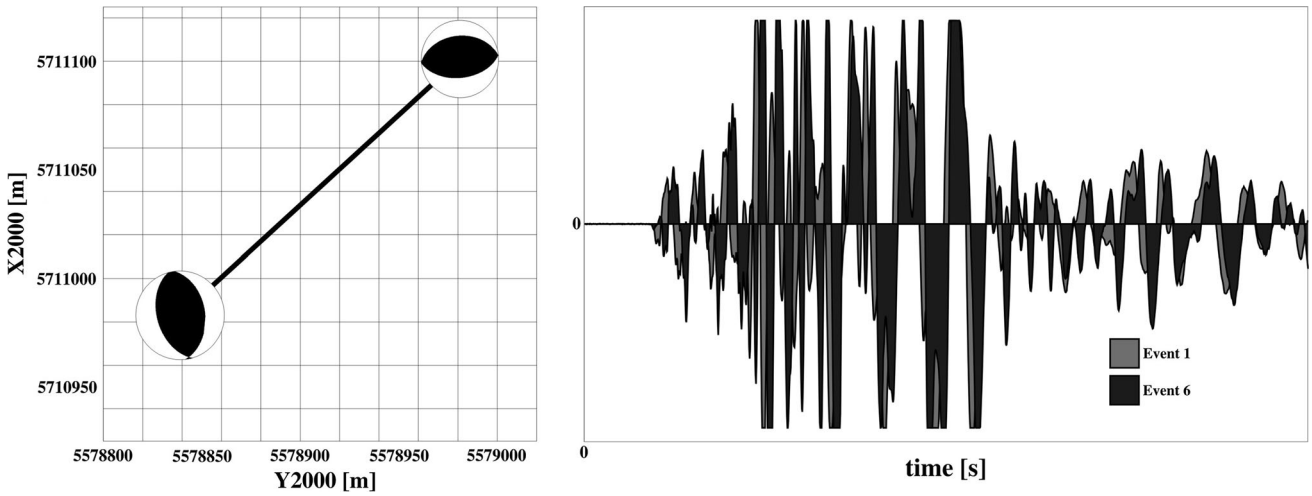


Fig. 6 Doublet Ev1–Ev6. DC solutions (*left*) and seismograms overlapping recorded on the same station

Table 3 Doublets selection based on distances between hypocenter locations ΔD and differences in moment magnitude ΔM_w

Doublet	ΔD	ΔM_w
Ev1–Ev6	193 m	0
Ev2–Ev4	80 m	0.1
Ev3–Ev5	48 m	0.1
Ev7–Ev8	83 m	0.1

Two groups of doublets can be now considered. The first one consists of three pairs: Ev2–Ev4, Ev3–Ev5 and Ev7–Ev8. The second one contains just Ev1–Ev6 doublet. For selected doublets, we propose the following seismic signals analysis. On the first step let us choose one the most adequate seismic station and roughly compare the seismograms obtained for two events from each pair. The selected sensor should be far enough from clusters centroid to avoid some recordings problems especially seismograms clipping in S-wave coda. To show how the waveforms overlap each other, we used a station located about 6 km away from the XX/1 mining panel (black triangle in Fig. 1). The results are gathered in Figs. 3, 4 and 5 (first group) and in Fig. 6 for the last pair.

The first group of events exhibits very interesting features in both focal mechanisms and waveform similarities. It should also be noted that the rough line connecting two events forming doublet, corresponds to directions of dominant discontinuities in the mining panel (Mining Geophysical Survey personal information 2015). This observation suggests that for this group the DC-part rupture on pre-existing discontinuities. The most important conclusion is that waveforms for doublets observed in this group are almost identical, even though tremors have, among pairs, other magnitude ranges.

Some other conclusions follow from Fig. 6. The pair Ev1–Ev6 in comparison with the previous doublets has different fault plane orientation and the waveforms not overlap each other. On the other hand, the complete MT results plotted on a lower hemisphere are identical between Ev1 and Ev6 (Fig. 2). Moreover, based on miners observations, the line between these two events does not correspond to any existing geological discontinuity. These two features are very interesting, when we take into account almost the same MT decomposition and suggest that the focal mechanisms for both events are rather connected with pillar destruction without contribution dealing with pre-existing lines of weakens, what is not exceptional at the Rudna mine (Lizurek and Wiejacz 2011). One can make a statement that Ev1–Ev6 should not be treated as a doublet.

It is worth to calculate the parameter η which measures the degree of fault rupture overlapping (Kagan and Jackson 1999),

$$\eta = \frac{L_1 + L_2}{2D} \tag{5}$$

where L represents the rupture length, considered as double source radii R while D is a distance between events hypocenters. The value of η larger than 1.0 suggests that the rupture zones overlap. Table 4 includes η for all analyzed pairs. All doublets, including Ev1–Ev6, have η larger than 1 strongly suggesting overlapping of its rupture zones. For the first group η coefficients, however, are more than four times greater that for pair Ev1–Ev6. Let us also note that if we consider the distances including the higher location errors in epicentral coordinates, only pair Ev1–Ev6 has a η less than 1.

The previous part was mostly based on visual inspection of seismic signals. Now, we are going to investigate waveforms similarities using signal analysis. This approach has many advantages, nevertheless very serious problem, namely signal clipping (Fig. 7) caused by network low

Table 4 The parameter η which measures degree of fault rupture overlapping

Doublet	L1 + L2	D	η
Ev1–Ev6	732	193	1.89
Ev2–Ev4	746	80	4.66
Ev3–Ev5	746	48	7.77
Ev7–Ev8	792	83	4.77

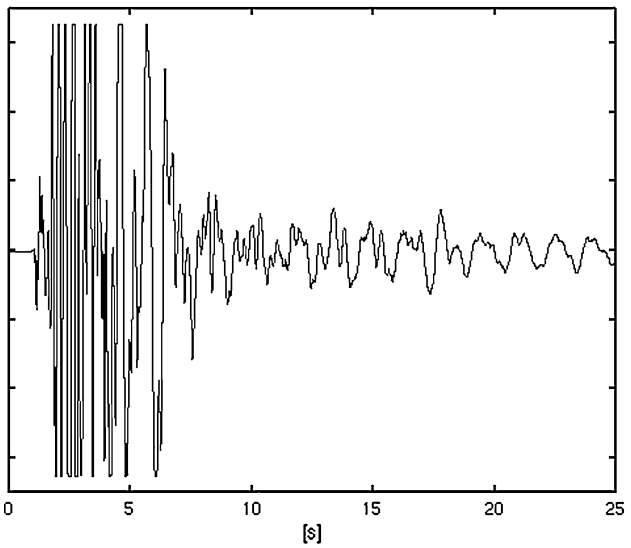


Fig. 7 Example of clipped signal recorded on selected seismic station

dynamic range is a factor influencing our investigation. To minimize this issue, we used not only normalized signal cross-correlation $C(\tau)$ (CC) (e.g., Schaff and Waldhauser 2005 and references therein):

$$C(\tau) = \frac{1}{\sqrt{\sum_{n=0}^{N-1} X^2(t_n) \sum_{n=0}^{N-1} Y^2(t_n)}} \sum_{n=0}^{N-1} X(t_n) Y(t_n + \tau) \quad (6)$$

where X and Y are two discrete time series of length N and τ denotes the time lag but also binary cross-correlation (BCC) and frequency cross-correlation (FCC) both eliminate saturation effects. Analysis was done with seismic traces for all possible doublets—stations pairs. During analysis with CC method, to avoid signal clipping problems, correlation measurements were made just for 5 s windows lengths, 1 s before and 4 s after first arrivals. Next maximal values of correlation coefficients were used for further study.

The resulting CC coefficients were gathered into two types of figures. The first one includes histograms, which show a number of stations versus coefficient intervals. The second one describes spatial distributions of waveform similarities in the form of the map, where correlation coefficient intervals are sketched for each station. An overview of the normalized cross-correlation results are given in Figs. 8, 9, 10 and 11. Further all possible event pairs (28 cases without coefficients for pairs consist of the same event) were investigated toward average CC values. The values range between 0.3 and 0.8. Only pairs Ev2–Ev4, Ev3–Ev5, among all studied pairs were characterized by average CC equal or exceeding 0.6. In Fig. 12 (left), average CC coefficients are presented for all events.

As in the previous analysis, different doublets are characterized by different CC features. The pair Ev2–Ev4 exhibits not only high signal overlapping in selected channel but has also the highest average CC. The next two pairs, Ev3–Ev5 and Ev7–Ev8, have average CC > 0.6, even though for doublet Ev7–Ev8 eight records have very small CC—below 0.2. On the other hand some stations for the pairs Ev7–Ev8 and Ev2–Ev4 reached CC values above 0.9 while for Ev3–Ev5 most signals have an CC 0.7. Let us note that northeastern stations exhibit smaller CC, especially in case of Ev1–Ev6; we expect that it can be, at least partially, caused by some random environmental and cultural seismic noise background. In some way, differences in background noises can be explained by a very large mining waste neutralization repository. The facility is located in the eastern part of the Rudna's area (see Fig. 1). The Ev1–Ev6 pair is characterized by the lowest average CC, although the high similarity of the full MT solution with dominance of compensated linear vector dipole (CLVD) term and almost identical T-axis orientation. The main difference which can influence the signal CC is the nodal plane and P-axis orientation whose strike differ by about 90° in this case. These features of the rupture orientations were crucial for seismic wave radiation pattern of those events, which were obviously different according to the results of average CC of the studied signals. Since there is not a sufficient similarity in waveforms, the pair Ev1–Ev6 is not a doublet.

For CC analysis we used just 5 s signals and for pair Ev7–Ev8 there is not direct evidence that obtained CC values denoted a doublet, we decided to support our double hypothesis by eliminating clipping effects. The technique is based on BCC methodology and in comparison with CC method uses the whole available signal data. This part of the data analysis consists of four steps concerning signals:

- Mean and trend removal,
- Ambient noise elimination by normalizing all amplitudes larger than a fixed value to zero,
- Median filtering
- Signal binarization using one-bit normalization method.

Figure 13 shows a raw seismograms (top), seismogram after signal processing (middle) and binary signal one-bit

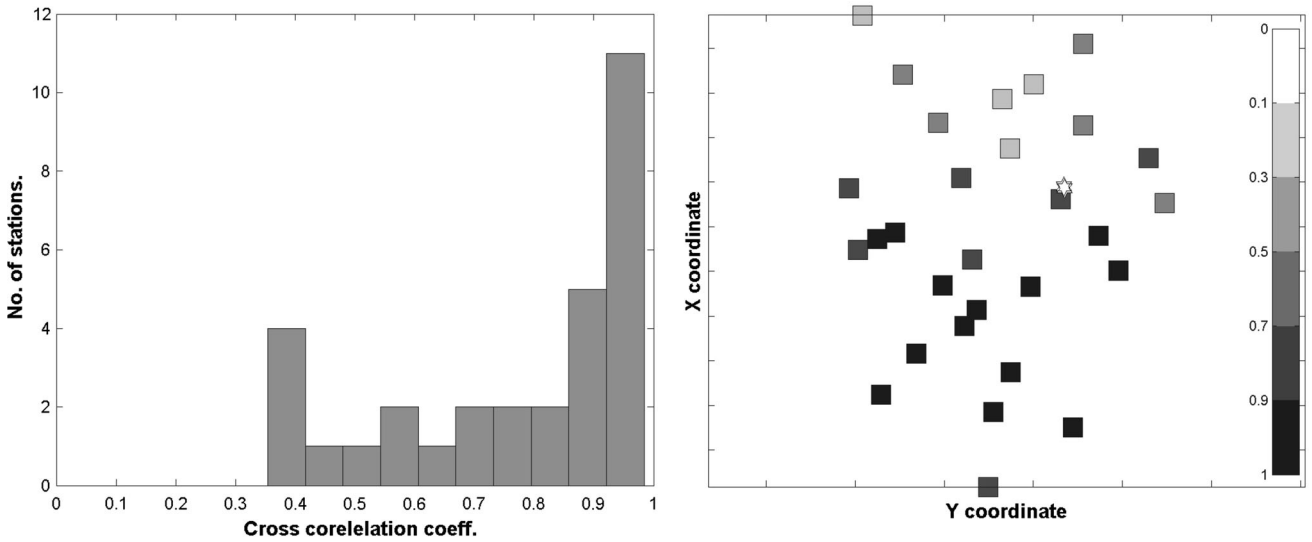


Fig. 8 Doublet Ev2–Ev4. Cross-correlation coefficients and its spatial distributions. The *black square* indicates the station with the highest correlation while *white stars* are epicenters

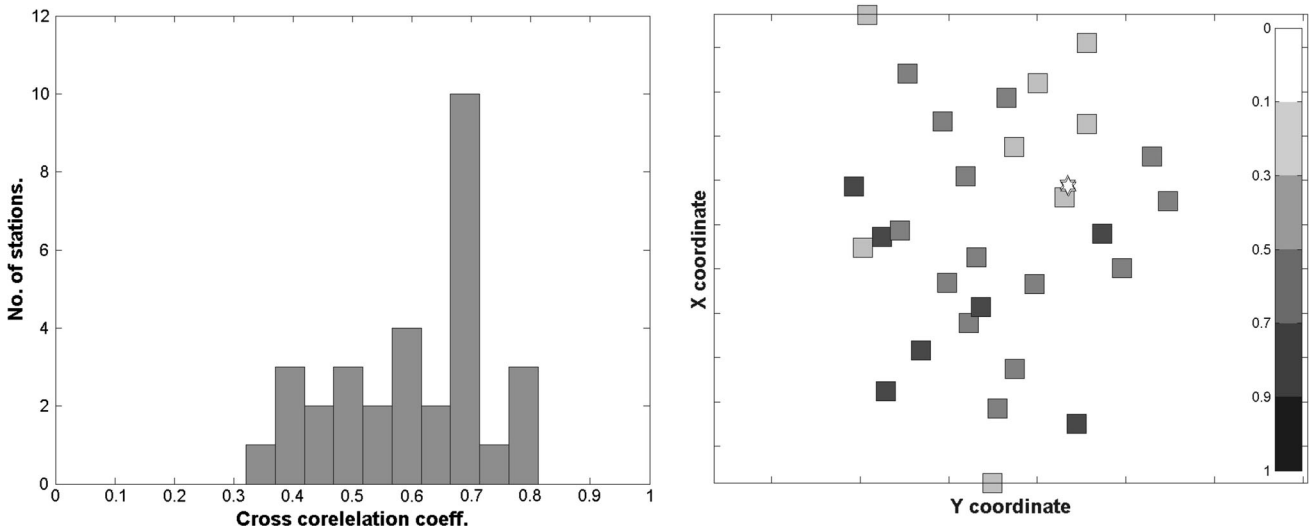


Fig. 9 Doublet Ev3–Ev5. Cross-correlation coefficients and its spatial distributions. The *black square* indicates the station with the highest correlation while *white stars* are epicenters

normalization (bottom) (Cupillard et al. 2011) for one selected event. For all possible events pairs maximal value of BCC coefficient were calculated for all stations. Results are collected in Fig. 12 (middle) as an average BCC coefficient. These results strongly support observation obtained with CC methodology. We noticed that the biggest BCC average values are connected with the proposed doublets including Ev7–Ev8 except Ev1–Ev6. Our results suggest that the BCC approach can be valuable analysis for data affected by clipping. It will be very important issue in case of clipped records within seconds after P wave first arrival.

Third method used in our studies was the FCC. In this method authors calculate correlation between histograms of frequencies of signals obtained in previous step. Frequency vector was calculated as $1/(k_i \cdot \mathrm{d}t)$, where $\mathrm{d}t$ is sampling rate, and k_i is number of following samples without changing a sign. Histogram was calculated with step 0.05 Hz. Results are gathered in Fig. 12 (right). The results obtained with the FCC are more complicated in comparison with the CC and the BCC solutions. The similarities for pairs Ev2–Ev4 and Ev7–Ev8 are clearly visible. On the other hand the pair Ev3–Ev5 has average coefficient even smaller than pair Ev1–Ev6. Moreover, there are few more possible pairs like Ev1–Ev7 or Ev1–Ev8 with quite high coefficients. Nevertheless, when the results from tree methods are gathered together, we clearly see that doublets Ev2–Ev4, Ev3–Ev5 and Ev7–Ev8 are similar in the sense of seismograms similarities.

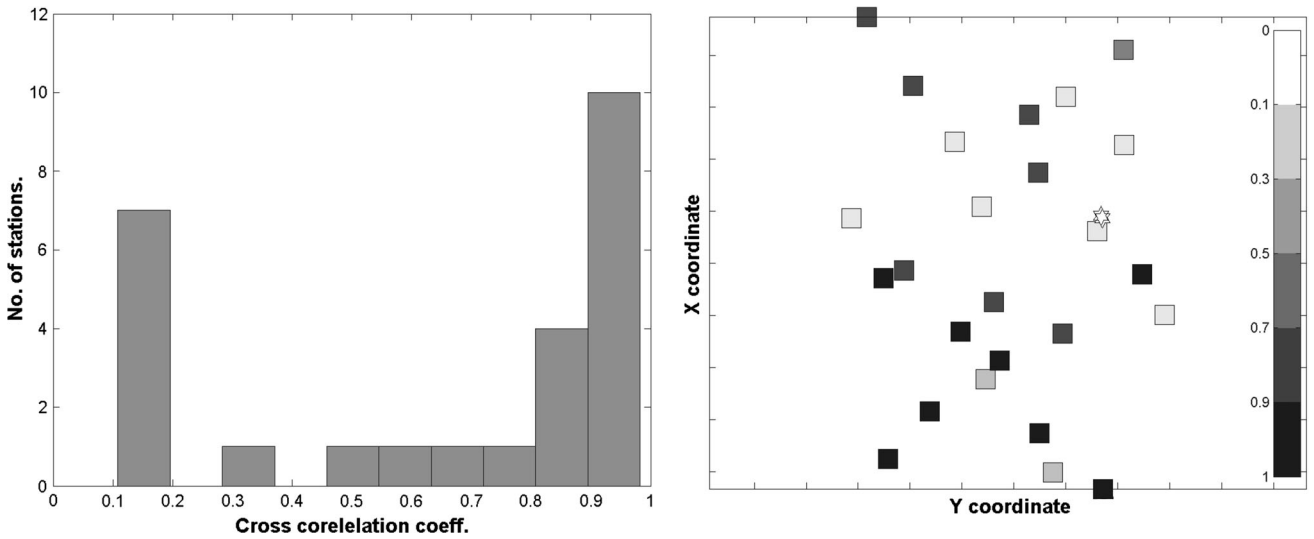


Fig. 10 Doublet Ev7–Ev8. Cross-correlation coefficients and its spatial distributions. The *black square* indicates the station with the highest correlation while *white stars* are epicenters

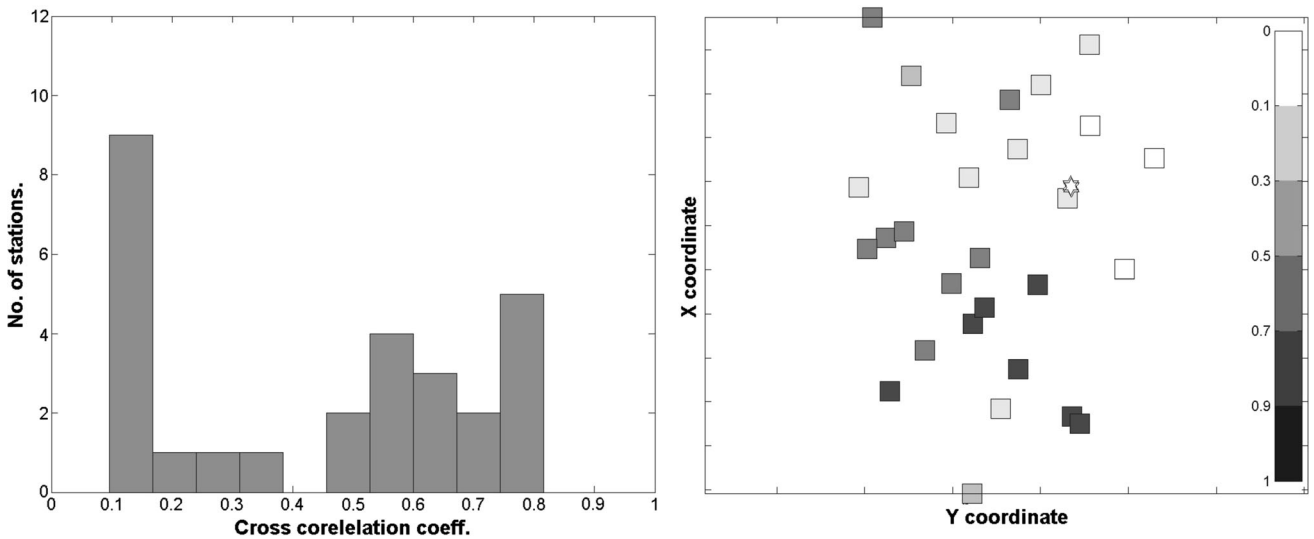


Fig. 11 Doublet Ev1–Ev6. Cross-correlation coefficients and its spatial distributions. The *black square* indicates the station with the highest correlation while *white stars* are epicenters

Discussion and conclusions

Our study shows that an analysis of waveform similarities can be a robust method to distinguish two similar induced seismic events occurring on mining area even in case of data set recorded on network with low dynamic range. Among all possible combinations of events pairs, just three doublets exhibit signals analysis features to be treated as doublets and all of them belong to the first group. Our expectation is that average similarity threshold should be above $CC > 0.6$. The same conclusions can be derived if we used BCC/FCC analysis.

Some very interesting features observed within an analysis of particular doublets should be concluded. Apart from the most obvious, namely the influence of events separations and general types of focal mechanisms to waveform similarities, we can observe very strong influence coming from nodal planes orientations resulting from the main stress axes orientation. It is especially interesting since the non-DC part of the full MT solutions within the studied pairs was not influencing the result of average BCC/CC/FCC analysis as much as the nodal plane orientations differences. This observation suggests that non-DC parts in the full MT could be contaminated by noise but on

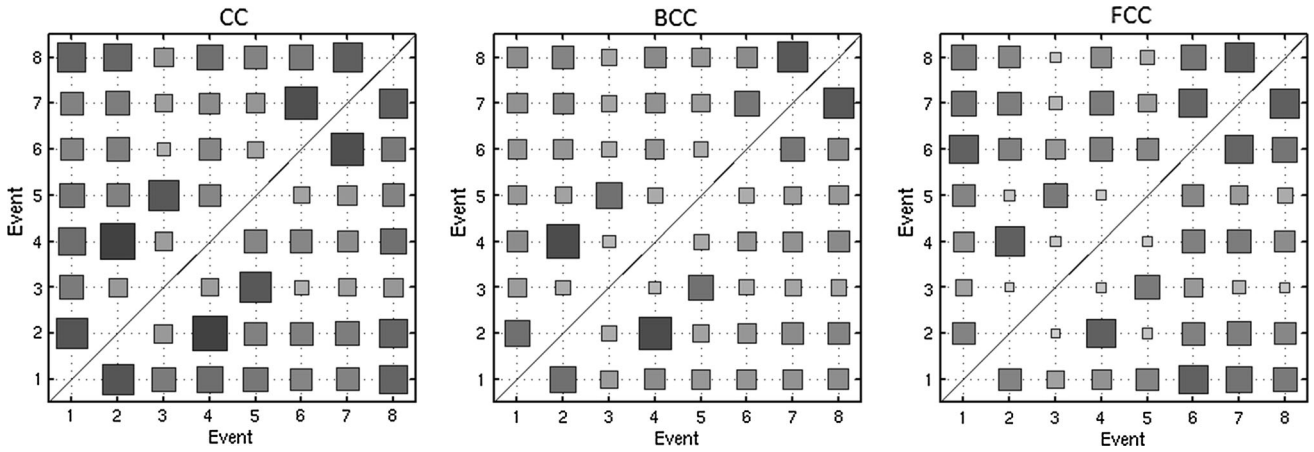


Fig. 12 Graphical comparison of averaged correlation coefficients using all three methods

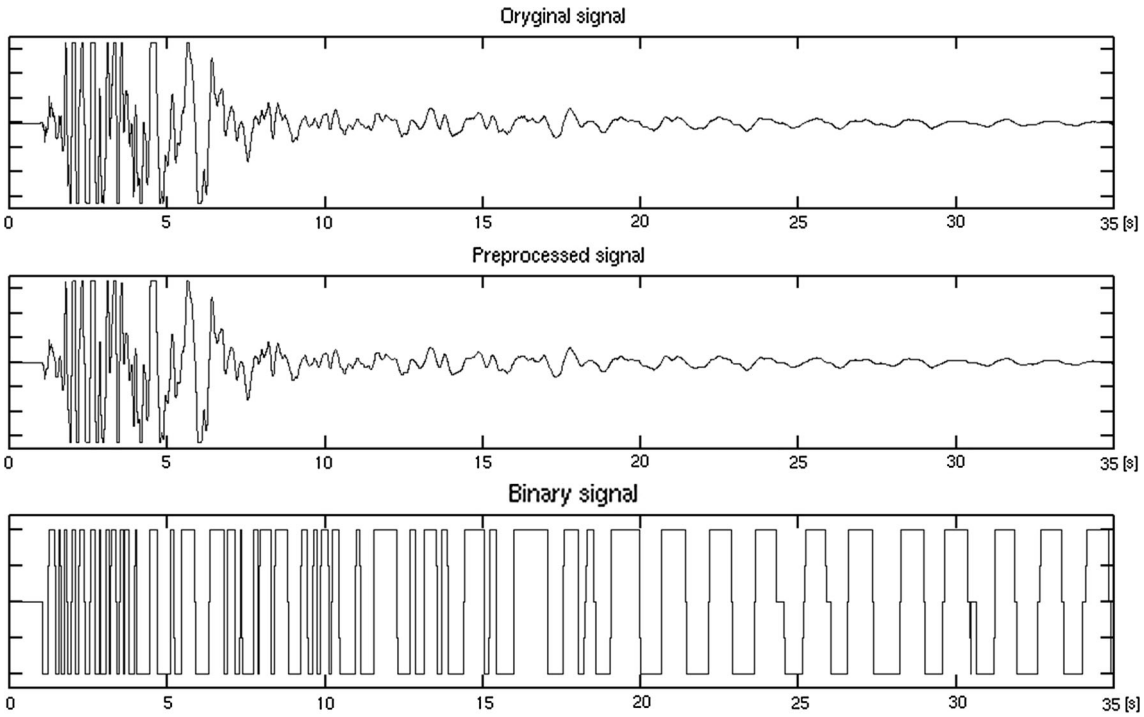


Fig. 13 Seismic signal binarization. *Top*—raw seismogram, *middle*—seismogram after first three steps of analysis (see for details in text), *bottom*—binary signal

the other hand the thorough study provided by Orlecka-Sikora et al. (2014) states that source mechanisms occurred on the panel XX/1 are better explained by the full MT (tensile source mechanism) and extended source rather than DC part of point source full MT. Let us note that the difference in strike of the nodal plane was crucial. From four pairs taken under investigation, the only pair with low average BCC/CC/FCC was the one with significantly different nodal plane strikes. Knowing that, we calculate a radiation patterns for all pairs caused by slip on fault. We followed with method described by Kwiatek and Ben-Zion (2013) and in Fig. 14 we present the radiation patterns for Ev1 and Ev6. The pair has small CC, especially on northeastern part of the network (Fig. 11). It can be noticed that the radiation pattern can be another feature responsible for that. The inter event time is a less influencing factor. Even though the doublet with the highest CC occurred within just 28 days, next two doublets have time differences up to 127 days and measured CC higher than 0.6. We can conclude that the P and T axis orientation and, therefore, the nodal planes of DC solution of the full MT plays the most significant role in the similarity analysis. The

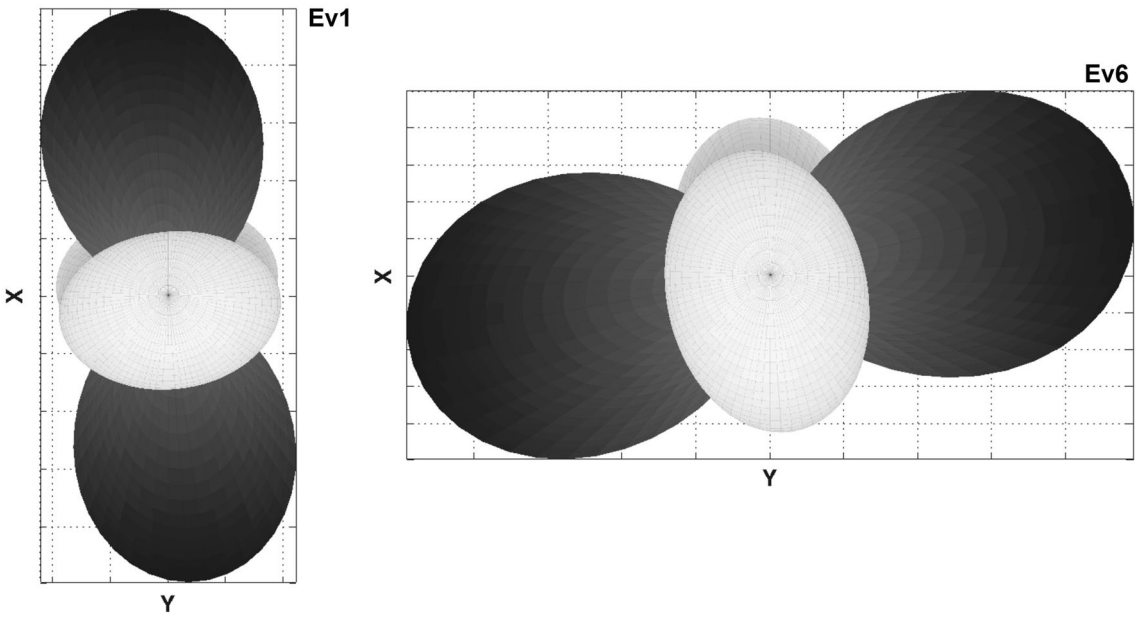


Fig. 14 Radiation patterns dealt with fault orientations of pair Ev1 and Ev6. The differences of pressure axis (*black axis*) could be considered as responsible for smallest CC on the northeastern part of the network

more similar nodal planes and P, T axis orientations the more similarity in waveforms we observed. Since the most similar pairs have nodal planes in accord with the main discontinuities in the mining panel, they probably followed the same rupture. This may be an important indicator of the rupture properties within the rocks in the particular mining environment.

The last but not least is that the BCC and FCC signal analysis work very well with clipped seismograms although the results obtained with the BCC much better support proposed doublets. The method should be especially valuable when clipping affects the whole record and CC with raw signal is hardly useful for studies. Results presented here could be treated as a first step to provide automated signals analysis to identify similar events on mining environments.

Acknowledgements We are very thankful to the "Rudna" copper mine for providing seismic data (catalogue and seismic signals) and technological information. The work was supported by statutory activities No. 3841/E-41/S/2017 of the Ministry of Science and Higher Education of Poland. Most of figures were prepared using the GMT package (Wessel and Smith 1998). Anonymous reviewers are appreciated for their valuable comments which were helpful in improving the article.

References

Aki K, Richards PG (2002) Quantitative seismology. University Science Books, Sausalito

Andrews DJ (1986) Objective determination of source parameters and similarity of earthquakes of different size. In: Das S, Boatwright J, Scholz CH (eds) Earthquake source mechanisms, Maurice Ewing, vol 6. American Geophysical Union, Washington, D.C, pp 259–267

Anooshehpoor A, Brune JN (2001) Quasi-static slip-rate shielding by locked and creeping zones as an explanation for small repeating earthquakes at Parkfield. Bull Seismol Soc Am 91(2):401–403

Awad H, Kwiatek G (2005) Focal mechanism of earthquakes from June 1987 swarm in Aswan, Egypt, calculated by the moment tensor inversion. Acta Geophys Pol 53(3):275–291

Baisch S, Ceranna L, Harjes H-P (2008) Earthquake cluster: what can we learn from waveform similarity? Bull Seismol Soc Am 98(6):2806–2814. doi:10.1785/0120080018

Brune JN (1970) Tectonic stress and the spectra seismic shear waves from earthquakes. J Geophys Res 75:4997–5009

Brune JN (1971) Correction. J Geophys Res 76:5002

Cesca S, Dham T, Tolg Sen A (2013) Moment tensor clustering: a tool to monitor mining induced seismicity. Geophys Res Abstr 15, EGU2013-12871

Cupillard P, Stehly L, Romanowicz B (2011) The one-bit noise correlation: a theory based on the concepts of coherent and incoherent noise. Geophys J Int 184(2011):1397–1414. doi:10.1111/j.1365-246X.2010.04923.x

De Natale G, Iannaccone Martini M, Zollo A (1987) Seismic sources and attenuation properties at the Campi Flegrei volcanic area. Pure Appl Geophys 125(6):883–917

Evans KF, Moriya H, Niitsuma H, Jones RH, Phillips WS, Genter A, Sausse J, Jung R, Baria R (2005) Microseismicity and permeability enhancement of hydrogeologic structures during massive fluid injections into granite at 3 km depth at the Soultz HDR site. Geophys J Int 160(1):388–412. doi:10.1111/j.1365-246X.2004.02474.x

Fitch TJ, McCowan DW, Shields MW (1980) Estimation of seismic moment tensor from teleseismic body wave data with application

to intraplate and mantle earthquakes. J Geophys Res 85:3817–3828
Geller RJ, Mueller CS (1980) Four similar earthquakes in central California. Geophys Res Lett 7:821–824
Gibowicz SJ (2006) Seismic doublets and multiplets at Polish coal and copper mines. Acta Geophys 54(2):142–157. doi:10.2478/s11600-006-0014-y
Gibowicz SJ (2009) Seismicity induced by mining: recent research. Adv Geophys 51:1–53
Gibowicz SJ, Kijko A (1994) An introduction to mining seismology. Academic Press, San Diego
Gibowicz SJ, Lasocki S (2001) Seismicity induced by mining: ten years later. Adv Geophys 44:39–181
Godano M, Bernard P, Dublanchet P (2015) Bayesian inversion of seismic spectral ratio for source scaling. Application to a persistent multiplet in the western Corinth rift. J Geophys Res 120(11):7683–7712. doi:10.1002/2015JB012217
Got JL, Frechet J, Klein FW (1994) Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea. J Geophys Res 99:15375–15386
Gregersen S, Wiejacz P, Dębski W, Domanski B, Assinovskaya B, Guterch B, Mäntyniemi P, Nikulin VG, Pacesa A, Puura V, Aronov AG, Aronova TI, Grünthal G, Husebye ES, Sliaupa S (2007) The exceptional earthquakes in Kaliningrad District, Russia on September 21, 2004. Phys Earth Planet Inter 164(1–2):64–74. doi:10.1016/j.pepi.2007.06.005
Haskell NA (1953) The dispersion of surface waves in multilayered media. Bull Seismol Soc Am 43:17–34
Kagan YY, Jackson DD (1999) Worldwide doublets of large shallow earthquakes. Bull Seismol Soc Am 89:1147–1155
Kwiatek G, Ben-Zion Y (2013) Assessment of P and S wave energy radiated from very small shear-tensile seismic events in a deep South African mine. J Geophys Res 118:3630–3641. doi:10.1002/jgrb.50274
Kwiatek G, Martinez-Garzon P, Bohnhoff M (2016) HybridMT: a MATLAB/shell environment package for seismic moment tensor inversion and refinement. Seismol Res Lett. doi:10.1785/0220150251
Lizurek G, Wiejacz P (2011) Moment tensor solution and physical parameters of selected recent seismic events at Rudna copper mine. In: Idziak AF, Dubiel R (eds) Geophysics in mining and environmental protection, Geoplanet: Earth and Planetary Sciences 2. Springer, Berlin
Madariaga R (1976) Dynamics of an expanding circular fault. Bull Seismol Soc Am 66:636–666
Niewiadomski J (1997) Spectral analysis and seismic source parameters. In: Mendecki AJ (ed) Seismic monitoring in mines. Chapman & Hall, London, pp 144–158
Orlecka-Sikora B, Cesca S, Lasocki S, Lizurek G, Wiejacz P, Rudziński Ł (2014) Seismogenesis of exceptional ground motion due to a sequence of mining induced tremors from Legnica-Głogów Copper District in Poland. Geophys J Int. doi:10.1093/gji/ggu109
Poupinet G, Ellsworth WL, Frechet J (1984) Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras fault, California. J Geophys Res 89:5719–5731
Quintanar L, Rodríguez-González M, Campos-Enríquez O (2004) A shallow crustal earthquake doublet from the Trans-Mexican Volcanic Belt (Central Mexico). Bull Seismol Soc Am 94:845–855. doi:10.1785/0120030057
Rudziński Ł, Dębski W (2011) Extending the double—difference location technique to mining applications. Part I numerical study. Acta Geophys 59(4):785–814. doi:10.2478/511600-011-0021-5
Rudziński Ł, Cesca S, Lizurek G (2016) Complex rupture process of the 19 March 2013, Rudna Mine (Poland) induced seismic event and collapse in the light of local and regional moment tensor inversion. Seismol Res Lett 87:274–284. doi:10.1785/0220150150
Schaff P, Waldhauser F (2005) Waveform cross—correlation—based differential travel—time measurements at the Northern California seismic network. Bull Seismol Soc Am 95(6):2446–2461. doi:10.1785/0120040221
Sen AT, Cesca S, Bischoff M, Meier T, Dahm T (2013) Automated full moment tensor inversion of coal mining-induced seismicity. Geophys J Int 195:1267–1281. doi:10.1093/gji/ggt300
Spottiswoode SM, Milev AM (1998) The use of waveform similarity to define planes of mining—induced seismic events. Tectonophysics 289:51–60
Waldhauser F, Ellsworth WL (2000) A double—difference earthquake location algorithm: method and application to the northern Hayward Fault, California. Bull Seismol Soc Am 90:1353–1363
Wehling-Benatelli S, Becker D, Bischoff M, Frederich W, Meier T (2013) Indication for different types of brittle failure due to active coal mining using waveform similarities of induced seismic events. Solid Earth Discuss 5:655–698. doi:10.5194/sed-5-655-2013
Wessel P, Smith WHF (1998) New, improved version of the Generic Mapping Tools released. EOS Trans AGU 79:579
Wiejacz P (1992) Calculation of seismic moment tensor for mine tremors from the Legnica—Głogów Copper Basin. Acta Geophys Pol 40:103–122
Wyss M, Brune JN (1968) Seismic moment, stress and source dimensions for earthquakes in the California-Nevada region. J Geophys Res 73:4681–4694

Acta Geophys. (2017) 65:299–308
DOI 10.1007/s11600-017-0022-0

RESEARCH ARTICLE

Time reversal seismic source imaging using peak average power ratio (PAPR) parameter

Anna Franczyk[1] · Andrzej Leśniak[1] · Damian Gwiżdż[1]

Received: 20 February 2017 / Accepted: 21 February 2017 / Published online: 7 March 2017

Abstract The time reversal method has become a standard technique for the location of seismic sources. It has been used both for acoustic and elastic numerical modelling and for 2D and 3D propagation models. Although there are many studies concerning its application to point sources, little so far has been done to generalise the time reversal method to the study of sequences of seismic events. The need to describe such processes better motivates the analysis presented in this paper. The synthetic time reversal imaging experiments presented in this work were conducted for sources with the same origin time as well as for the sources with a slight delay in origin time. For efficient visualisation of the seismic wave propagation and interference, a new coefficient—peak average power ratio—was introduced. The paper also presents a comparison of visualisation based on the proposed coefficient against a commonly used visualisation based on a maximum value.

Keywords Time reversal · Seismic wave modelling · Source location

✉ Anna Franczyk
franczyk@geol.agh.edu.pl

Andrzej Leśniak
lesniak@agh.edu.pl

Damian Gwiżdż
damian.gwizdz@gmail.com

[1] AGH University of Science and Technology, Krakow, Poland

Introduction

Measurement of seismic emissions related to natural or human activity is an important tool in the recognition of internal structures and the dynamic behaviour of rock mass. The energy, source location and focal mechanism of such emissions are analysed. The seismic source determines how seismic energy radiates through the rock mass, meaning it has a crucial role in the analysis of kinematic and dynamic changes in a given area.

One of the key parameters used in an analysis of these changes is the location of emission sources.

In mining practice, source location is performed separately for each event, independently of other recorded events. Another method is the relative location method, for which the location of the source is determined relative to the location of adjacent sources. There are several variants of this method (see, e.g., Waldhauser and Ellsworth 2000; Rudzinski and Debski 2012; Douglas 1967). In oint event location, events are relocated according to the location of the so-called "master event" (Douglas 1967). Relative location methods usually improve the accuracy of source location; however, they are strongly influenced by data quality, similarly to non-relative methods of location (Rudzinski and Debski 2011; Debski and Klejment 2016). The crucial factors are determination errors of the aforementioned parameters, like P or S wave onsets.

The superimposing of two or more events in a seismic records, together with the signal-to-noise ratio of the data and the velocity model, are considered as the most important factors that hamper not only event location, but also analysis of its source mechanism and energy estimation. The coda component of the first event usually has larger amplitudes than the P wave of the second wave. Thus, the evaluation of the second events' parameters that

are needed for location is difficult or even impossible. Superposition of seismic events can happen for high energy-induced seismic emissions recorded in underground mines. Smaller, secondary events are recorded after the high energy event, and frequently overlap the main waveform. In such cases, it is only possible to locate the first event and the total energy of the group of events.

A promising solution in such cases could be utilisation of the whole waveform recorded by all the sensors (even for superimposed events) for source location. The time reversal imaging (TRI) technique utilises the fact that the wave equation is reversible in time, meaning that it does not change form when the time direction is reversed, for example $t \rightarrow -t$. This gives rise to the assumption that the observed wave propagation is symmetrical in time when energy dissipation is ignored. Positive orientation of the time axis allows the observation of the emission of waves from sources located beneath the surface and propagation toward sensors located on the surface. For negative time direction, the time-reversed recorded signals are the sources of energy in sensors' positions and back propagate through the rock mass to their sources. The energy is then focused at the point where the source is located. Similarly, for superimposed events from a few sources, time-reversed propagation from sensors focuses at the individual sources. The quality of the focusing depends largely on the knowledge of the velocity structure of the geological media where the wave propagates.

The idea of TRI was proposed in the 1960s by Parvulescu and Clay (1965) for submarine communication. These ideas were then further expanded by Fink (Fink et al. 1989; Fink 1992, 1997), who made TRI better known. The development of theoretical aspects of the method and its application in non-destructive testing were continued by Anderson et al. (2009a, b) and Saenger et al. (2011). The TRI technique in seismology has been used in source location and identification of source mechanisms (Gajewski and Tessmer 2005; Larmat et al. 2006, 2010; Kawakatsu and Montagner 2008; Steiner and Saenger 2010; Artman et al. 2010; Debski 2015). In seismic exploration the TRI technique has been applied to wave field migration (Baysal et al. 1983; McMechan 1983; Tarantola 1988; Fichtner et al. 2006) and structure imaging in complex geological conditions like salt domes (Willis et al. 2006). Most of these projects exploited TRI for an assumed single point source location. TRI methods have also been successfully applied to the location of multiple simultaneous sources (Saenger et al. 2011). However, according to previous results, exploration of the potential of TRI to detect the details of finite rupture processes does not provide encouraging results (Blomgren et al. 2002; Kremers et al. 2011). The main problem is caused by the limits of the source size that could be reconstructed (Anderson et al. 2011).

In the present paper, we discuss the possibilities of using TRI in the location of sequences of seismic events using the new imaging coefficient: peak average power ratio. The former definition of this parameter is presented in "Quantitative assessment of focusing for multiple source point simulations with the same excitation time". We use synthetic seismic data generated in stratified two-dimensional geological media. The modelled sequences consist of several events with the same origin time and with relatively small delays between them. In this work, a comparison between the location of seismic sources and a visualisation based on the proposed parameter and maximum absolute value is also presented.

The paper is organised as follows: In the second section, the methodology used for location of sequences of seismic events together with a description of the PAPR imaging parameters is presented. Subsequently, we present synthetic tests of the proposed imaging technique. Numerical modelling was performed for different source distribution and different temporal source emission. The first example simulated random distribution of seismic events. The second example simulated a deterministic case: a rupture process with sources laid out perpendicular to the receiver array and with a slight time lag between subsequent source emissions.

Outline of the methodology

For our TRI of the sequence of seismic events, we employ a finite difference algorithm to model wave propagation in 2D acoustic media. The location of source coordinates was performed in three steps based on the PARP coefficient value. In the first step, the synthetic seismograms for the given geological model and the exact locations of the source points and receivers were computed. In the second step of the computational algorithm, the full synthetic seismograms were reversed in time in order to perform the role of source functions. These seismograms were used in the backward propagation of the seismic wave that was the third stage of the TRI rupture process imaging experiment. The backward propagation was performed iteratively and at each time step of the propagation algorithm the value of the PARP coefficient was updated and stored.

Forward simulation

The goal of the forward simulation in our case was to create a synthetic dataset for acoustic emissions in the given geological model. Synthetic seismograms were

computed for constant density using a two-dimensional acoustic wave field equation described by

$$\frac{\partial^2 P}{\partial t^2} = \kappa \left[\frac{1}{\rho} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} \right) P + \nabla \cdot f \right], \quad (1)$$

where $P(x, z)$ is the pressure field, f denotes body force, $\rho = \rho(x, z)$ denotes density of the medium and $\kappa = \kappa(x, y)$ is the bulk modulus.

The acoustic wave field equation was transformed to a system of first-order hyperbolic linear equations (Virieux 1986).

The equation was solved numerically using a staggered-grid, finite difference method that is stable for all values of Poisson's ratio and characterised by relatively small grid dispersion and grid anisotropy (Levander 1988). In this work, second order temporal and fourth order spatial operators under Cartesian coordinates were used.

For semi-infinite space modelling, free surface boundary conditions at $z = 0$ were applied. The other boundaries at the grid periphery were coded to satisfy the wave absorbing conditions (Cerjan et al. 1985). The source wavelet in the forward modelling was estimated with the Ricker wavelet, the second derivative of a Gaussian. We restricted modelling to non-dissipative media for simplicity.

Source location reverse modelling

The backward propagation of the seismic wave presented in this work is based on the same acoustic numerical modelling in two-dimensional space as the forward modelling. In contrast to standard seismic wave modelling, which uses seismic explosions as the initial boundary condition, time-reversed recordings are used in the backward propagation as the sources of seismic waves. A modelling scheme the same as the modelling scheme used in the first stage of computation is used to propagate the wave field backward into the assumed model. Assuming that the receiver configuration is sufficiently dense and regular, the time-reversed wave field will focus in space and time in the computational node corresponding to the coordinates of the source point. The focusing point should be noticeable as an area of abnormal values by visualisation of the seismic field obtained during backpropagation of the inversed seismograms.

The imaging parameters: peak average power ratio coefficient and the maximum value

The time reversal imaging methods that have so far been carried out are mostly based on the maximum value of a given parameter. The following were tested as imaging parameters: maximum horizontal and vertical displacement components (Hu and McMechan 1988; Steiner and Saenger 2012; Saenger 2011), maximum particle velocity (Steiner et al. 2008), strain components (Blomgren et al. 2002), maximum amplitude of pressure value (Gajewski and Tessmer 2005), stress components' energy density (Gajewski and Tessmer 2005; Saenger 2011), maximum P- and S wave energy density, maximum energy density, and

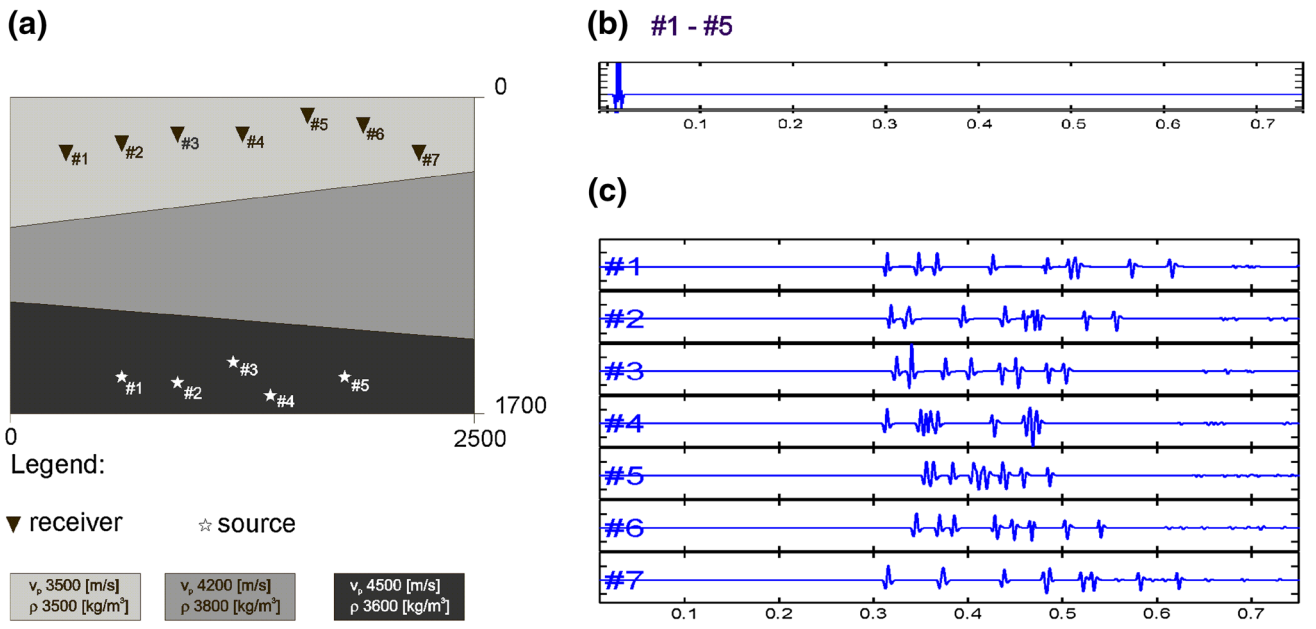


Fig. 1 Computational model (**a**), assumed source function (**b**), and synthetic seismograms (**c**). Seismograms (**c**) were computed as a result of forward simulation of the seismic signals (**b**) propagated from sources (#1–#5) located in the *bottom layer* of the computational model (**a**) to the assumed receivers network (*triangles* #1–#7) located in the *upper layer*

maximum stress components (Saenger 2011). With regard to the acoustic wave field modelling, the maximum absolute pressure value (MAPV) applied as an imaging condition appears to be a good indicator of the point source location (Eq. 2).

$$\mathrm{MAPV}(x) = \max_{t \in T} |P(x,t)|. \quad (2)$$

However, TRI using maximum pressure value can be problematic in the reconstruction in the seismic events that are not separated in time or space (Anderson et al. 2011).

The application of the PAPR coefficient is proposed in this paper to improve the spatial focus focal of seismic event detection.

The value of the PAPR coefficient can be computed from Eq. 3:

$$\mathrm{PAPR}(x) = \frac{P^2_{\max}(x)}{P^2_{\mathrm{total}}(x)/T}, \qquad P^2_{\mathrm{total}}(x) = \sum_{t=1}^{T} P^2(x,t), \quad (3)$$

where $P^2_{\max}$ is the maximum energy of the signal recorded in the given computational node, P^2_{total} is the total energy in the given node, $P(x,t)$ is the pressure value in each computational node, t is time index, and T is the number of computational time steps of the backward propagation of the seismic wave algorithm.

For both imaging conditions, the enormous values computed in the given computational node corresponding to the source point location are maintained during the whole process of backward wave propagations. Although both imaging parameters are calculated in much the same manner, the TRI with PAPR coefficient shows a higher spatial resolution that can improve the location of seismic event sequences.

Numerical case study

Synthetic multiple source point simulations with the same emission time

The results presented in this paper were obtained for an inhomogeneous, layered model. We assume that the P wave velocity increases from 3500 m/s in the first layer, through 4200 m/s in the second, up to 4500 m/s in the third layer (Fig. 1a). For the numerical tests, the five sources with coordinates summarised in Table 1 and the same excitation time $t = 0$ were used. The location of the sources is also depicted in Fig. 1a. The Ricker signal was used as a source function in the forward modelling (see Fig. 1b). The coordinates of the receivers are summarised in Table 2.

The receiver network reflects the locations and distances of surface seismic networks that are typical in Polish copper mines. Synthetic seismograms from these receivers computed in the forward simulation stage are presented in Fig. 1c.

The length of the source function is much shorter than the thickness of the geological layers in the assumed medium. As a result, the evaluated seismograms are mainly composed of isolated Ricker signals and interference between particular components is usually low (or medium for seismogram #4). The highest amplitude components are direct waves and the waves reflected from the surface. The secondary waves are almost invisible in the seismograms.

Pressure changes and the PAPR ratio were computed for every node of the computational grid during the second stage of modelling when the time-reversed seismograms from Fig. 1c were used as source functions.

Table 1 Coordinates of the source points assumed in numerical modelling

No.	x (m)	z (m)	f (Hz)
#1	600	1500	140
#2	900	1530	170
#3	1200	1420	130
#4	1400	1600	143
#5	1800	1500	147

Table 2 Coordinates of receivers used in numerical simulations

No.	x (m)	z (m)
#1	300	300
#2	600	250
#3	900	200
#4	1250	200
#5	1600	100
#6	1900	150
#7	2200	300

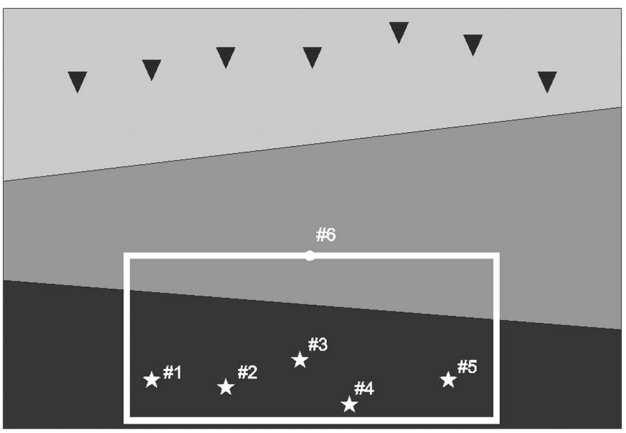


Fig. 2 Location of the analysis region (*white rectangle*) and the whole computational model

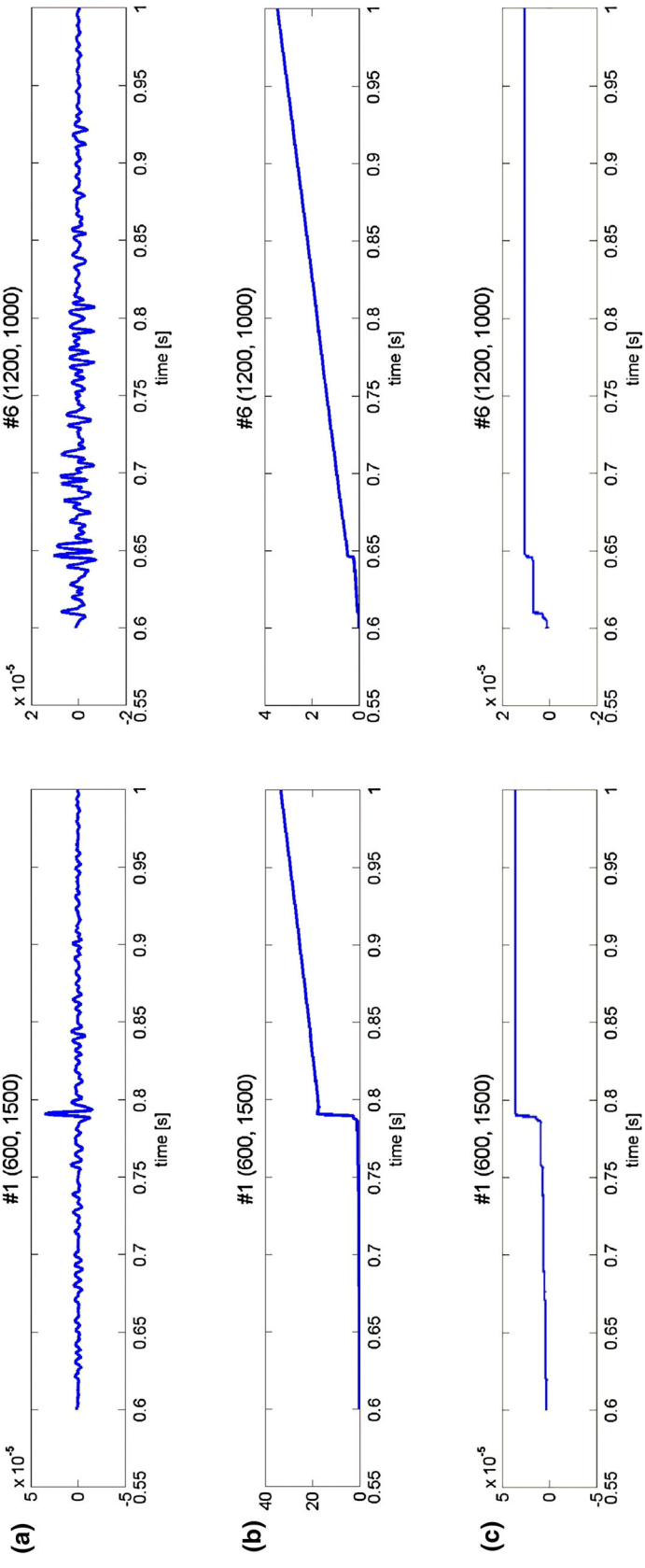


Fig. 3 Pressure changes (**a**) computed during backward propagation of the recorded signal, MAPV (**b**) and PAPR ratio changes (**c**) for source points #1 and for a given point (#6) placed outside the source area

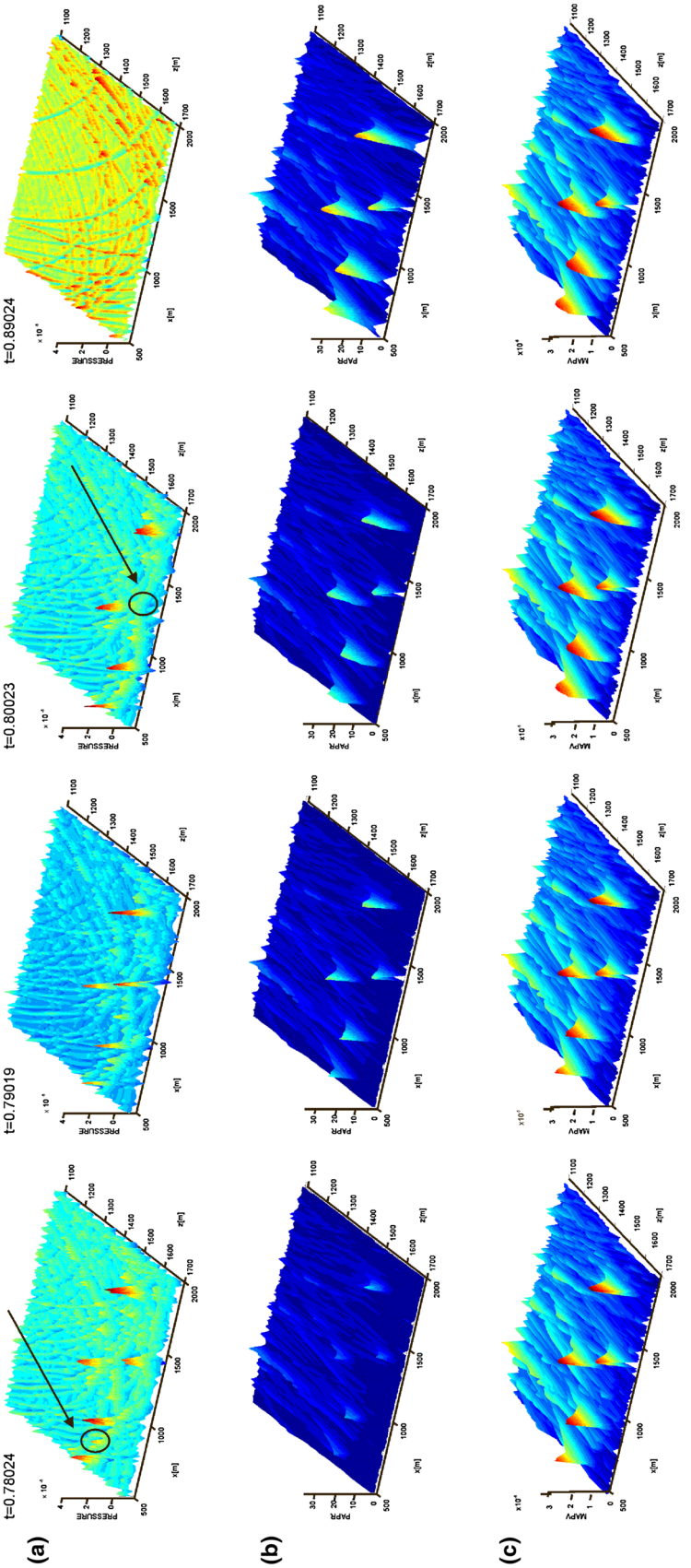


Fig. 4 Pressure values (**a**), PAPR ratio (**b**), and MAPV (**c**) computing during backward propagation stage for time steps corresponding to the wave excitation times (*first column*) and to time steps 0.01, 0.02, and 0.1 s after excitation time (*second*, *third*, and *fourth column*, respectively)

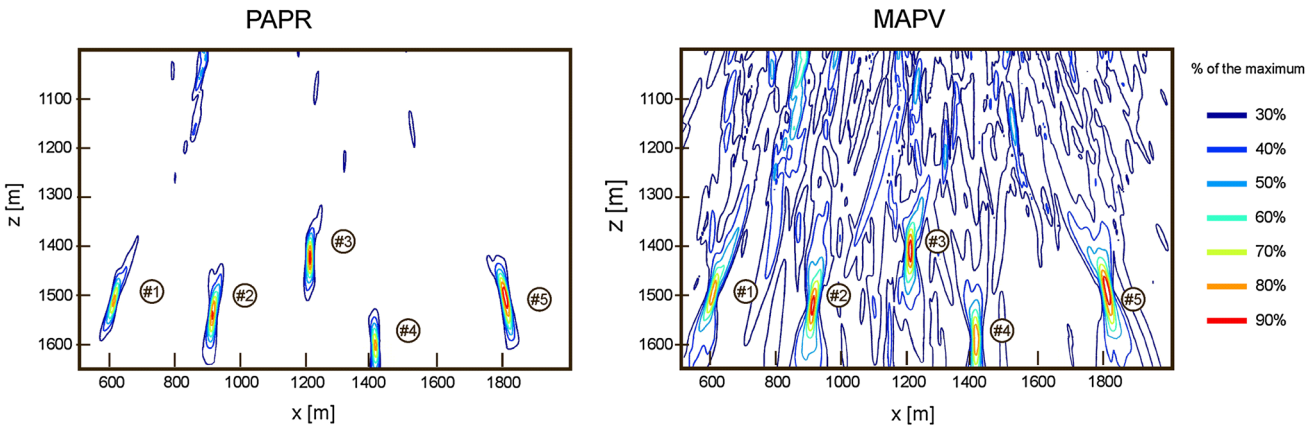


Fig. 5 The contour plot obtained after the last time step of the backward propagation stage, with areas of the highest values of PAPR ratio and MAPV normalised to the maximum values obtained during backward propagation of the seismic wave recorded for simulation of multiple sources emission

Quantitative assessment of focusing for multiple source point simulations with the same excitation time

The analysis presented in the paper was limited to the area in the vicinity of seismic sources where we expected the focusing effect of the seismic waves. The location of the study area superimposed on the entire computational model is presented in Fig. 2.

The pressure, the PAPR ratio and MAPV changes in time obtained during backward propagation of the seismic wave are presented in Fig. 3a–c, respectively. Plots were prepared for the computational nodes corresponding to the assumed source point location #1 (left column) and for one computational node placed in a source free point (plot #6). The location of the computational nodes used in visualisation is also shown in Fig. 2.

For all types of plots, it is easy to indicate the moment of concentration of the energy at the source point (plot #1, left column in Fig. 3). For the seismic signal plotted for computational node placed outside the source point there are no such significant change (plot #6, right column in Fig. 3). However, there are essential differences between pressure plots and plots of MAPV and PAPR value as MAPV and values of the PAPR ratio remains extremely large in the source point, whereas the increased pressure value fades over time. These relationships remain true for all the plots prepared for source points and for the plots prepared for the points that are outside the source area (Fig. 4).

The main difference between TRI using the PAPR ratio and TRI using MAPV is spatial resolution. As shown in Fig. 3b, the values achieved by the PAPR ratio in the source area are an order of magnitude greater than the values that it achieves outside this area. Values of MAPV outside the source area (Fig. 3c) are reduced only a few times in relation to the values that they take for the source area. In Fig. 5, the contour plots of the PAPR parameter values normalised to unity for given time steps are presented. Due to the transparency of the results, only values corresponding to 30, 40, …, 90% of the maximum value of the PAPR parameter are presented.

In both maps, only sources #2, #3, and #4 are indicated by the contours corresponding to the 90% of the maximum value. Both TRI parameters correctly locate all the sources with the threshold of 80 and 70% of maximum value. In the normalised PAPR map, a few false locations appear only for the threshold of 40 and 30% of maximum value, whereas for this threshold noise hampering can be observed in the correct location in the normalised MAPV maps.

Quantitative assessment of focusing for multiple source point simulations with different excitation times

The advantage of TRI using the PAPR ratio value in the detection of coordinates of seismic sources was tested also for the model with different excitation times of seismic sources. Figure 6 presents plots for the computational model with interfering seismic signals. We assume the same numerical model of subsurface rock structure with the same P wave velocities as the previous case (Fig. 1a). The five sources with coordinates presented in Table 3 with different excitation times were used in the numerical tests. The location of the sources is also depicted in Fig. 6a. The excitation time of the seismic waves in the particular point sources were so chosen to imitate the rupture process (Udias et al. 2014). Synthetic seismograms computed during the forward simulation are presented in Fig. 6c. The Ricker signal assumed as a source functions in all five

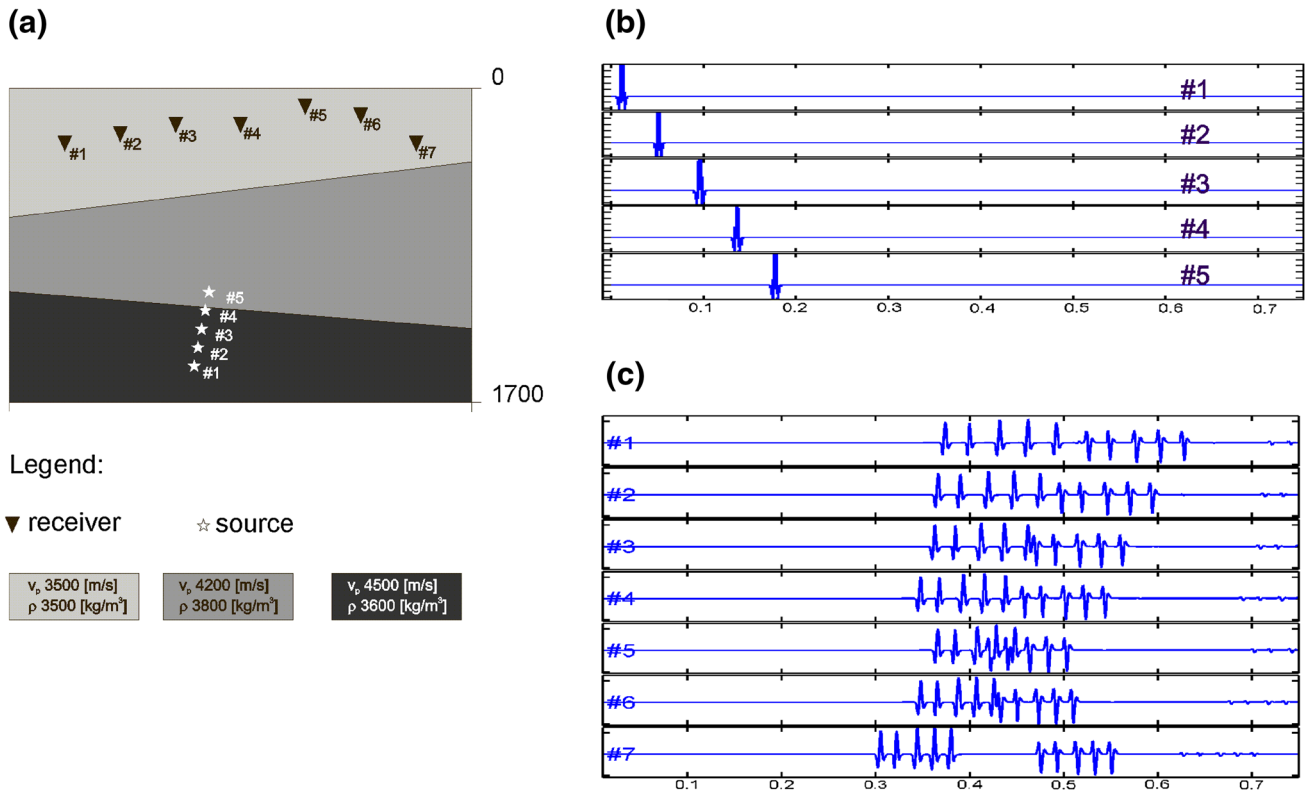


Fig. 6 Computational model (**a**), assumed source function (**b**), and synthetic seismograms (**c**). Seismograms (**c**) were computed as a result of the forward simulation of the seismic signals (**b**) propagated from sources (#1–#5) (**a**) to the assumed receiver network (*triangles* #1–#7) located in the *upper layer*

Table 3 Coordinates of the source points assumed in numerical modelling

No.	x (m)	z (m)	f (Hz)	Real time (s)
#1	1000	1500	140	0
#2	1020	1400	170	0.042
#3	1040	1300	130	0.083
#4	1060	1200	143	0.125
#5	1080	1100	147	0.166

computational nodes corresponding to the source points is presented in Fig. 6b. We assumed the same receiver network with coordinates presented in Table 3.

Results of the visualisation of backward propagation of seismograms reversed in time with the usage of MAPV and the PAPR ratio value are presented in Fig. 7. Once again, significant differences between both types of visualisation can easily be seen.

In both maps, only sources #3, #4, and #5 are indicated by the contours corresponding to 90% of the maximum value. The contour plotted for 80% of maximum values correctly indicates the location of the additional source (seismic source #1), whereas the contour plotted for 70% of the maximum correctly refocused all assumed seismic sources. The differences appear for the contour of 60% of maximum value. In the MAPV maps, all five source areas are merged into one big area, whereas for the PAPR map such a low resolution is obtained only for the threshold of 30% of maximum value.

Discussion and conclusions

This paper introduces a new imaging condition that improved spatial resolution in the problem of seismic sources location. The application of PAPR ratio as an imaging condition improved the refocus for multiple sources with both the same excitation time and with a slight time lag between subsequent source emissions that simulate the rupture process. The results presented in this paper refer to synthetic data with a well-known subsurface structure velocity model. The almost perfect focusing obtained in the two examples presented above was achieved using detailed knowledge of the subsurface velocity structure and application of narrow broadband source signals (see Figs. 1b, 6b). For real data, only the first factor can be improved by adjustment of the numerical model of the subsurface structure to the real velocity model. The lack of the detailed velocity model is

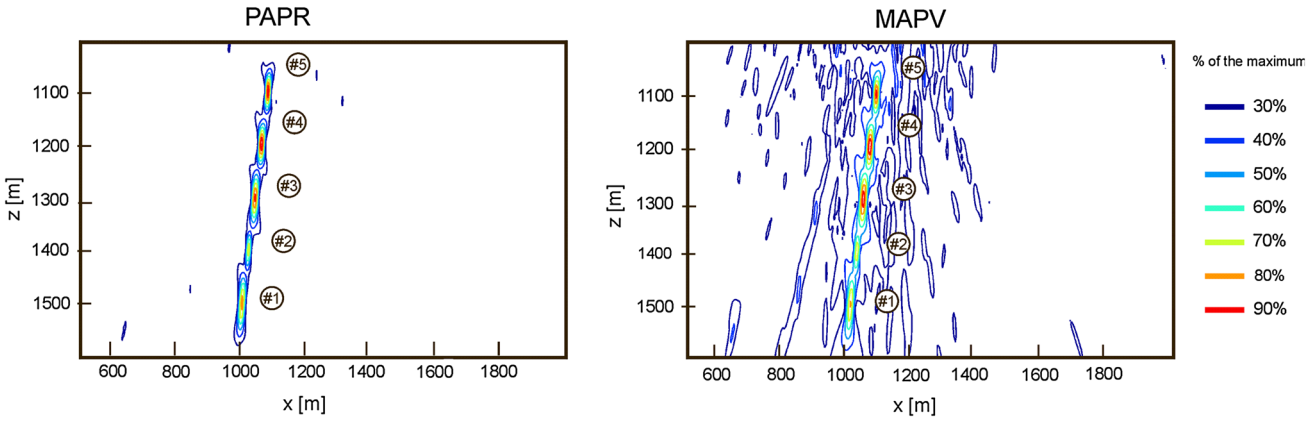


Fig. 7 The contour plot obtained after the last time step of the backward propagation stage, with areas of the highest values of PAPR ratio and MAPV normalised to the maximum values obtained during backward propagation of the seismic wave recorded for simulation of the rupture process

the undeniable limitation of the application of the proposed imaging condition to the real recordings. This is the common factor negatively affecting all location procedures.

The results presented in this paper do not consider the presence of noise in the seismic records. Although the low values of the signal-to-noise ratio are regarded as one of the important conditions for correct seismic event location, the TRI using the PAPR ratio may be resistant to the presence of noise. The value of the PAPR ratio is the averaged value obtained in the given point, additionally decreasing in time. The noise is usually uncorrelated so the averaging procedure decreases its amplitude. Based on the above statements one can say that the use of the PAPR ratio coefficient can improve the location results comparing to other imaging coefficients used in the TRI.

It is important to mention that the source location as a function of space and time cannot be determined simultaneously. As shown in the article, the application of the PAPR ratio improves spatial location of seismic sources. Having better location of seismic sources, we can go back through the time snapshots to examine the time at which the amplitude achieves its maximum value: this is the origin time of the seismic event. It makes it possible to determine the time history of the source emissions, and time sequence of multiple sources.

The proposed method of TRI with PAPR ratio can be extended to elastic data. In this case, more imaging parameters can be used instead of the pressure value. The effects of geometry of the receiver network on the refocusing results of the multiple source emission are currently under investigation. The application of the proposed imaging condition to real data is also planned for the future.

Acknowledgements This work was partly support by the AGH University of Science and Technology, Faculty of Geology, Geophysics and Environmental Protection, as a part of statutory Project 11.11.140.613. This paper was partially supported by Grant No. 2011/01/B/ST10/07305 from the National Science Centre, Poland.

References

Anderson BE, Guyer RA, Ulrich TJ, Johnson PA (2009a) Time reversal of continuous-wave, steady-state signals in elastic media. Appl Phys Lett 94(11):111908. doi:10.1063/1.3097811

Anderson BE, Ulrich TJ, Griffa M, Le Bas P-Y, Scalerandi M, Gliozzi AS, Johnson PA (2009b) Experimentally identifying masked sources applying time reversal with the selective source reduction method. J Appl Phys 105(8):083506. doi:10.1063/1.3079517

Anderson BE, Griffa M, Ulrich TJ, Johnson PA (2011), Time reversal reconstruction of finite sized sources in elastic media. J Acoust Soc Am 130(4):EL219–EL225. doi:10.1121/1.3635378

Artman B, Podladtchikov I, Witten B (2010) Source location using time-reverse imaging. Geophys Prospect 58(5):861–873. doi:10.1111/j.1365-2478.2010.00911

Baysal E, Kosloff D, Sherwood JWC (1983) Reverse time migration. Geophysics 48(11):1514–1524. doi:10.1190/1.1441434

Blomgren P, Papanicolaou G, Zhao H (2002) Super-resolution in time-reversal acoustics. J Acoust Soc Am 111(1):230–248. doi:10.1121/1.1421342

Cerjan C, Kosloff D, Kosloff R, Reshef M (1985) A non-reflecting boundary condition for discrete acoustic and elastic wave equations. Geophysics 50(4):705–708. doi:10.1190/1.1441945

Debski W (2015) Using meta-information of a posteriori Bayesian solutions of the hypocentre location task for improving accuracy of location error estimation. Geophys J Int 201(3):1399–1408. doi:10.1093/gji/ggv083

Debski W, Klejment P (2016) The new algorithm for fast probabilistic hypocenter locations. Acta Geophys 64(6):2382–2409. doi:10.1515/acgeo-2016-0111

Douglas A (1967) Joint epicentre determination. Nature 215(5096):47–48. doi:10.1038/215047a0

Fichtner A, Bunge H-P, Igel H (2006) The adjoint method in seismology I. Theory. Phys Earth Planet Inter 157(1–2):86–104. doi:10.1016/j.pepi.2006.03.016

Fink M (1992) Time reversal of ultrasonic field—part I: basic principles. IEEE Trans Ultrason Ferroelectr Freq Control 39(5):555–566. doi:10.1109/58.156174

Fink M (1997) Time reversed acoustics. Phys Today 50(3):34–40. doi:10.1063/1.881692

Fink M, Prada C, Wu F, Cassereau D (1989) Self-focusing in inhomogeneous media with time reversal acoustic mirrors. IEEE Ultras Symp Proc 1(2):681–686. doi:10.1109/ULTSYM.1989.67072

Gajewski D, Tessmer E (2005) Reverse modelling for seismic event characterization. Geophys J Int 163(1):276–284. doi:10.1111/j.1365-246X.2005.02732.x

Hu LZ, McMechan GA (1988) Elastic finite difference modelling and imaging for earthquake sources. Geophys J Int 95(2):303–313. doi:10.1111/j.1365-246X.1988.tb00469.x

Kawakatsu H, Montagner J-P (2008) Time-reversal seismic source imaging and moment-tensor inversion. Geophys J Int 175(2):686–688. doi:10.1111/j.1365-246X.2008.03926.x

Kremers S, Fichtner A, Brietzke GB, Igel H, Larmat C, Huang L, Kaser M (2011) Exploring the potentials and limitations of the time reversal imaging of finite seismic sources. Solid Earth 2(1):95–105. doi:10.5194/se-2-95-2011

Larmat C, Montagner J-P, Fink M, Capdeville Y, Tourin A, Clévédé E (2006) Time-reversal imaging of seismic sources and application to the great Sumatra earthquake. Geophys Res Lett 33(19):L19312. doi:10.1029/2006GL026336

Larmat C, Guyer RA, Johnson PA (2010) Time-reversal methods in geophysics. Phys Today 63(8):31–35. doi:10.1063/1.3480073

Levander A (1988) Fourth-order finite-difference P-SV seismograms. Geophysics 53(11):1425–1436. doi:10.1190/1.1442422

McMechan G (1983) Migration by extrapolation of time-dependent boundary values. Geophys Prospect 31(3):413–420. doi:10.1111/j.1365-2478.1983.tb01060.x

Parvulescu A, Clay CS (1965) Reproducibility of signal transmission in the ocean. Radio Electron Eng 29(4):223–228. doi:10.1049/ree.1965.0047

Rudzinski L, Debski W (2011) Extending the double-difference location technique to mining applications—part I: numerical study. Acta Geophys 59(4):785–814. doi:10.2478/s11600-011-0021-5

Rudzinski L, Debski W (2012) Extending the double difference location technique—improving hypocenter depth determination. J Seismol. 17(1):83–94. doi:10.1007/s10950-012-9322-7

Saenger EH (2011) Time reverse characterization of sources in heterogeneous media. NDT E Int 44(8):751–759. doi:10.1016/j.ndteint.2011.07.011

Saenger EH, Kocur GK, Jud R, Torrilhon M (2011) Application of time reverse modelling on ultrasonic non-destructive testing of concrete. Appl Math Model 35(2):807–816. doi:10.1016/j.apm.2010.07.035

Steiner B, Saenger EH (2010) Comparison of 2D and 3D time reverse modelling for tremor source localization. SEG Tech Program Expand Abstr 2010:2171–2175. doi:10.1190/1.3513275

Steiner B, Saenger EH (2012) Comparison of 2D and 3D time-reverse imaging—a numerical case study. Comp Geosci 46:174–182. doi:10.1016/j.cageo.2011.12.005

Steiner B, Saenger EH, Schmalholz SM (2008) Time reverse modelling of low-frequency microtremors: application to hydrocarbon reservoir localization. Geophys Res Lett 35(3):L03307. doi:10.1029/2007GL032097

Tarantola A (1988) Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation. Pure Appl Geophys 128(1):365–399. doi:10.1007/BF01772605

Udias A, Madariaga R, Buforn E (2014) Source mechanisms of earthquakes. Theory and practice. Cambridge University Press, Cambridge

Virieux J (1986) P-SV wave propagation in heterogeneous media: velocity-stress finite-difference method. Geophysics 51(4):889–901. doi:10.1190/1.1442147

Waldhauser F, Ellsworth W (2000) A double-difference earthquake location algorithm: method and application. Bull Seismol Soc Am 90(6):1353–1368. doi:10.1785/0120000006

Willis ME, Lu R, Burns DR, Toksoz MN, Campman X, de Hoop M (2006) A novel application of time reversed acoustics: salt dome flank imaging using walk away VSP surveys. Geophysics 71(2):A7–A11. doi:10.1190/1.2187711

Acta Geophys. (2017) 65:309–330
DOI 10.1007/s11600-017-0032-y

RESEARCH ARTICLE

Site-specific uniform hazard spectrum in Eastern Turkey based on simulated ground motions including near-field directivity and detailed site effects

Aida Azari Sisi[1] · Ayşegül Askan[1] · Murat Altuğ Erberik[1]

Received: 27 December 2016 / Accepted: 14 March 2017 / Published online: 24 March 2017

Abstract In this study, stochastic earthquake catalog of the Erzincan region in Turkey is generated based on synthetic ground motions. Monte Carlo simulation method is used to identify the spatial and temporal distribution of events. Ground motion time histories are generated using stochastic simulation methodology. Annual exceedance rate of each ground motion amplitude is calculated through statistical distribution of the complete set of ground motions. The results are compared with classical probabilistic seismic hazard analysis (PSHA). Classical PSHA generally produces larger spectral amplitudes than the proposed study due to wide range of aleatory variability. The effects of near-field forward directivity and detailed site response are also investigated on the results.

Keywords Ground motion simulation · Probabilistic seismic hazard analysis · Monte Carlo simulation · Near-field forward directivity · Site amplification function

Introduction

Probabilistic seismic hazard analysis is a common approach for modeling potential seismicity of a region. The method was first introduced by Cornell (1968) and extended later by numerous researchers (e.g., McGuire and Arabasz 1990; Kramer 1996; Thenhaus and Campbell 2003; McGuire 2004). The concept is widely used in seismic loss estimation methodologies (e.g., Cao et al. 1999; Luco et al. 2007; Eads et al. 2013).

PSHA is a powerful tool for estimating potential seismicity in a region, but it is known to have a few shortcomings: Naeim and Lew (1995) referred to unrealistic energy content of classical uniform hazard spectrum as PSHA considers a wide range of aleatory variability.

According to Bommer and Crowley (2006) PSHA also neglects intra-event variability. These authors proposed stochastically generated ground motion scenarios using Monte Carlo simulation (MCS), which were later employed by Crowley and Bommer (2006) for earthquake loss estimation. Assatourians and Atkinson (2013) defined a PSHA toolbox called EqHaz, which also employed the MCS approach. Wen and Wu (2001) followed the same concept to generate earthquake catalog for western, central and eastern United States. The difference was that Wen and Wu (2001) computed ground motion intensity parameters (GMIPs) through synthetic time histories instead of ground motion prediction equations (GMPEs). This methodology was later shown to produce agreeable estimates of linear and nonlinear structural demands by Gu and Wen (2007). Hence, it was used for derivation of fragility functions for building structures (Ellingwood et al. 2007). Hashash and Moon (2011) also developed hazard functions based on the same methodology for central US. Similarly, Papoulia et al. (2015) performed PSHA for Greece using synthetic ground motions.

Musson (2000) stated that the most important advantages of MCS-based hazard functions are the powerful handling of uncertainty and simplicity of disaggregation. Assatourians and Atkinson (2013) also pointed to

✉ Aida Azari Sisi
aida.sisi@metu.edu.tr

Ayşegül Askan
aaskan@metu.edu.tr

Murat Altuğ Erberik
altug@metu.edu.tr

[1] Civil Engineering Department, Middle East Technical University, Ankara, Turkey

flexibility and simplicity of the Monte Carlo simulation approach to be implemented in PSHA calculations.

Using synthetic ground motions rather than ground motion prediction equations to estimate GMIP enables one to consider the complex source effects (such as directivity), path effects (such as duration) and detailed local site effects in seismic hazard and risk assessment studies. Besides, ground motion prediction equations are not always capable of producing satisfactory results in regions with sparse data (e.g., Akansel et al. 2014; Raschke 2014). The effective role of site response in seismic hazard assessment was also highlighted in previous studies (e.g., Cramer 2006; Hashash and Moon 2011). Most attenuation models consider rough site categories as rock and soil (e.g., Ambraseys et al. 2005; Akkar and Bommer 2010). As a result, use of simulated ground motions instead of ground motion prediction equations in probabilistic seismic hazard analyses becomes a valid option. In that regard, validation of the simulations for representing the regional seismicity in an accurate manner is necessary.

In general, there are three categories for ground motion simulation methods. Deterministic methods mostly involve numerical solutions of wave propagation equation in heterogeneous media (e.g., Frankel 1993; Olsen et al. 1996). These techniques model the physical process accurately; however, they require well-resolved velocity models and considerable computational effort to simulate the ground motions. Stochastic approaches for modeling ground motions that account for randomness in ground motions are practical, but do not involve velocity models as refined as in the deterministic approaches (e.g., Beresnev and Atkinson 1997; Boore 2003). Yet, stochastic methods have been employed effectively in particular for higher frequencies (higher than 1 Hz) (e.g., Roumelioti et al. 2004; Yalcinkaya 2005; Ugurhan and Askan Gundogan 2010; Tahghighi 2012; Ugurhan et al. 2012; Demartinos and Faccioli 2012; Hung and Kiyomiya 2013). For regions without detailed source descriptions and well-resolved velocity models, stochastic methods are particularly preferred (e.g., Zafarani et al. 2009; Chopra et al. 2012; Askan et al. 2013). As the third option, hybrid methods have been developed more recently to overcome shortcomings related to both deterministic and stochastic techniques (e.g., Graves and Pitarka 2010; Nicknam et al. 2010). These techniques mainly combine motions derived from deterministic and stochastic methods for the low and high frequencies, respectively, and generate reliable broadband synthetics.

In this study, a recent approach is employed to derive site-specific uniform hazard spectrum (UHS) based on a stochastic earthquake catalog and ground motions simulated by the stochastic approaches. Monte Carlo simulation method is applied to determine temporal and spatial distribution of the earthquakes in a selected region. Ground motion amplitudes related to each scenario are obtained from simulated ground motion time histories. Stochastic finite-fault method as outlined in Motazedian and Atkinson (2005) is employed to generate the ground motions. Annual exceedance rate of each scenario is then calculated leading to hazard functions. As the main purpose of PSHA is to model regional seismicity, it is important to include regional seismological parameters within the calculations. This study focuses particularly on this issue by considering the effect of detailed local parameters in seismic hazard assessment through ground motion simulations. This study is quite a novel approach regarding Turkey, however, similar methodology has been applied for other countries (e.g., Wen and Wu 2001). The most important difference is that in this study, there is lack of recorded ground motions in Erzincan region but there are sufficient actual ground motions in the regions studied by Wen and Wu (2001). Hence Wen and Wu (2001) had the opportunity to compare the simulated ground motions with the actual ones.

After presenting initial results obtained from the proposed method, we perform sensitivity analyses in terms of selected parameters that affect the amplitude and frequency content of the generated ground motions. In order to include and test the effects of forward directivity, the model proposed by Mavroeidis and Papageorgiou (2003) is applied to the hazard functions. Then, the hazard curves and uniform hazard spectra with and without forward directivity are compared with each other.

Two alternative approaches are applied in this paper to characterize site response. The first one is generic site amplification proposed by Boore and Joyner (1997), which is based on local V_{s30}[1] at sites of interest. The second approach is the theoretical site amplification, which is based on transfer functions computed for layered earth models. One-dimensional (1D) standard site response analysis is employed for soil profiles to develop transfer functions. The results from both techniques are compared to figure out the effect of detailed local site response on seismic hazard.

Study area and methodology

The Erzincan region in Eastern Turkey is selected as the case study area in this paper. The region is in the relatively less-studied and sparsely monitored Eastern part of the North Anatolian Fault zone (NAFZ). Erzincan city is located in a tectonically very complex regime, close to the conjunction of three active faults, namely North Anatolian,

[1] Average shear wave velocity for the top 30 m of the subsurface profile.

North East Anatolian and East Anatolian Fault Zones (EAFZ) (Fig. 1). NAFZ displays right-lateral strike–slip faulting, whereas EAFZ and North East Anatolian Fault Zones (NEAFZ) have left-lateral strike–slip faulting in the area (Askan et al. 2013). This region is particularly selected since until very recently there were only a few strong motion stations around Erzincan. Indeed, the destructive 1992 Erzincan ($M_w = 6.6$) mainshock was recorded only by 3 stations within 200 km epicentral distance. It is thus difficult to select a suitable GMPE based on a comparison between the limited past dataset and predictive models. Thus, the region is a good candidate for the proposed approach where ground motion simulations are employed instead of GMPEs.

Three sites are selected in this region. Site 1 is Erzincan city center, which is very close to NAFZ and is located on soft soil. Site 2 is near Ovacik to southwest of Erzincan city, which is exposed to forward directivity effects more than the two other sites. Site 3 is inside Erzincan city near Cumhuriyet district, which is located on stiff soil. The sites are selected with different distances from NAFZ, different soil conditions and different rupture directivity characteristics. Table 1 presents the coordinates of selected sites together with the site classes according to the NEHRP classification. In this study, the seismic zones inside an effective area are considered. The effective area is defined as a circle with radius of 150 km around the site of interest. The coordinates and other properties of the seismic zones are derived from Deniz (2006). There are nine seismic zones consisting of five faults and four areal sources in the region of interest. Table 2 lists seismic properties of seismic zones and Fig. 2 shows the locations of these seismic zones along with the sites under study.

As the first step, the events are distributed within time spans using Monte Carlo simulation method. This approach is defined as a controlled selection of a random number from a probability distribution. The number of earthquakes within a specified time span related to each source is obtained assuming Poisson distribution (Wu and Wen 2000):

$$\sum_{X=0}^{n_k-1} \frac{(tv_k)^X}{X!}\exp(-tv_k) < u_k \leq \sum_{X=0}^{n_k} \frac{(tv_k)^X}{X!}\exp(-tv_k), \quad (1)$$

where v_k is activity rate of kth source, n_k is number of earthquakes inside kth source, t is time span which is taken as 10 years herein and u_k is a random number between 0 and 1 with uniform distribution. Equation (1) inserts uncertainties in calculations due to the generation of random numbers (u_k). To overcome large errors resulting from this randomness, the Monte Carlo simulations in Eq. (1) are repeated until a complete catalog is acquired. In this study, 1000 simulations are used so the catalog period is 10,000 years. Similarly, Wu and Wen (2000) discussed 9000 simulation years to be adequate. Papoulia et al. (2015) also used 10,000-year catalog for hazard calculations. To check the sufficiency of catalog period, total number of generated events per catalog period must be approximately equal to the sum of activity rates regarding the whole seismic zones. After identifying the total number of events, magnitude of each event is calculated through Gutenberg–Richter recurrence model:

$$\log(N) = a - bM, \quad (2)$$

Table 1 Location and site classes of the selected sites

Site ID	Coordinates	Site class (according to NEHRP)
Site 1	39.7464°N 39.4914°E	D
Site 2	39.6200°N 39.2000°E	D
Site 3	39.7566°N 39.4925°E	C

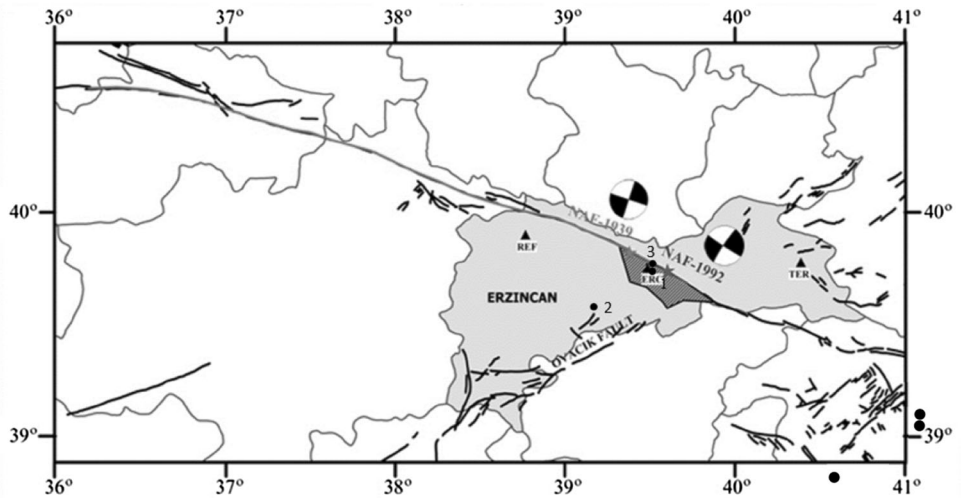


Fig. 1 Regional map showing the epicenters, rupture zones and the mechanisms of the 1939 and 1992 earthquakes (epicenters are indicated with *stars*) and strong ground motion stations that recorded 1992 Erzincan earthquake are indicated with *triangles*. The sites, which are used in this study, are indicated with *solid circles* with site numbers beside them (Adopted from Askan et al. 2013)

Table 2 Seismic parameters of the seismic sources in this study

No.	Name	M_{max}	M_{min}	Depth	λ	β
1	North Anatolian fault system segment D	8	4.5	25.04	1.07	1.347
2	East Anatolian fault system	7.5	4.5	24.29	2.14	2.161
3	North East Anatolian fault system	7.8	4.5	22.15	1.141	2.162
4	Central Anatolian fault system	7.1	4.5	20.1	0.56	2.74
5	Yazyurdu–Goksun fault zone	7	4.5	20.27	1.008	3.431
6	Background inner 3	5.4	4.5	6.67	0.075	2.197
7	Background inner 4	5.4	4.5	22.22	0.636	2.625
8	Background north	5.8	4.5	18.51	0.738	3.27
9	Background inner 5	5.6	4.5	36.62	1.996	2.395

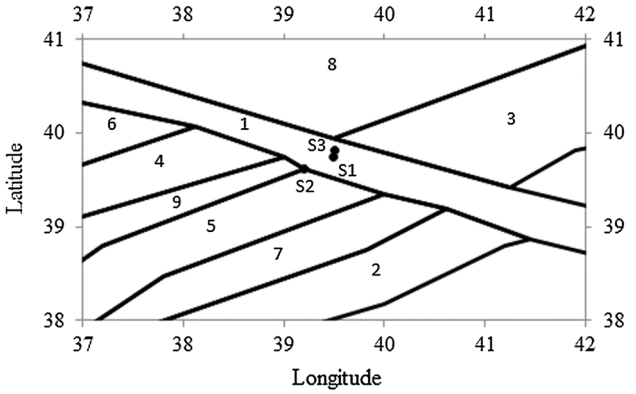


Fig. 2 Locations of seismic zones which are defined in Table 2. The three sites (S1, S2 and S3) in this study are shown with solid circles

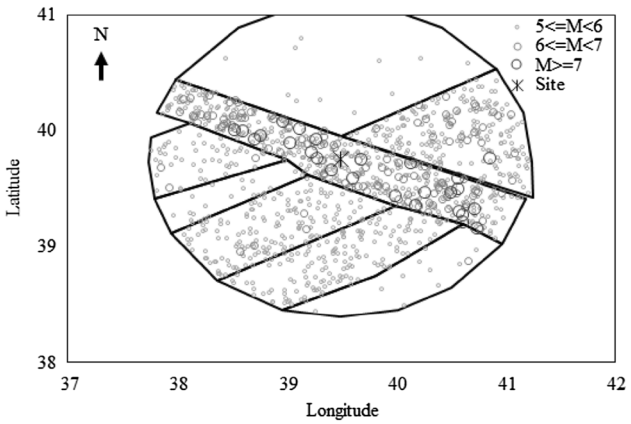


Fig. 3 Distribution of events inside effective area in a 3000-year stochastic earthquake catalog related to Site 1. The seismic zones inside the effective area are displayed with solid lines

where N is number of earthquakes with magnitudes larger than M; a and b are recurrence parameters corresponding to each source.

The epicenters of events are distributed randomly inside each seismic zone. Two random numbers for latitude and longitude are generated inside the borders of each source. A random number between 0 and seismogenic depth is generated for the depth parameter related to small events. On the other hand, for large events ($M_w > 6$) surface rupture is considered. Figure 3 shows distribution of epicenters inside effective area for a catalog period of 3000 years, related to Site 1. The seismic zones inside effective area and the site are displayed in Fig. 3 as well. The same 3000-year seismicity catalog as well as the seismic zones is shown in a broad area through Fig. 4.

Next, ground motion time series due to seismic waves propagating from epicenters to the site of interest, are simulated. The ground motions of events that occur on the faults are modeled using stochastic finite-fault model based on dynamic corner frequency proposed by Motazedian and Atkinson (2005). In this approach, the entire fault is divided into subfaults where each subfault is considered as a stochastic point source (Boore 2003). The contribution of each subfault to the amplitude of ground acceleration is then obtained from Eq. (3). The motions from subfaults are then summed up using an appropriate time delay as given in Eq. (4).

$$A_{ij}(f) = \left\{ CM_{0ij}H_{ij}(2\pi f)^2 / \left[1 + (f_{0ij})^2\right] \right\} \times \left\{ \exp(-\pi f \kappa) \exp(-\pi f R_{ij}/Q(f)\beta) D(f) G(R_{ij}) \right\}, \quad (3)$$

$$a(t) = \sum_{i=1}^{nl} \sum_{j=1}^{nw} a_{ij}\left(t + \Delta t_{ij}\right). \quad (4)$$

In these equations, C is a scaling factor, M_{0ij} is the seismic moment of ijth subfault, H_{ij} is the high-frequency scaling factor of ijth subfault to conserve high-frequency energy level, f_{0ij} is the dynamic corner frequency of ijth subfault, Δt_{ij} is the time delay of the ijth subfault, n_l and n_w are the number of subfaults along length and width of extended fault, respectively, κ is the high-frequency spectral decay (kappa) factor (Anderson and Hough 1984), R_{ij} is the distance to site related to ijth subfault, $Q(f)$ is frequency-dependent quality factor, β is shear wave velocity, $D(f)$ is site amplification and $G(R_{ij})$ is geometric spreading function.

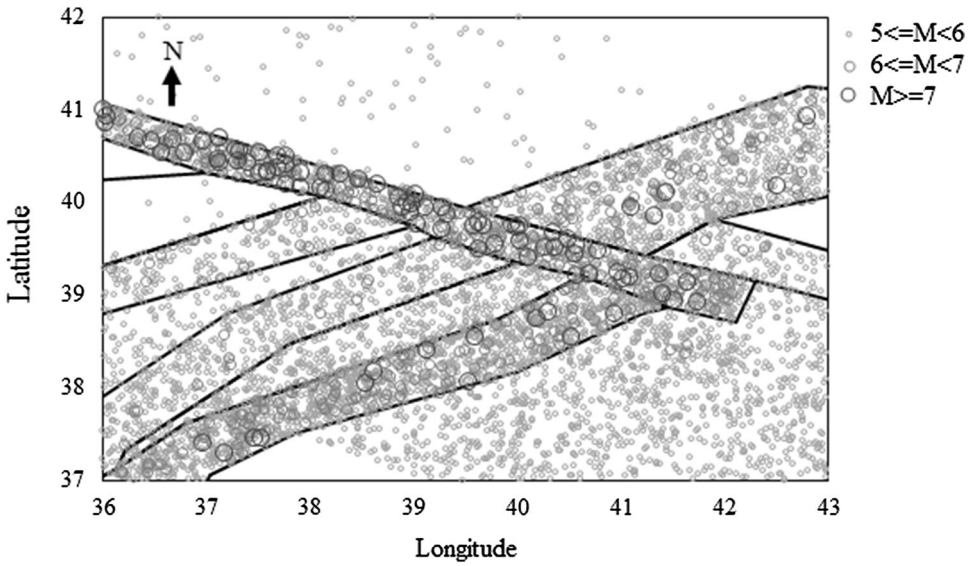


Fig. 4 Distribution of events in a 3000-year stochastic earthquake in a broader area around the effective area. The seismic zones in this area are displayed with solid lines

Dynamic corner frequency is a relatively new approach which models the dependency of corner frequency upon time. This approach results in conservation of the rupture energy regardless of subfault size. The dynamic corner frequency as a function of time and source parameters is shown as follows:

$$f_{0ij}(t) = N_R(t)^{-1/3} 4.9E + 6\beta(\Delta\sigma/M_0/N), \quad (5)$$

where $N_R(t)$ is number of ruptured subfaults up to time t, M_0 is total seismic moment, $\Delta\sigma$ is stress drop and N is total number of subfaults.

Hence, in this study local seismicity parameters are considered in hazard calculations, which is not the case in classical PSHA. Values of geometric spreading, quality factor, high-frequency decay factor and ground motion duration are adopted from Askan et al. (2013) (Table 3). Rupture dimensions are estimated from the empirical relationships defined by Wells and Coppersmith (1994). Stress drop is estimated from the empirical relations in Mohammadioun and Serva (2001) that relate its value to rupture dimensions as follows:

$$\Delta\sigma = 8.9 \times W^{0.8}, \quad (6)$$

where W is width of rupture. Finally, EXSIM computer program is used to model the extended faults (Motazedian and Atkinson 2005). This program has limitations while modeling ground motion duration and phase. However, it is currently the most effective approach for stochastic ground motion simulation and it has been used widely by researchers (e.g., Nicknam et al. 2010; Demartinos and Faccioli 2012).

For areal sources, stochastic point source method is used following the approach outlined in Boore (2003). The formulation of acceleration amplitudes is similar to Eq. (3) and SMSIM computer program (Boore 2003) is applied. Point source model is mostly preferred when dimensions of the source are negligible with respect to distance to site (Boore 2009). This point is taken into account in this study for areal sources that are far away from the site with relatively smaller magnitudes. Among all simulation parameters, value of the stress drop for point source simulations is different from the corresponding value for extended fault models.

As the final step, response spectrum of the synthetic ground motions is calculated for specified periods. Then ground motion amplitudes related to each period are sorted from largest to smallest, the jth ground motion amplitude has annual exceedance rate (AER) as follows:

$$\mathrm{AER}_j = \frac{j}{n}, \quad (7)$$

where n is catalog period in terms of year. The same approach for calculation of annual exceedance rate was also effectively employed by Assatourians and Atkinson (2013). The ground motion amplitudes related to the same hazard level for the entire period range yields site-specific

Table 3 Seismic parameters which are used in this study (Adopted from Askan et al. 2013)

Parameter	Value
Crustal shear wave velocity	3700 m/s
Rupture velocity	3000 m/s
Crustal density	2800 kg/m^3
Pulsing area percentage	50
Quality factor	$Q = 122f^{0.68}$
Geometrical spreading	$R^{-1.1}$ $R \leq 25$km
	$R^{-0.5}$ $R > 25$km
Duration model	$T = T_0 + 0.05R$
Windowing function	Saragoni–Hart
Kappa factor	Regional kappa model ($\kappa_0 = 0.066$)

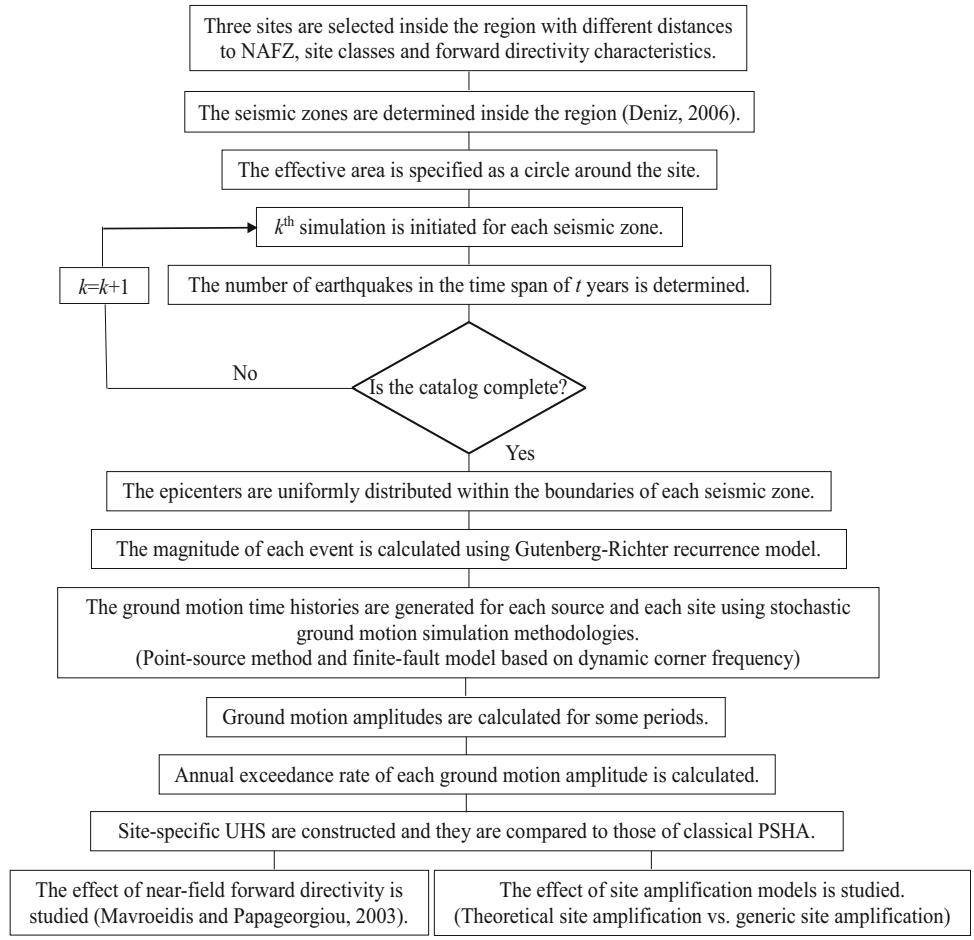


Fig. 5 Flowchart for step-by-step procedures in this paper

UHS. The whole procedures of the proposed approach in this study are coded with MATLAB programming software by the authors, also displayed in Fig. 5 within a flowchart.

Figures 6, 7 and 8 show hazard curves of PGA and PSA (at selected period values of $T = 0.1$, 0.2, 0.5, 0.7, 1.0, 1.5 and 2.0 s) as well as uniform hazard spectra of 2, 10 and 50% exceedance probability in 50 years for Site 1, Site 2, and Site 3, respectively. The results from the proposed study are compared with the results of classical PSHA. EZFRISK software is applied to perform classical PSHA in this study. The GMPE by Akkar and Bommer (2010) is used for classical PSHA herein as follows:

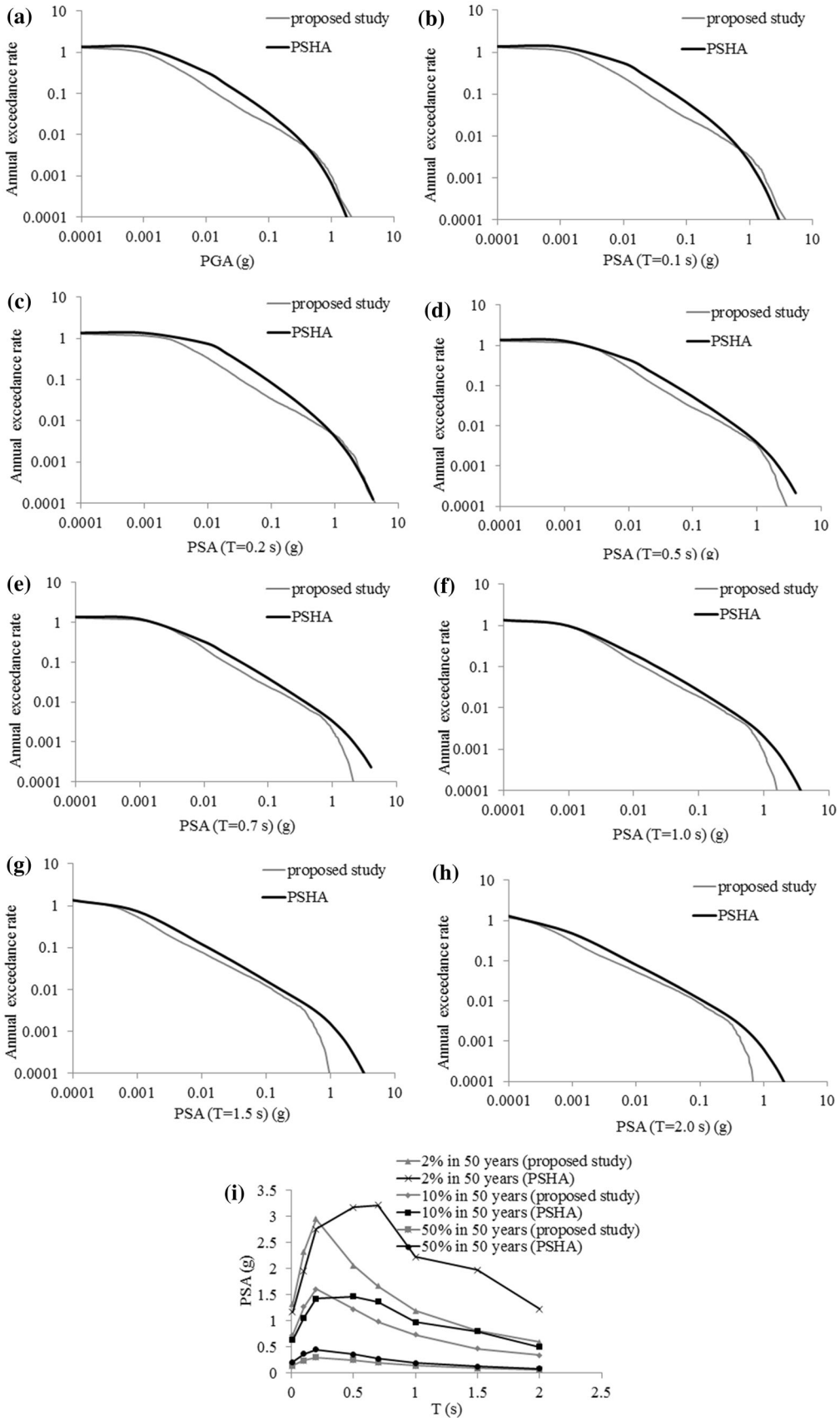
$$\log(\mathrm{PSA}) = b_1 + b_2 M + b_3 M^2 + (b_4 + b_5 M)\log\sqrt{R_{jb}^2 + b_6^2} + b_7 S_S + b_8 S_A + b_9 F_N + b_{10} F_R + \varepsilon\sigma, \quad (8)$$

where S_S and S_A take the value of 1 for soft and stiff sites, respectively, and zero for rock sites. F_N and F_R take value of 1 for normal and reverse focal mechanisms, respectively, and zero for strike–slip faults.

In Fig. 6, the ground motion amplitudes from the proposed study are observed to be slightly larger than PSHA for lower exceedance rates (2 and 10% in 50 years) and lower periods (shorter than 0.5 s). However, classical PSHA produces larger results for higher exceedance rates (higher than 10% in 50 years) within the whole period range. As period increases, hazard curves from the proposed method and PSHA converge to each other for higher exceedance rates while they diverge from each other for lower exceedance rates.

Fig. 6 Hazard curves for **a** PGA, **b** PSA ($T = 0.1$ s), **c** PSA ($T = 0.2$ s), **d** PSA ($T = 0.5$ s), **e** PSA ($T = 0.7$ s), **f** PSA ($T = 1.0$ s), **g** PSA ($T = 1.5$ s), **h** PSA ($T = 2.0$ s) and **i** UHS of proposed study and classical PSHA for Site 1

In order to observe the variance of simulated motions from the corresponding median values (ε) for Site 1, Site 2 and Site 3, the number of standard deviations versus simulated GMIPs are plotted in Figs. 9, 10 and 11, respectively. The median PSAs are obtained from predictive equations of Akkar and Bommer (2010). Figure 9 augments the above discussions. It is inferred from Fig. 9 that synthetic ground motions yield larger spectral amplitudes for larger GMIPs (generally larger than 0.1 g). These higher values refer to major events which may not be predicted closely by the ground motion prediction


(a)
proposed study
PSHA
Annual exceedance rate
PGA (g)
(b)
PSA (T=0.1 s) (g)
(c)
PSA (T=0.2 s) (g)
(d)
PSA (T=0.5 s) (g)
(e)
PSA (T=0.7 s) (g)
(f)
PSA (T=1.0 s) (g)
(g)
PSA (T=1.5 s) (g)
(h)
PSA (T=2.0 s) (g)
(i)
2% in 50 years (proposed study)
2% in 50 years (PSHA)
10% in 50 years (proposed study)
10% in 50 years (PSHA)
50% in 50 years (proposed study)
50% in 50 years (PSHA)
PSA (g)
T (s)

equations due to inherent data scarcity from large events. This discrepancy seems to be insignificant for larger periods where positive εs are observed for smaller PSAs. This observation confirms the previous discussions to some extent. For smaller periods and larger PSA values, εs of Fig. 9 are larger than or equal to dominant ε of PSHA. For larger periods, however, simulated PSA for low hazard levels decrease so that εs of Fig. 9 become less than the dominant ε of PSHA.

It is observed from Fig. 7 that at Site 2, PSHA gives larger PSAs compared to those from the proposed method for the entire period range and for each hazard level contrary to Site 1, which had an exception for low periods and low exceedance rates. There seems a remarkable difference between two approaches especially for the 2475-year return period in the longer period range. As the period increases, the difference between two approaches becomes more evident for low exceedance rates while the difference is not as obvious for the high exceedance rates.

Figure 10 is similar to Fig. 9 in terms of the decreasing and increasing trends of εs corresponding to larger and smaller ground motion amplitudes, respectively, with respect to period, which clarifies the previous observation. The only visible discrepancy is that number of standard deviations is less than 2 even for lower periods regarding Site 2, whereas it is larger than 2 for Site 1.

The results related to Site 3 in Fig. 8 are similar to those at Site 2, as PSHA produces considerably higher ground motion amplitudes than the proposed study especially for 2% in 50 years. Similarly, the hazard curves of the two methods become close to each other for higher exceedance rates and far apart for lower exceedance rates, respectively, as period increases. This observation is supplemented by Fig. 11: as period increases, εs for all of the three sites increase for lower GMIPs and decrease for higher GMIPs, lying in a plateau around $\varepsilon = 0$ in long-period region.

Overall, traditional PSHA yields larger ground motion amplitudes than the proposed methodology due to large aleatory variability of GMPE (sigma) as well as the wide range of standard deviations involved in classical PSHA (-3 to $+3$ standard deviations). Stochastic ground motion simulation introduces aleatory variability via random phase spectrum and it is believed to compute the ground motion amplitudes in a more physical manner than GMPEs. Hence there is no need to insert additional aleatory variability in seismic hazard calculations in the proposed study.

Sensitivity analysis

In this section, we initially study the ratio of the (equivalent) stress drop value used in point sources to that used in finite faults to calibrate the stress drop of point sources.

Fig. 7 Hazard curves for **a** PGA, **b** PSA ($T = 0.1$ s), **c** PSA ($T = 0.2$ s), **d** PSA ($T = 0.5$ s), **e** PSA ($T = 0.7$ s), **f** PSA ($T = 1.0$ s), **g** PSA ($T = 1.5$ s), **h** PSA ($T = 2.0$ s) and **i** UHS of proposed study and classical PSHA for Site 2

Then, we investigate the effect of several parameters on proposed method.

Stress drop for point sources

Atkinson et al. (2009) state that the stress drop parameter that controls the strength of the high-frequency radiation does not actually have the same meaning in EXSIM and SMSIM. In SMSIM, it is directly related to the Brune source model for a given stress parameter where the stress drops, corner frequency and seismic moment control the spectral amplitudes. However, in EXSIM, it has this meaning only for a particular subfault (Atkinson et al. 2009). Hence, stress drop must be larger for point source simulations in order to be consistent with the corresponding amplitudes of finite-fault simulations as discussed by Moghaddam et al. (2010).

To investigate this point, some comparisons are made between finite-fault and point source models using different values of stress drop for point source simulations. 13 March 1992 Erzincan earthquake recordings are used in the comparisons.

To quantify the differences, a misfit term is defined as follows:

$$E(f) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{A_i(f)_{\mathrm{FF}}}{A_i(f)_{\mathrm{PS}}}\right), \tag{9}$$

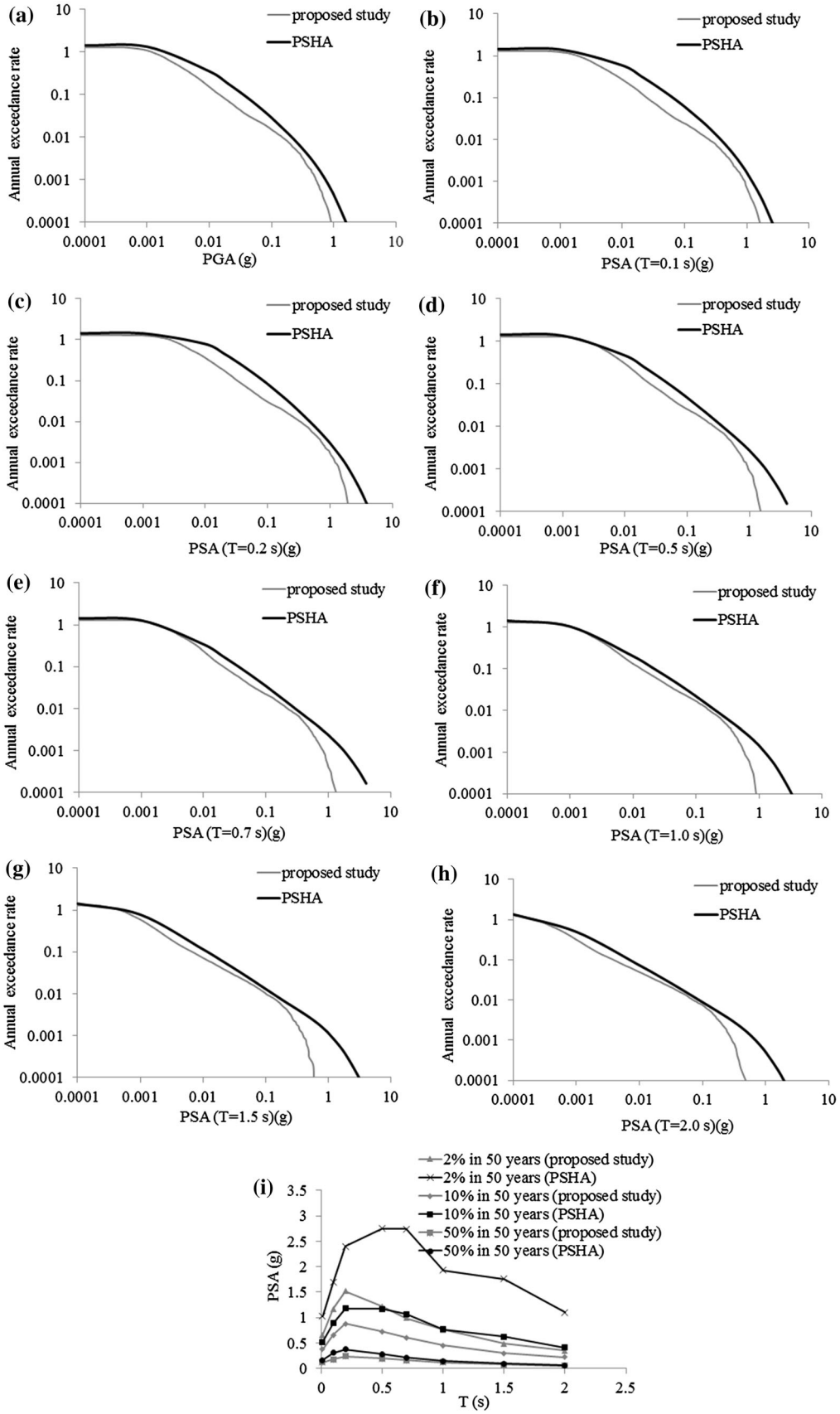
where n is number of stations which is 3 in this case study (abbreviated as ERC, TER and REF and shown in Fig. 1) since the 1992 Erzincan mainshock was recorded only by these stations within 200 km epicentral distance. $A_i(f)$ is the amplitude of response spectrum (PSA) or Fourier amplitude spectrum (FAS) related to finite-fault (FF) and point source (PS) models.

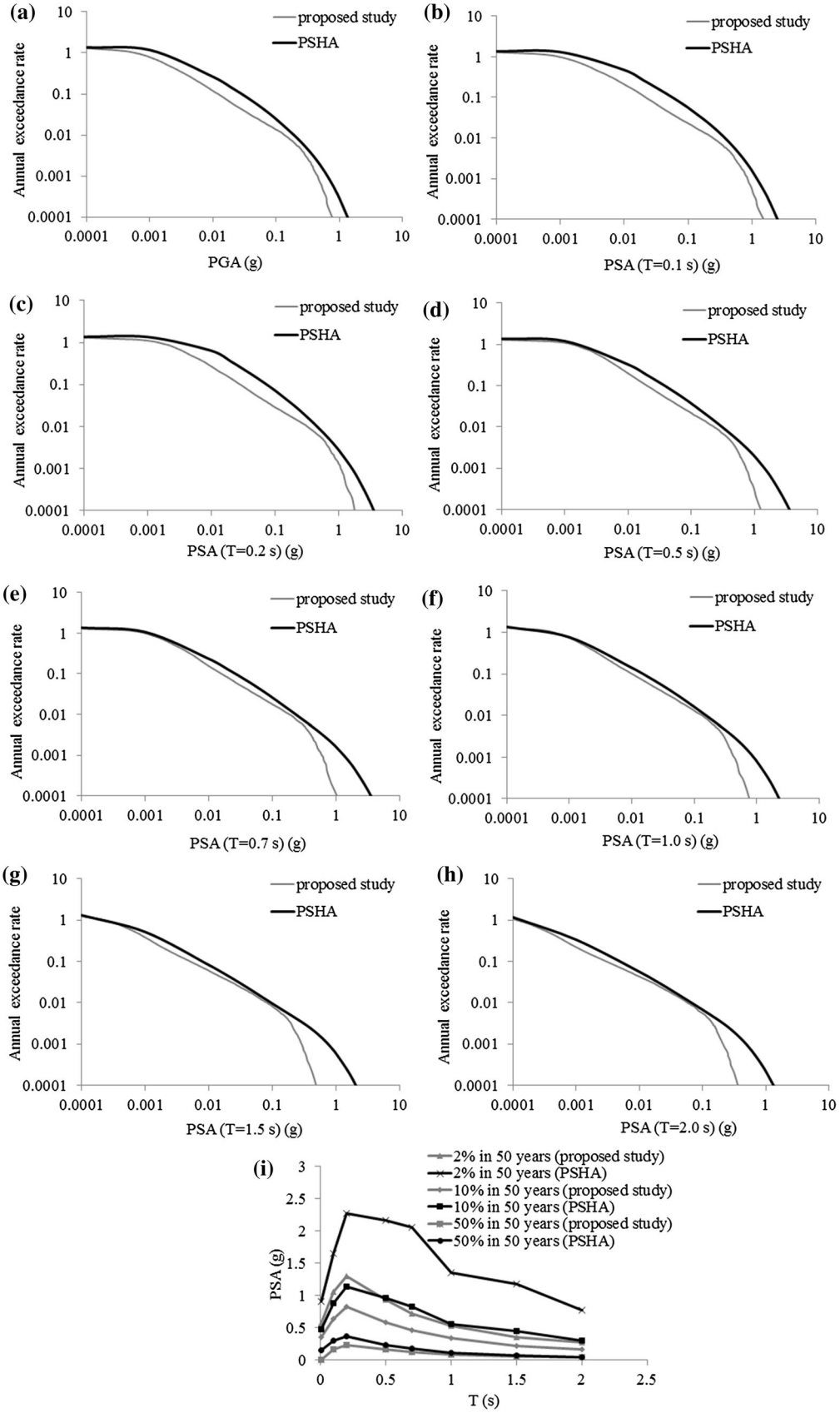
Next, a sensitivity index in the frequency domain is computed for each station as follows:

$$\mathrm{SI} = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{A_i(f)_{\mathrm{FF}}}{A_i(f)_{\mathrm{PS}}}\right), \tag{10}$$

where n is number of discrete frequencies considered and the other terms are as defined previously.

Ground motion time histories of the 1992 Erzincan earthquake are initially simulated using finite-fault model. The corresponding ground motions are also modeled using point source method with five different values of stress drop as 1, 1.25, 1.5, 1.75 and 2 times the stress drop of finite-fault model. All of the other parameters are kept


(a)
proposed study
PSHA
Annual exceedance rate
PGA (g)
(b)
PSA (T=0.1 s)(g)
(c)
PSA (T=0.2 s)(g)
(d)
PSA (T=0.5 s)(g)
(e)
PSA (T=0.7 s)(g)
(f)
PSA (T=1.0 s)(g)
(g)
PSA (T=1.5 s)(g)
(h)
PSA (T=2.0 s)(g)
(i)
2% in 50 years (proposed study)
2% in 50 years (PSHA)
10% in 50 years (proposed study)
10% in 50 years (PSHA)
50% in 50 years (proposed study)
50% in 50 years (PSHA)
PSA (g)
T (s)


(a)
proposed study
PSHA
Annual exceedance rate
PGA (g)
(b)
PSA (T=0.1 s) (g)
(c)
PSA (T=0.2 s) (g)
(d)
PSA (T=0.5 s) (g)
(e)
PSA (T=0.7 s) (g)
(f)
PSA (T=1.0 s) (g)
(g)
PSA (T=1.5 s) (g)
(h)
PSA (T=2.0 s) (g)
(i)
2% in 50 years (proposed study)
2% in 50 years (PSHA)
10% in 50 years (proposed study)
10% in 50 years (PSHA)
50% in 50 years (proposed study)
50% in 50 years (PSHA)
PSA (g)
T (s)

◀**Fig. 8** Hazard curves for **a** PGA, **b** PSA ($T = 0.1$ s), **c** PSA ($T = 0.2$ s), **d** PSA ($T = 0.5$ s), **e** PSA ($T = 0.7$ s), **f** PSA ($T = 1.0$ s), **g** PSA ($T = 1.5$ s), **h** PSA ($T = 2.0$ s) and **i** UHS of proposed study and classical PSHA for Site 3

constant. Figure 12 shows misfit and sensitivity indices of PSA and FAS from point source simulations with different stress drop ratios with respect to finite-fault models. Stress drop of SMSIM is assumed to be 1.5 times the corresponding value for extended fault as defined in Eq. (6), based on the observations in Fig. 12 as well as previous discussions of Atkinson et al. (2009) and Moghaddam et al. (2010).

Effect of near-field forward directivity

In this section, effect of forward directivity model on the hazard functions is investigated for Site 2. Site 2 is selected in this section since it is anticipated to be exposed to potential rupture directivity effects of NAFZ more than two other sites. The main reason for this selection is the location of Site 2 with respect to NAFZ and right-lateral strike–slip mechanism of NAFZ (Fig. 2). The analytical formulation of Mavroeidis and Papageorgiou (2003) is used to characterize pulse-like motion related to near-field scenarios. EXSIM program already has this option and the procedures are thoroughly described by Motazedian and Atkinson (2005). The hazard scenarios that are affected by directivity pulses involve rupture distances less than 15 km with epicenters located to the right of Site 2 as NAFZ is a right-lateral strike–slip fault zone. Mavroeidis and Papageorgiou (2003) proposed the following mathematical expression for the acceleration time history of the long-period pulse:

$$a(t) = \begin{cases} -\dfrac{A\pi f_p}{\gamma}\left[\begin{array}{c} \sin\left(\dfrac{2\pi f_p}{\gamma}(t-t_0)\right)\times\cos\left(2\pi f_p(t-t_0)+\nu\right) \\ +\gamma\sin(2\pi f_p(t-t_0)+\nu)\times\left[1+\cos\left(\dfrac{2\pi f_p}{\gamma}(t-t_0)\right)\right] \end{array}\right] & t_0-\dfrac{\gamma}{2f_p}\le t\le t_0+\dfrac{\gamma}{2f_p}, \\ 0 \quad \text{otherwise} \end{cases} \tag{11}$$

where A, f_p, γ, ν and t_0 are pulse amplitude, pulse frequency, oscillatory character, phase angle and time shift to specify time history peak, respectively. These are the basic input parameters for modeling near-field pulse which are determined by fitting previously recorded pulse-like ground motions to analytical function in Eq. (11). 13 March 1992 Erzincan earthquake, ERC station recording displays impulsive behavior mostly in the North–South component (Askan et al. 2013) hence it is used for the calibration of parameters herein.

The parameter t_0 is estimated such that peak of pulse and actual time history occur at the same time. Pulse amplitude, A, is determined in order to make the peak ground velocity (PGV) and peak pseudo-velocity spectrum (PSv) values of analytical pulse consistent with the corresponding values of recorded ground motion. Pulse period, $T_p = 1/f_p$, is calibrated so that peak values of PSv from analytical pulse and recorded ground motions occur at the same frequency. Finally, velocity and displacement time histories are attempted to be fitted in order to estimate γ and ν parameters. It should be noted that the parameters are determined through a simultaneous trial and error process. The calibration method is described extensively in Mavroeidis and Papageorgiou (2003). The final optimized parameters are demonstrated in Table 4. Figure 13 shows PSv, velocity and displacement time histories using the input parameters listed in Table 4.

Figure 14 exhibits Fourier amplitude spectrum (FAS) and pseudo-acceleration spectrum (PSa) of observed and synthetic ground motion with and without forward directivity. It is observed that, synthetic ground motion agrees better with the observed one for low-frequency region, after implementing forward directivity parameters.

Among the parameters stated in Table 4, A and T_p are varied for different scenarios. It is not possible to estimate these two parameters using the aforementioned trial and error procedures for each scenario due to lack of recorded data. However, as it is highlighted by several authors, pulse period is dependent on magnitude (Somerville 1998; Mavroeidis and Papageorgiou 2003; Shahi and Baker 2011). Mavroeidis and Papageorgiou (2003) expressed a predictive model for pulse period as follows:

$$\log T_p = -2.9 + 0.5M_w. \tag{12}$$

This relationship is already implemented in EXSIM. Besides, authors of the same study pointed to a strong relationship between A and PGV. Hence attenuation models for ground motions with near-field effect can be applied to predict PGV and as a result to predict A. In this

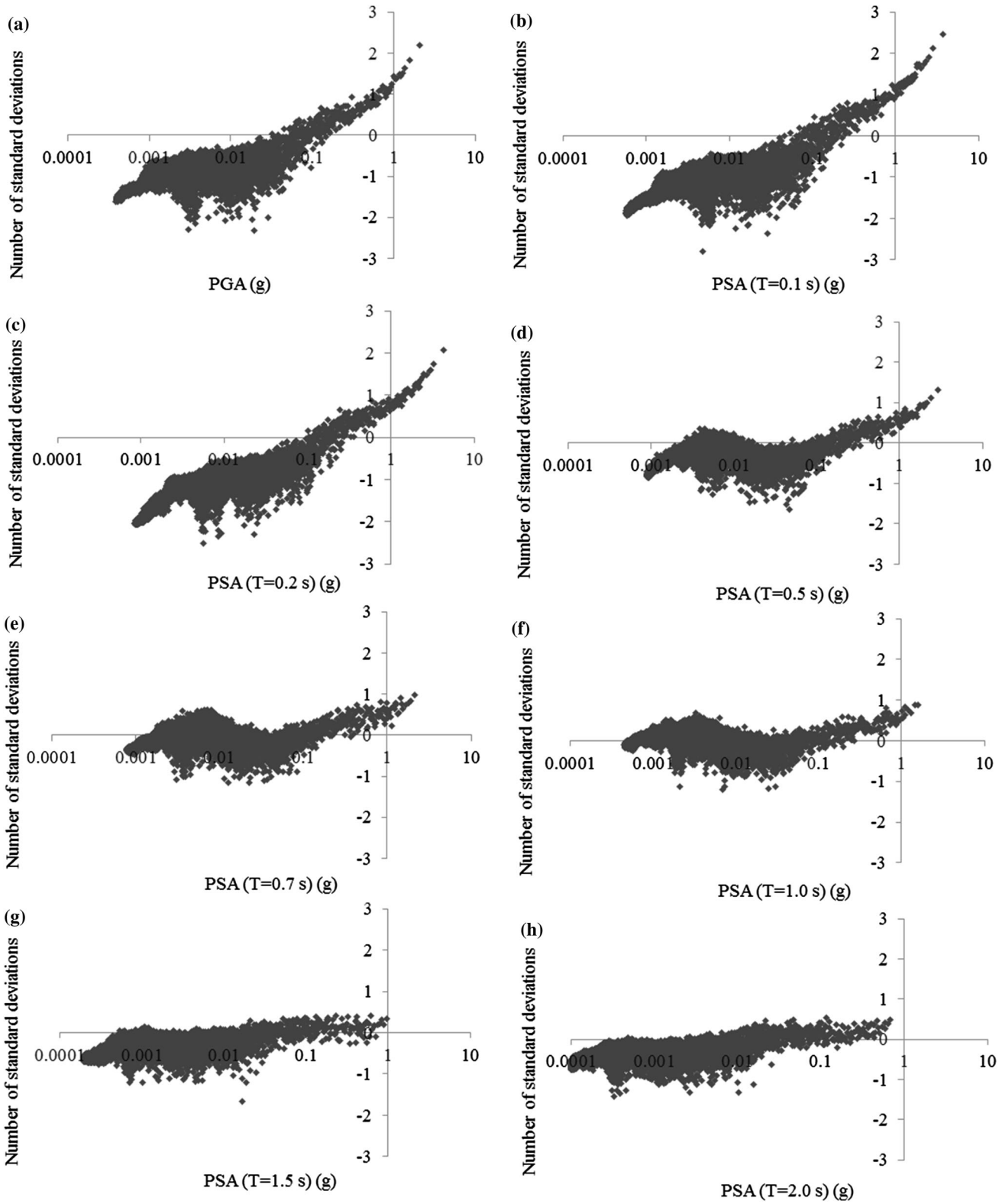


Fig. 9 Number of standard deviation of simulated GMIPs from the corresponding median values (Akkar and Bommer 2010) versus simulated GMIPs for **a** PGA, **b** PSA ($T = 0.1$ s), **c** PSA ($T = 0.2$ s), **d** PSA ($T = 0.5$ s), **e** PSA ($T = 0.7$ s), **f** PSA ($T = 1.0$ s), **g** PSA ($T = 1.5$ s) and **h** PSA ($T = 2.0$ s) for Site 1

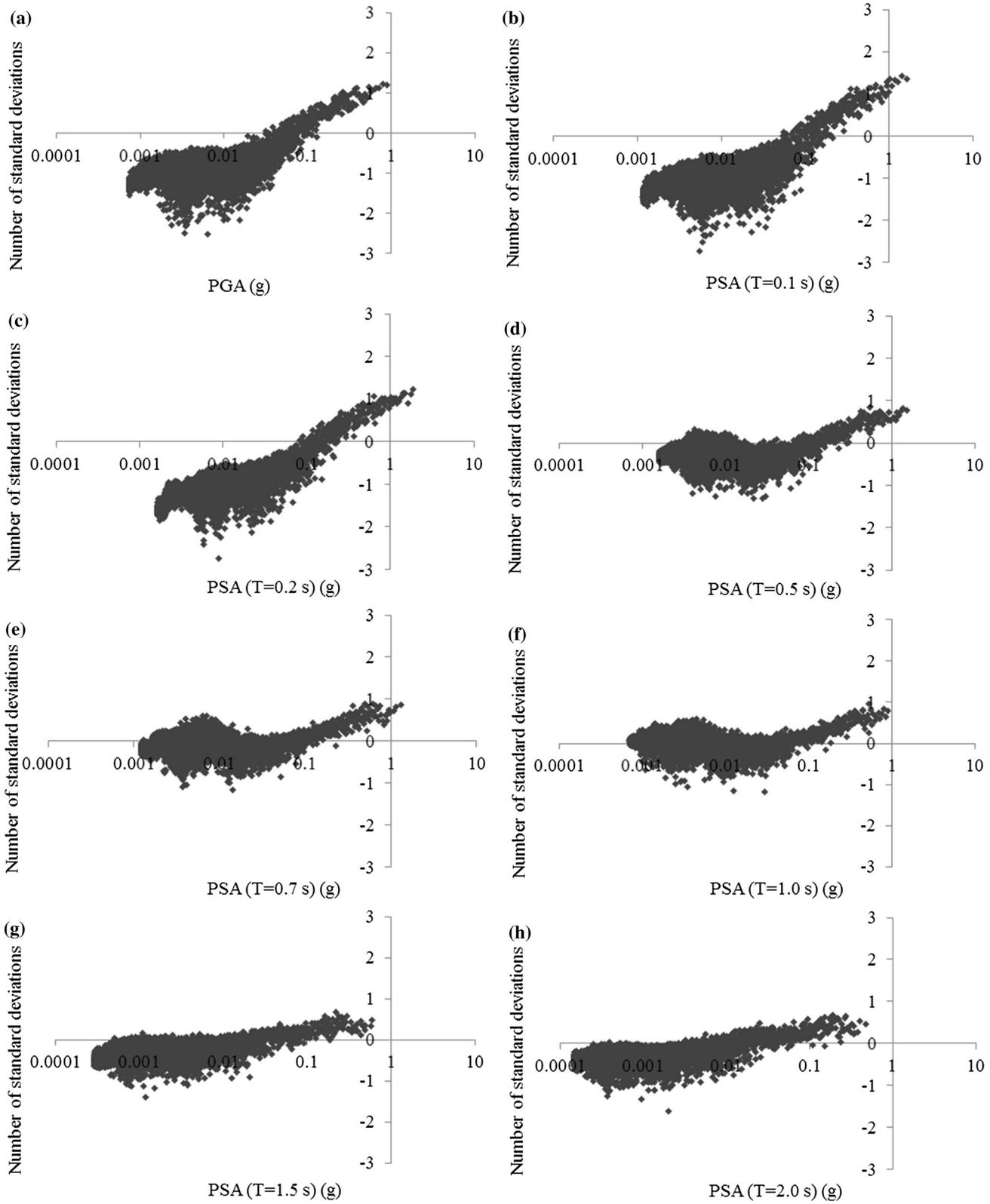


Fig. 10 Number of standard deviation of simulated GMIPs from the corresponding median values (Akkar and Bommer 2010) versus simulated GMIPs for **a** PGA, **b** PSA ($T = 0.1$ s), **c** PSA ($T = 0.2$ s), **d** PSA ($T = 0.5$ s), **e** PSA ($T = 0.7$ s), **f** PSA ($T = 1.0$ s), **g** PSA ($T = 1.5$ s) and **h** PSA ($T = 2.0$ s) for Site 2

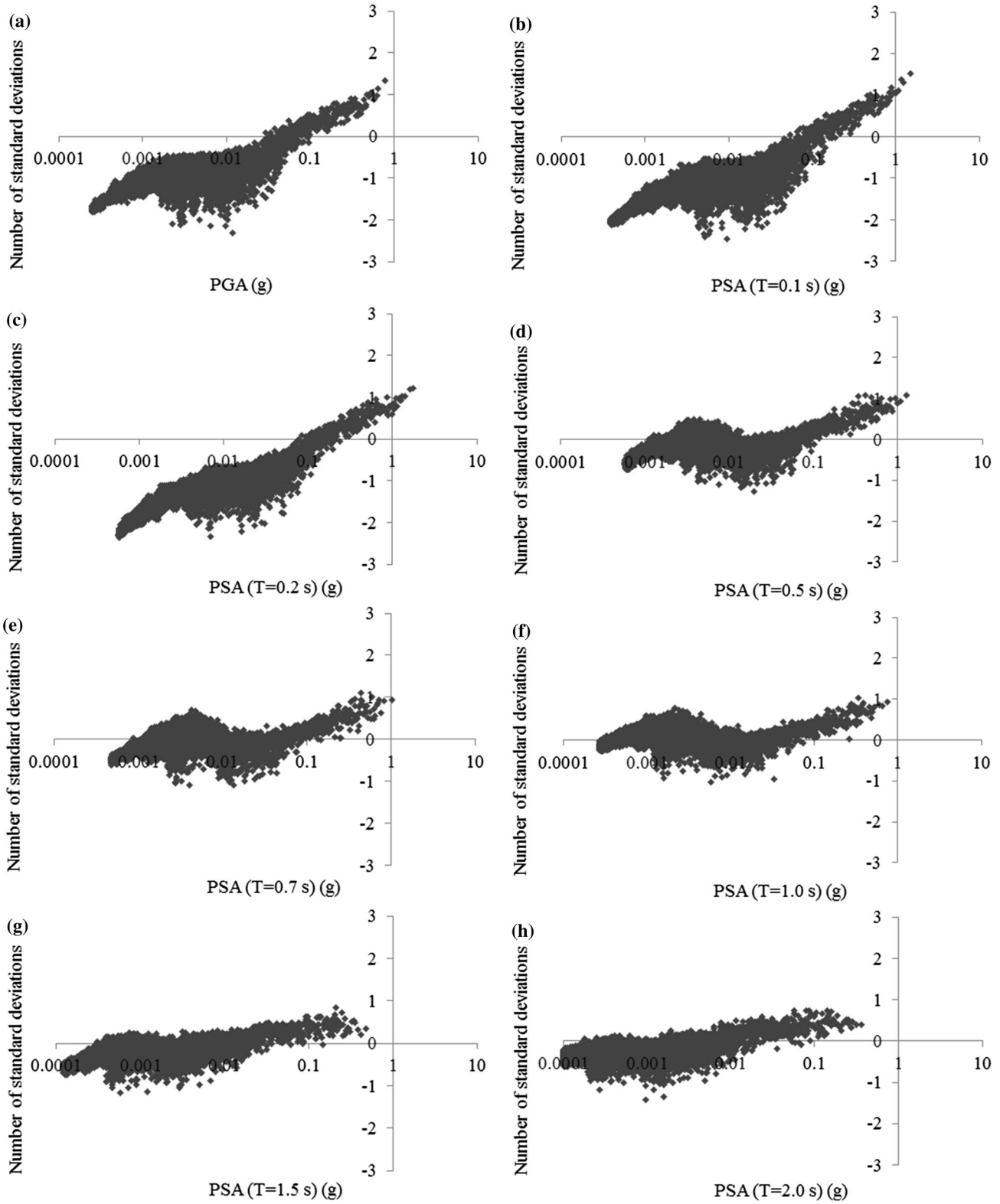


Fig. 11 Number of standard deviation of simulated GMIPs from the corresponding median values (Akkar and Bommer 2010) versus simulated GMIPs for **a** PGA, **b** PSA ($T = 0.1$ s), **c** PSA ($T = 0.2$ s), **d** PSA ($T = 0.5$ s), **e** PSA ($T = 0.7$ s), **f** PSA ($T = 1.0$ s), **g** PSA ($T = 1.5$ s) and **h** PSA ($T = 2.0$ s) for Site 3

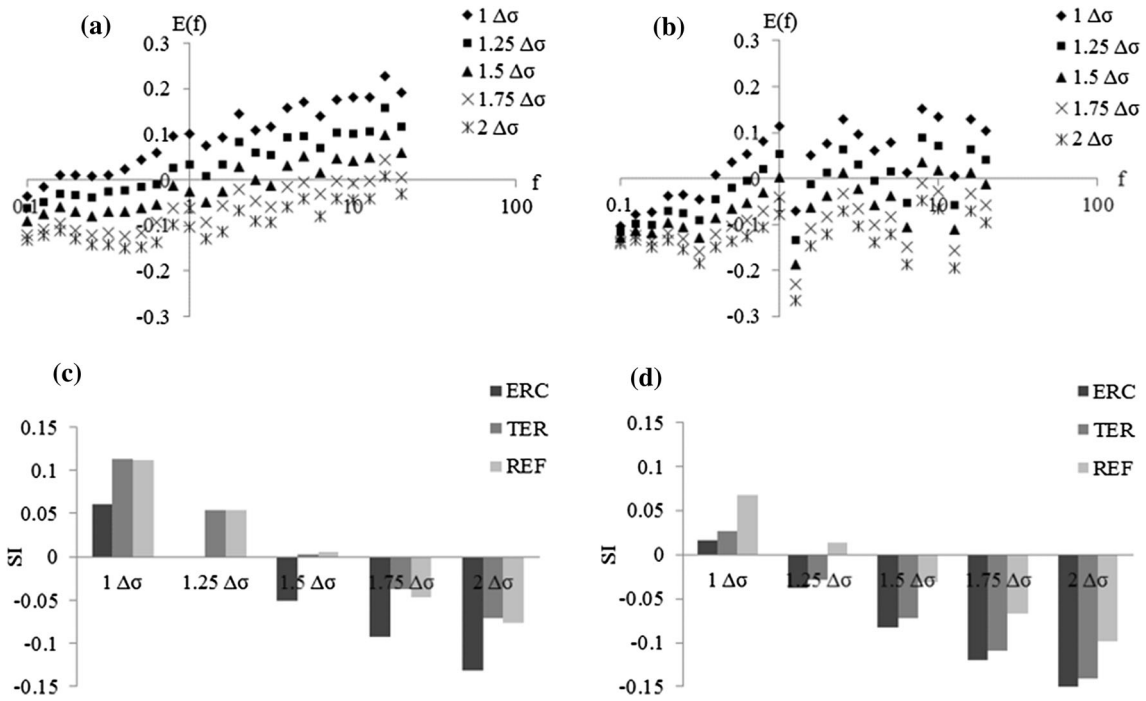


Fig. 12 Misfit (E(f)) related to **a** PSa, **b** FAS, and sensitivity index (SI) related to **c** PSa, **d** FAS for point source simulations with different stress drop values with respect to finite-fault model

Table 4 Input parameters for Mavroeidis and Papageorgiou (2003) analytical pulse model

Parameter	Value
A	80 cm/s
T_p	2 s
γ	2.3
ν	180 degrees

study, PGV is estimated from the predictive model of Bray and Rodriguez-Marek (2004) for near-field ground motions given as:

$$\ln \mathrm{PGV} = 4.58 + 0.34M_w - 0.58\mathrm{Ln}(R^2 + 7^2). \quad (13)$$

Equation (13) is applicable to M_w range of 6–8 and closest distance range of 0–15 km. Bray and Rodriguez-Marek (2004) studied a large database for the regression analysis. Finally, this predictive model yields an agreeable estimation of median PGV as 71 cm/s at ERC station of the 13 March 1992 Erzincan earthquake, which is along with the calibrated pulse amplitude in Table 4.

Figure 15 shows uniform hazard spectra for Site 2 with and without considering a forward directivity model. It is obvious from Fig. 15 that near-field forward directivity pulses tend to increase long-period PSAs for 2 and 10% exceedance probabilities in 50 years. The results from two approaches display differences that start from 1 s period and increase monotonically with period. The forward directivity analytical pulse which is applied for ground motions with magnitudes larger than $M_w = 6$ since Eq. (13), which is used to estimate pulse amplitude, is valid for magnitude range of 6–8. Because there is lack of recorded ground motions with pulse-like behavior and magnitude smaller than 6, the minimum pulse period is 1.25 s according to Eq. (12) therefore this analytical pulse model is associated with long periods. The difference is larger for hazard spectra with 2475-year return period compared to that with 475-year one. The increase in spectral amplitude at $T = 2$ s for 2% in 50 years hazard level due to forward directivity is 32% which is to some extent consistent with 30% amplification for $T = 3$ s and 1500 years given by Abrahamson (2000). In addition, this percentage agrees well with amplification ratios calculated by Shahi and Baker (2011), which varied between 1.1 and 1.4.

Response spectrum, which does not involve duration effects, is known to be inadequate to characterize directivity (Somerville 1998). Yet, duration is directly involved with directivity effects such as pulse-like behavior. Besides, duration plays an important role in identifying structural response due to nonlinear degradation (Bolt 1973; Novikova and Trifunac 1994). Empirical relations were developed to predict duration based on ground motion parameters such as magnitude, distance and site class (e.g., Novikova and Trifunac 1994; Bommer and Martinez-Pereira 1999; Kempton and Stewart 2006). These studies presented the possibility of implementing duration in PSHA, but none of them calculated duration directly from ground motion time history.

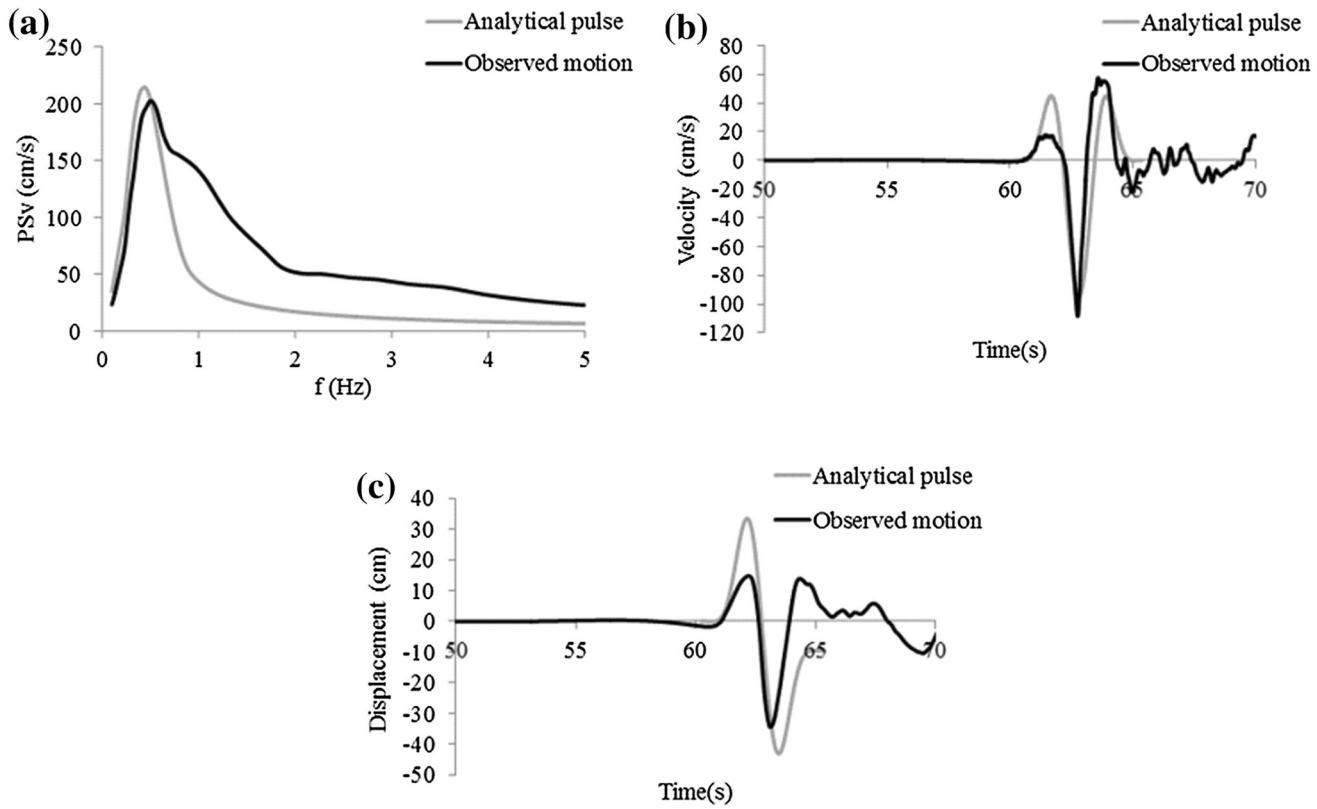


Fig. 13 Calibration of parameters in Table 4 with respect to 13 March 1992 earthquake (ERC recording) in terms of **a** PSv, **b** velocity time series and **c** displacement time series

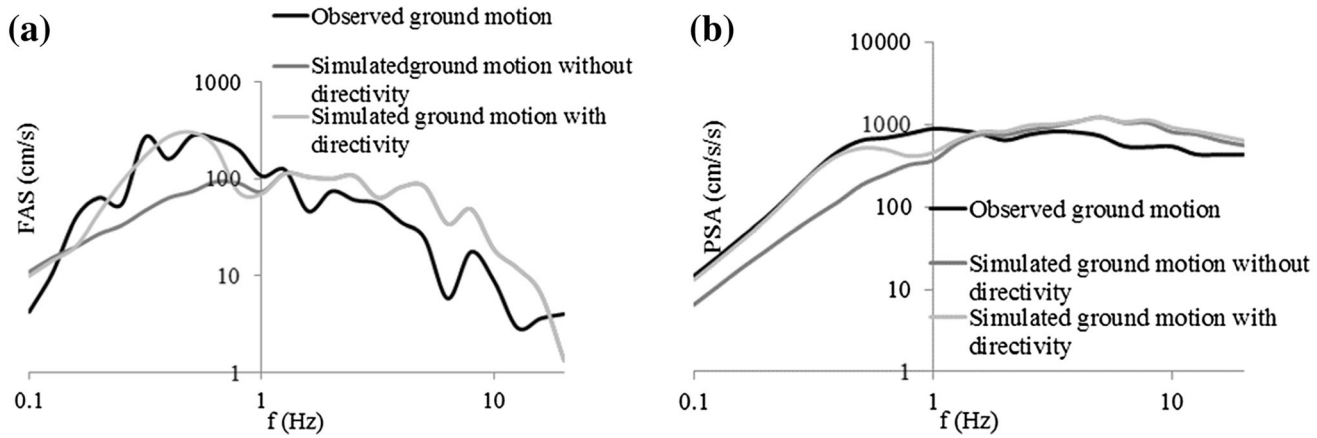


Fig. 14 Observed and simulated ground motions of 13 March 1992 earthquake (ERC recording) with and without considering forward directivity in terms of **a** FAS and **b** PSa considering the parameters in Table 4

In this study, significant duration is calculated via Arias intensity from the simulated time histories (Arias 1970):

$$I_{\mathrm{A}} = (\pi/2g)\int_0^{t_{\mathrm{d}}} a^2(t)\mathrm{d}t. \tag{14}$$

Effective duration, defined by Trifunac and Brady (1975), is computed for all scenarios as the time span between 5 and 95% of maximum Arias intensity. Then the scenarios are divided into two categories: ground motions with effective duration less than 10 s (i.e., short duration) and above (i.e., long duration). The hazard curves are disaggregated for 2 and 10% exceedance probability in 50 years according to these two groups. Contribution ratios of ground motions with high duration with respect to these two hazard levels are illustrated in Fig. 16 with and without taking rupture directivity into account.

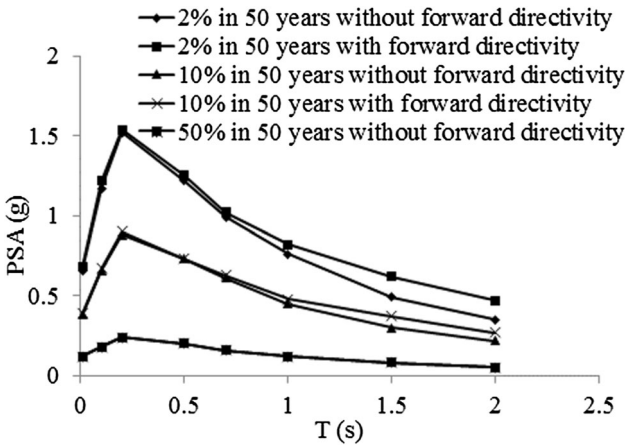


Fig. 15 Uniform hazard spectra (UHS) for Site 2 with and without considering forward directivity

Figure 16 shows that, ground motions with long duration are dominant for long-period spectral ordinates without considering near-field effects. Contribution of long duration is approximately 80% for long-period PSAs and both hazard levels, whereas for short-period ground motion amplitudes, this percentage reduces to about 40 and 60% for 2475 and 475 years, respectively. There is also monotonic and gradual growth in contribution ratio with respect to period for 475-year return period.

Forward directivity model leads to a lower contribution ratio of long-duration ground motion for periods larger than $T = 1$ s. There is even less than 20% contribution for 2475 and 40% for 475 return periods in the long-period range. The reason is that, long-period pulses cause ground motion to release the majority of energy in a short duration. Ground motions with short duration contribute less to 475-year return period, since this hazard level is less affected by rupture directivity. The same conclusion was derived in terms of UHS in Fig. 15. However, considering the entire period range and both return periods, taking near-fault effects into account leads to smaller ground motion durations. The hazard in terms of PGV as well exhibits less contribution of long duration when adding directivity pulse but not as small as long-period response. However, its amplitude increases at longer periods ($T = 1.5$ and 2 s) after considering the near-field forward directivity pulse.

Effect of site response

In the previous sections, generic site amplification based on local V_{s30} measurements was used to derive site amplification. This methodology was proposed by Boore and Joyner (1997) applying quarter-wavelength approximation (Joyner et al. 1981). The technique does not produce peaks and troughs in amplification function since detailed seismic wave propagation within geotechnical layers is neglected. Although it provides an agreeable estimate of average response, it can underestimate the response of models with velocity gradients (Boore 2013).

In this section, UHS of Site 1 and 3 are recalculated via theoretical site amplification based on transfer function. One-dimensional (1D) site response analysis is required for soil profiles to develop these transfer functions. The soil profile is idealized to be composed of infinite horizontal layers located on uniform half space. In this study, SHAKE software (Schnabel et al. 1972) is used to compute frequency-dependent site amplification factors based on local soil conditions. Velocity profiles and the geotechnical input parameters of soil layers were obtained via detailed field observations at nine sites in Erzincan (Askan 2015). Two bedrock motions are used as input ground motions in the program with PGAs of 0.002 and 0.4 g. These motions are called as small and large input motions, respectively, from this point onward. The closest points to Site 1 and 3 are selected and ground response is computed by making use of the small and large input motions at bedrock level. Figure 17 represents 1D velocity profiles of the closest

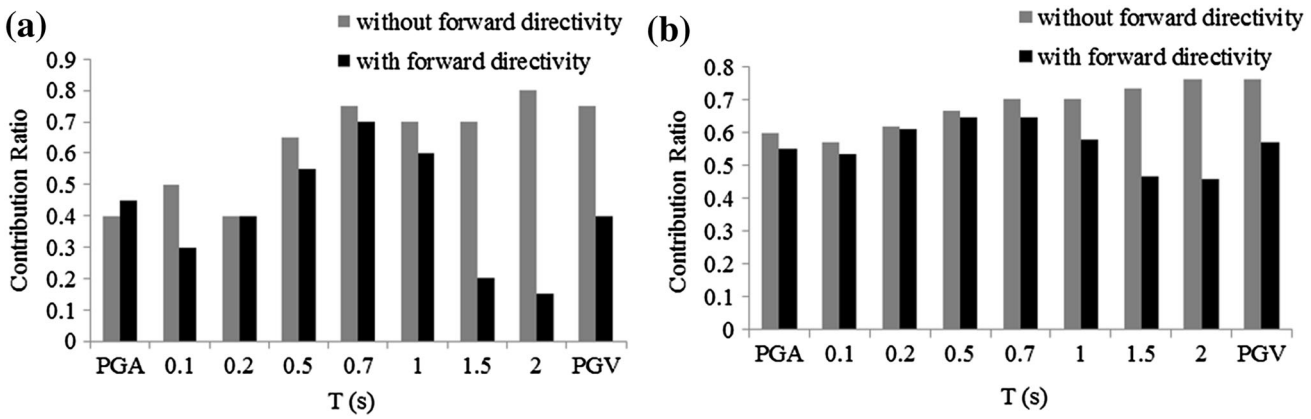


Fig. 16 Contribution of long-duration ground motions in **a** 2% in 50 years and **b** 10% in 50 years hazard levels for Site 2 with and without considering forward directivity

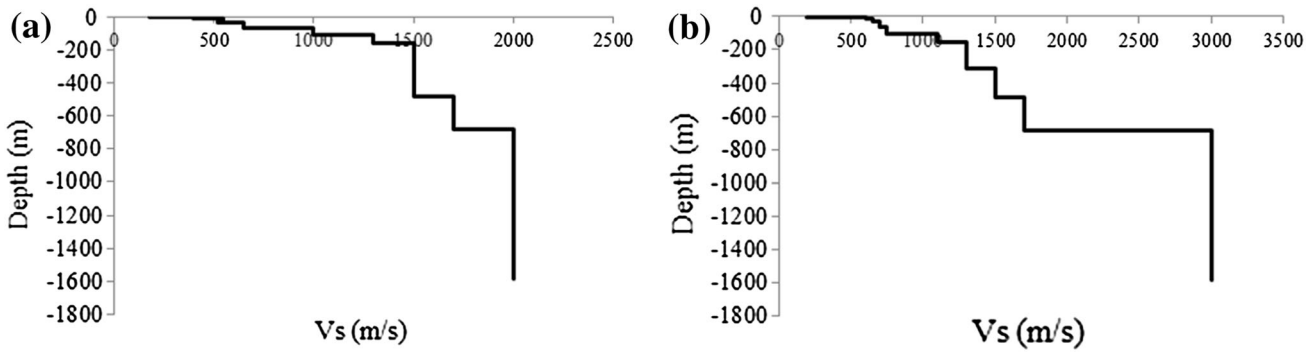


Fig. 17 1D shear wave velocity profile of **a** Site 1 and **b** Site 3 (Adopted from Askan 2015)

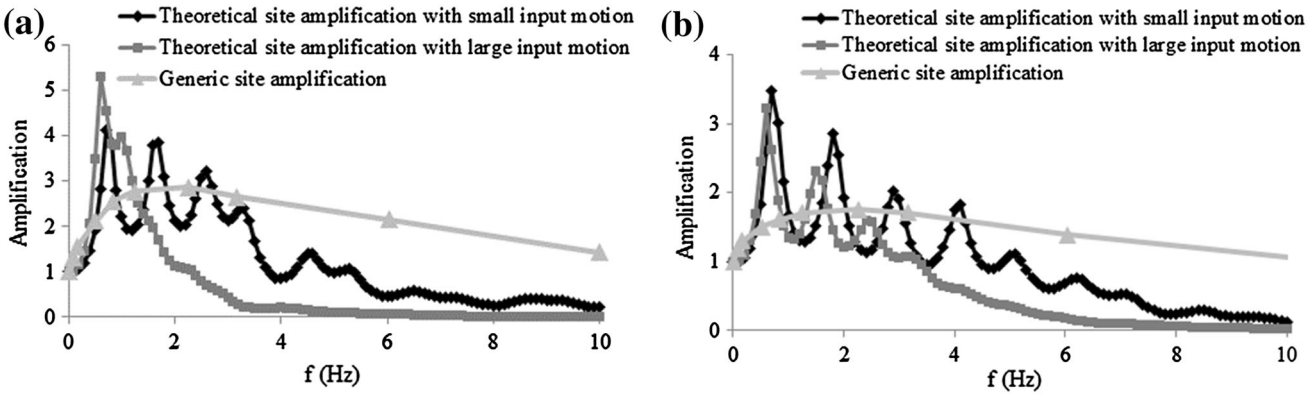


Fig. 18 Theoretical and generic attenuated site amplification factor as a function of frequency for **a** Site 1 and **b** Site 3

points to Site 1 and 3, which are used in this study. Figure 18 shows theoretical amplification functions and the generic ones at Site 1 and 3. High-frequency kappa factor is as well applied in Fig. 18 on the amplification factors.

As expected, generic amplification function is a smoothed form of theoretical transfer function corresponding to small input motion. Transfer function displays completely different trend when the large rock motion is employed. It exhibits deamplification for larger frequencies which is addressed as soil nonlinearity in literature (e.g., Beresnev et al. 1998; Khaheshi Banan et al. 2012). This phenomenon was included in recent NGA attenuation models as well (e.g., Abrahamson and Silva 2008). The definition of soil nonlinearity is that, soil amplification becomes less significant for larger ground motion amplitudes and smaller periods.

It is also observed in Fig. 18 that, site deamplification regarding Site 3 begins at frequencies larger than Site 1 since Site 3 is stiffer. Two approaches are followed in this section for calculating hazard spectra to see the effects of soil nonlinearity in hazard calculations. First, theoretical site response resulting from small input motion accounts for the whole scenarios (i.e., soil nonlinearity is disregarded). Then, theoretical site response resulting from large and small motion are implemented for scenarios with median PGA larger and smaller than 0.1 g, respectively (i.e., soil nonlinearity is regarded). The proposed UHS using generic and theoretical site amplifications with and without soil nonlinearity effects are presented in Fig. 19 at Site 1 and 3.

According to Fig. 19, theoretical transfer function leads to larger ground motion amplitudes than generic site response for larger periods (between 0.2 and 2 s) while disregarding soil nonlinearity at Site 1. The difference between the two methods becomes larger for periods around 0.5 and 1.5 s since there are obvious peaks in theoretical amplification function around these periods. Design spectrum of Turkish Seismic Code (TSC 2007) corresponding to the same site conditions is also illustrated in Fig. 19 along with UHS. It is observed that the shape of UHS considering theoretical site response is more similar to design spectrum than that of generic site amplification. Code-based design spectrum is derived from UHS of 10% exceedance probability in 50 years. It is noted that the UHS related to 10% probability of exceedance in 50 years and theoretical site amplification is about 50% larger than the design spectrum.

Theoretical response with nonlinearity considerably underestimates the spectral ordinates for periods less than

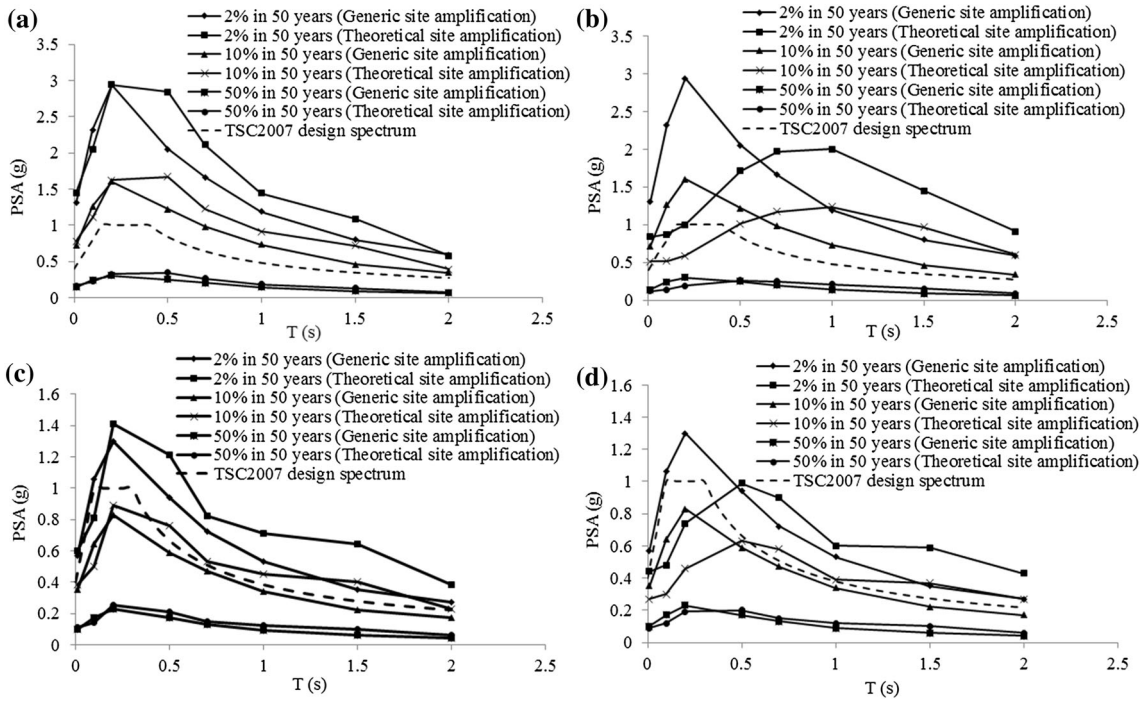


Fig. 19 Uniform hazard spectra (UHS) for generic and theoretical site amplification related to Site 1: **a** without soil nonlinearity effects, **b** with soil nonlinearity effects. Uniform hazard spectra (UHS) for generic and theoretical site amplification related to Site 3: **c** without soil nonlinearity effects, **d** with soil nonlinearity effects

0.6 s, while it overestimates them for periods above 0.6 s as compared to generic site response. The vast underestimation for smaller periods is the result of deamplification due to soil nonlinearity which was discussed previously. The shape of UHS from theoretical site amplification with soil nonlinearity is completely different from the code-based spectrum. The underestimation of theoretical site response for short periods seems to be more obvious for lower hazard probabilities. This finding augments the concept of soil nonlinearity which is more significant for larger ground motion amplitudes.

The general behavior of the UHS at Site 3 is similar to those of Site 1. Theoretical site response with small input motion gives amplified GMIPs as compared to generic site amplification, for periods of 0.2–2 s more obviously for $T = 0.5$ and 1.5 s. Code-based design spectrum and UHS related to 475 years and theoretical site amplification at Site 3 are not as similar as the ones at Site 1 in terms of shape, but the spectral amplitudes are more comparable at Site 3.

Again the general trend for two sites is similar while considering soil nonlinearity. The underestimation of theoretical site response with respect to generic site response at Site 3 is smaller and it terminates at a smaller period than that of Site 1. The main reason is that soil nonlinearity of Site 3 is not as dominant as that of Site 1. Site 3 yields more complex spectra than Site 1 after implementing theoretical site amplification due to the impedance differences among the soil layers (Fig. 17).

The previous observations also highlight the significance of site amplification function in seismic hazard assessments and seismic design approaches. The spectral ordinates of TSC-based design spectrum remain constant for different site classes which require further modifications.

Conclusions

In this study, an alternative approach to PSHA is proposed to estimate regional seismic hazard using simulated ground motions and an MCS-based stochastic earthquake catalog. Erzincan region in eastern Turkey is selected as the case study area. The results obtained at three selected sites in Erzincan are compared with those from traditional PSHA to see the differences in between. The effects of near-field forward directivity and alternative site amplification functions are also investigated in this study. The main observations in this study are as follows:

- Classical PSHA generally yields larger spectral ordinates than the proposed study most probably due to the wide aleatory variability involved (considering −3 and +3 standard deviations). There is an exception regarding Site 1 that the proposed study yields slightly larger response spectra for low exceedance rates and periods

less than 0.5 s. This observation may be due to major events at this site as it is in the close vicinity of North Anatolian Fault Zone and located on soft soil. As a common finding for all three sites, classical PSHA produces considerably larger results than the proposed study for 2 and 10% exceedance probability levels in 50 years for periods larger than $T = 0.2$ s.

- Near-field forward directivity pulse model increases the spectral ordinates related to return periods of 2475 and 475 years for natural periods of larger than 1 s. Effective duration of ground motion regarding the aforementioned natural periods and return periods is also involved such that contribution of long-duration ground motions decreases considerably in particular for 2475 years.
- Theoretical site amplification without soil nonlinearity, increases ground motion amplitudes for larger periods as compared to those from generic site amplification. It also improves the shape of UHS with 475-year return period regarding Site 1 in terms of similarity to TSC-based design spectrum, but the spectral ordinates are not comparable. The values of spectral ordinates related to UHS and design spectrum are more consistent at Site 3. Theoretical site amplification with soil nonlinearity, however, overestimates and underestimates ground motion amplitudes for larger and smaller periods, respectively, as compared to generic site amplification. The underestimation in low-period range is considerable for both sites, especially for Site 1 as it is softer than Site 3.

UHS from the proposed methodology is derived from individual records of scenario earthquakes, whereas classical UHS is the envelope of multiple earthquakes and may produce unphysically overestimated spectral ordinates. The proposed methodology also has the advantage of considering regional and local seismicity parameters such as path effects (including anelastic attenuation, geometric spreading, and duration) as well as high-frequency decay (in the form of the kappa factor) which is not the case in classical PSHA. Another advantage of this study is the development of a complete earthquake catalog consistent with the regional seismicity as the recorded catalogs are generally incomplete both temporally and spatially.

The proposed study in this paper also facilitates implementation of detailed site response inside probabilistic hazard studies. It is difficult, however, to account for such site response models via traditional PSHA as the site parameters are usually coarsely included in ground motion prediction models. Besides, near-field forward directivity is inserted into the proposed seismic hazard assessment methodology in a straight forward manner.

Finally, implementation of the anticipated ground motions in the region via detailed source and site-related parameters at selected sites shows that classical PSHA could be improved to include local seismic demand. Indeed in areas of sparse networks, scarce data and incomplete catalogs, the proposed approach could be employed effectively. Results of this study and similar studies can be useful for realistic seismic damage and loss estimations in the future since seismic demands of structures are defined based on local seismological parameters.

Data and resources

The ground motion recordings are acquired from Turkish National Strong Motion Network using http://daphne.deprem.gov.tr (last accessed on April 2014). EXSIM and SMSIM computer programs are downloaded from www.carleton.ca/~dariush and http://www.daveboore.com/software_online.html (last accessed on December 2014).

Acknowledgements Aida Azari Sisi was a graduate student fellowship recipient of TUBITAK-2215 Program. We are grateful for this support. We also thank Fatma Nurten Sisman Dersan for providing theoretical site amplification factors. We are grateful to David Boore and Dariush Motazedian for providing SMSIM and EXSIM as well as sharing their valuable comments during our initial applications in the past.

References

Abrahamson NA (2000) Effects of rupture directivity on probabilistic seismic hazard analysis. In: Proc. Sixth International Conference on Seismic Zonation: Managing Earthquake Risk in the 21st Century, Palm Springs, CA

Abrahamson NA, Silva W (2008) Summary of the Abrahamson & Silva NGA ground-motion relations. Earthq Spectra 24(1):67–97

Akansel VH, Ameri G, Askan A, Caner A, Erdil B, Kale Ö, Okuyucu D (2014) The 23 October 2011 Mw 7.0 Van (Eastern Turkey) earthquake: characteristics of recorded strong ground motions and post earthquake condition assessment of infrastructure and cultural heritage. Earthq Spectra 30(2):657–682

Akkar S, Bommer J (2010) Empirical equations for the prediction of PGA, PGV and spectral accelerations in Europe, the Mediterranean region and the Middle East. Seismol Res Lett 81(2):195–206. doi:10.1785/gssrl.81.2.195

Ambraseys NN, Douglas J, Sarma SK, Smit PM (2005) Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: horizontal peak ground acceleration and spectral acceleration. Bull Earthq Eng 3:1–53. doi:10.1007/s10518-005-0183-0

Anderson J, Hough SE (1984) A model for the shape of the fourier amplitude spectrum of acceleration at high frequencies. Bull Seismol Soc Am 74(5):1969–1993

Arias A (1970) A Measure of Earthquake Intensity. In: Hansen R (ed) Seismic Design for Nuclear Power Plants. MIT press, Cambridge, pp 438–483

Askan A (2015) Estimation of Potential Seismic Damage in Erzincan, Project TUJJB-UDP-01-12

Askan A, Sisman FN, Ugurhan B (2013) Stochastic strong ground motion simulations in sparsely-monitored regions: A validation and sensitivity study on the 13 March 1992 Erzincan (Turkey) earthquake. Soil Dyn Earthq Eng 55:170–181

Assatourians K, Atkinson G (2013) EqHaz: An open-source probabilistic seismic-hazard code based on the Monte Carlo simulation approach. Seismol Res Lett 84(3):516–524. doi:10.1785/0220120102

Atkinson G, Assatourians K, Boore MD, Campbell KW, Motazedian D (2009) A guide to differences between stochastic point-source and stochastic finite-fault simulations. Bull Seismol Soc Am 99(6):3192–3201. doi:10.1785/0120090058

Banan KK, Kolaj M, Motazedian D, Sivathayalan S, Hunter JA, Crow HL, Pugin AJ, Brooks GR, Pyne M (2012) Seismic site response analysis for Ottawa, Canada: A comprehensive study using measurements and numerical simulations. Bull Seismol Soc Am 102(5):1976–1993. doi:10.1785/0120110248

Beresnev I, Atkinson GM (1997) Modeling finite-fault radiation from the ωn spectrum. Bull Seismol Soc Am 87:67–84

Beresnev IA, Atkinson GM, Johnson PA, Field EH (1998) Stochastic finite-fault modeling of ground motions from the 1994 Northridge, California, earthquake. II. Widespread nonlinear response at soil sites. Bull Seismol Soc Am 88(6):1402–1410

Bolt BA (1973) Duration of strong motion. In: Proc. 5th World Conference of Earthquake Engineering, Rome, Italy

Bommer J, Crowley H (2006) The influence of ground-motion variability in earthquake loss modelling. Bull Earthq Eng 4:231–248. doi:10.1007/s10518-006-9008-z

Bommer J, Martinez-Pereira A (1999) The effective duration of earthquake strong motion. J Earthq Eng 3(2):127–172. doi:10.1080/13632469909350343

Boore MD (2003) Simulation of ground motion using the stochastic method. Pure Appl Geophys 160(3–4):635–676

Boore MD (2009) Comparing stochastic point-source and finite-source ground-motion simulations: SMSIM and EXSIM. Bull Seismol Soc Am 99(6):3202–3216. doi:10.1785/0120090056

Boore MD (2013) The uses and limitations of the square-root-impedance method for computing site amplification. Bull Seismol Soc Am 103(4):2356–2368. doi:10.1785/0120120283

Boore MD, Joyner WB (1997) Site amplifications for generic rock sites. Bull Seismol Soc Am 87(2):327–341

Bray J, Rodriguez-Marek A (2004) Characterization of forward-directivity ground motions in the near-fault region. Soil Dyn Earthq Eng 24:815–828. doi:10.1016/j.soildyn.2004.05.001

Cao T, Petersen MD, Cramer CH, Toppozada TR, Reichle MS, Davis JF (1999) The calculation of expected loss using probabilistic seismic hazard. Bull Seismol Soc Am 89(4):867–876

Chopra S, Kumar D, Choudhury P, Yadav RBS (2012) Stochastic finite fault modelling of Mw 4.8 earthquake in Kachchh, Gujarat, India. J Seismol 16:435–449. doi:10.1007/s10950-012-9280-0

Cornell CA (1968) Engineering seismic risk analysis. Bull Seismol Soc Am 58(5):1583–1606

Cramer CH (2006) Quantifying the uncertainty in site amplification modeling and its effects on site-specific seismic-hazard estimation in the Upper Mississippi embayment and adjacent areas. Bull Seismol Soc Am 96(6):2008–2020. doi:10.1785/0120060037

Crowley H, Bommer J (2006) Modelling seismic hazard in earthquake loss models with spatially distributed exposure. Bull Earthq Eng 4(3):249–273. doi:10.1007/s10518-006-9009-y

Demartinos K, Faccioli E (2012) Probabilistic seismic performance assessment of classes of buildings using physics-based simulations and ground-motion prediction equations. J Earthq Eng 16(1):40–60

Deniz A (2006) Estimation of Earthquake Insurance Premium Rates Based on Stochastic Methods, Master of Science, Middle East Technical University

Eads L, Miranda E, Krawinkler H, Lignos DG (2013) An efficient method for estimating the collapse risk of structures in seismic regions. Earthq Eng Struct Dyn 42:25–41. doi:10.1002/eqe.2191

Ellingwood BR, Celik OC, Kinali K (2007) Fragility assessment of building structural systems in mid-America. Earthq Eng Struct Dyn 36:1935–1952. doi:10.1002/eqe.693

Frankel A (1993) Three-dimensional simulations of the ground motions in the San Bernardino Valley, California, for hypothetical earthquakes on the San Andreas fault. Bull Seismol Soc Am 83:1020–1041

Graves RW, Pitarka A (2010) Broadband ground-motion simulation using a hybrid approach. Bull Seismol Soc Am 100(5A):2095–2123. doi:10.1785/0120100057

Gu P, Wen YK (2007) A record-based method for the generation of tridirectional uniform hazard-response spectra and ground motions using the Hilbert-Huang transform. Bull Seismol Soc Am 97(5):1539–1556

Hashash Y, Moon S (2011) Site amplification factors for deep deposits and their application in seismic hazard analysis for central US, Under USGS/NEHRP Grant: G09AP00123, University of Illinois at Urbana-Champaign

Hung TV, Kiyomiya O (2013) Source Parameter Estimation and Stochastic Ground Motion Simulation Based on Recorded Accelerograms in Northwestern Vietnam. J Earthq Eng 17(3):304–322

Joyner WB, Warrick RE, Fumal TE (1981) The effect of Quaternary alluvium on strong ground motion in the Coyote Lake, California, earthquake of 1979. Bull Seismol Soc Am 71:1333–1349

Kempton J, Stewart JP (2006) Prediction equations for significant duration of earthquake ground motions considering site and near-source effects. Earthq Spectra 22(4):985–1013

Kramer SL (1996) Geotechnical earthquake engineering. Prentice Hall Inc, Upper Saddle River

Luco N, Ellingwood BR, Hamburger RO, Hooper JD, Kimball JK, Kircher CA (2007) Risk-targeted versus current seismic design maps for the conterminous United States. In: Proc. SEAOC 2007 Convention Proceedings, Squaw Creek, California

Mavroeidis GP, Papageorgiou AS (2003) A mathematical representation of near-fault ground motions. Bull Seismol Soc Am 93(3):1099–1131

McGuire RK (2004) Seismic hazard and risk analysis. MNO-10, Earthquake Engineering Research Institute

McGuire RK, Arabasz WJ (1990) An introduction to probabilistic seismic hazard analysis. Soc Explor Geophys 1:333

Moghaddam H, Fanaie N, Motazedian D (2010) Estimation of stress drop for some large shallow earthquakes using stochastic point source and finite fault modeling. Sci Iran 17(3):217–235

Mohammadioun B, Serva L (2001) Stress drop, slip type, earthquake magnitude, and seismic hazard. Bull Seismol Soc Am 91(4):604–707

Motazedian D, Atkinson GM (2005) Stochastic finite-fault modeling based on a dynamic corner frequency. Bull Seismol Soc Am 95(3):995–1010. doi:10.1785/0120030207

Musson RMW (2000) The use of Monte Carlo simulations for seismic hazard assessment in the UK. Anadali Di Geofisica 43(1):1–9

Naeim F, Lew M (1995) On the use of design spectrum compatible time histories. Earthq Spectra 11(1):111–127

Nicknam A, Abbasnia R, Bozorgnasab M, Eslamian Y (2010) Synthesizing Broadband Time-Histories at Near Source Sites; Case Study, 2003 Bam Mw6.5 Earthquake. J Earthq Eng 14(6):898–917

Novikova EI, Trifunac MD (1994) Duration of strong ground motion in terms of earthquake magnitude, epicentral distance, site condition and site geometry. Earthq Eng Struct Dyn 23:1023–1043

Olsen KB, Archuleta RJ, Matarese JR (1996) Three-dimensional simulation of a magnitude 7.75 earthquake on the San Andreas fault. Science 270:1628–1632

Papoulia J, Fahjan YM, Hutchings L, Novikova T (2015) PSHA for Broad-Band Strong Ground-Motion Hazards in the Saronikos Gulf, Greece, from Potential Earthquake with Synthetic Ground Motions. J Earthq Eng 19(4):624–648

Raschke M (2014) Insufficient Statistical Model Development of Ground-Motion Relations for Regions with Low Seismicity. Bull Seismol Soc Am 104(2):1002–1005. doi:10.1785/0120130215

Roumelioti Z, Kiratzi A, Theodulidis N (2004) Stochastic strong ground-motion simulation of the 7 September 1999 Athens (Greece) earthquake. Bull Seismol Soc Am 94(3):1036–1052

Schnabel PB, Lysmer J, Seed HB (1972) SHAKE: a computer program for earthquake response analysis of horizontally layered sites. Earthquake Engineering Research Center, University of California, Berkeley, p 102

Shahi SK, Baker JW (2011) An empirically calibrated framework for including the effects of near-fault directivity in probabilistic seismic hazard analysis. Bull Seismol Soc Am 101(2):742–755. doi:10.1785/0120100090

Somerville PG (1998) Development of an improved ground motion representation for near fault ground motions. In: Proc. SMIP98 Seminar on Utilization of Strong-Motion Data, Oakland, CA

Tahghighi H (2012) Simulation of Strong Ground Motion Using the Stochastic Method: Application and Validation for Near-Fault Region. J Earthq Eng 16(8):1230–1247

Thenhaus PC, Campbell KW (2003) Seismic hazard analysis. In: Chan WF, Scawthorn C (eds) Earthquake engineering handbook. CRC Press, Boca Raton

Trifunac MD, Brady AG (1975) A study on the duration of strong earthquake ground motion. Bulletinof Seismological Society of America 65(3):581–626

Ugurhan B, Askan Gundogan A (2010) Stochastic strong ground motion simulation of the 12 November 1999 Düzce (Turkey) earthquake using a dynamic corner. Bull Seismol Soc Am 100(4):1498–1512. doi:10.1785/0120090358

Ugurhan B, Askan A, Akinci A, Malagnini L (2012) Strong-ground-motion simulation of the 6 April 2009 L'Aquila, Italy, earthquake. Bull Seismol Soc Am 102(4):1429–1445. doi:10.1785/0120110060

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

Wen YK, Wu CL (2001) Uniform hazard ground motions for mid-America cities. Earthq Spectr 17(2):359–384

Wu CL, Wen YK (2000) Earthquake Ground Motion Simulation and Reliability Implications, 630, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

Yalcinkaya E (2005) Stochastic finite-fault modeling of ground motions from the June 27, 1998 Adana-Ceyhan earthquake. Earth Planets Space 57:107–115

Zafarani H, Noorzad A, Ansari A, Bargi K (2009) Stochastic modeling of Iranian earthquakes and estimation of ground motion for future earthquakes in Greater Tehran. Soil Dyn Earthq Eng 29:722–741

Acta Geophys. (2017) 65:331–344
DOI 10.1007/s11600-017-0028-7

RESEARCH ARTICLE

Energy-based analysis of permanent strain behaviour of cohesive soil under cyclic loading

Wojciech Sas[1] · Andrzej Głuchowski[1] · Bartłomiej Bursa[2] · Alojzy Szymański[2]

Received: 30 December 2016 / Accepted: 10 March 2017 / Published online: 1 April 2017

Abstract In this paper the original results of uniaxial cyclic compression test on cohesive soil are presented. The shakedown phenomena in cohesive soil are described. Energy-based method highlights the change of soil material behaviour from plastic shakedown through plastic creep shakedown to incremental collapse. The samples were cyclically loaded under undrained conditions with the constant amplitude of stress in one-way test procedure. In this study the energy-based method was presented as a proper method to categorise response of cohesive soil to cyclic loading in uniaxial conditions. A shakedown criterion factor, S_E, was introduced to help understand the shakedown phenomena in cohesive soil. In cohesive soils the absence of a limit between plastic shakedown and plastic creep shakedown was pointed out.

Keywords Cohesive soil · Shakedown · Energy · Cyclic loading · Uniaxial compression · Soil mechanics

✉ Andrzej Głuchowski
andrzej_gluchowski@sggw.pl

Wojciech Sas
wojciech_sas@sggw.pl

Bartłomiej Bursa
bartlomiej_bursa@sggw.pl

Alojzy Szymański
alojzy_szymanski@sggw.pl

[1] Water Centre Laboratory, Faculty of Civil and Environmental Engineering, Warsaw University of Life Sciences – SGGW, Warsaw, Poland

[2] Department of Geotechnical Engineering, Faculty of Civil and Environmental Engineering, Warsaw University of Life Sciences – SGGW, Warsaw, Poland

Introduction

The behaviour of the soil under cyclic loading has been studied by many researches recently (Kokusho and Kaneko 2014; Feng et al. 2015; Cai et al. 2015; Sas et al. 2014, 2015). The most extensive studies were focused on liquefaction phenomena because of the danger that can be caused by its appearance (Kokusho and Kaneko 2014; Kokusho et al. 2012). One of the causes of liquefaction phenomena occurrence is the presence of high frequent cyclic loading which can be forced by traffic, earthquake or machinery vibrations. In opposite to this high frequent loading there also exists the slow quasi-static repeated loading excitated by, for instance, soil mass movements (there is no danger of liquefaction) (Zhou and Gong 2001). When it comes to cohesive soil the liquefaction phenomena does not exist. There is still a little knowledge about the cohesive soil behaviour under quasi-static cyclic loading. Quasi-static loading is characterised by time effect negligibility. In geotechnical field of studies such loading appears when its frequency is less than 2–5 Hz (Danne and Hettler 2015; Wichtmann 2005). The quasi-static phenomena occurs when harmonic excitation applied on a specimen causes displacement: $u = u^{ampl} cos(\omega t)$ with acceleration $\ddot{u} = -u^{ampl}\omega^2 \cos(\omega t)$, where $u^{ampl}\omega^2 << g$ (Danne and Hettler 2015).

Quasi-static loads are often encountered in industrial foundations, technical roads or local motorways. In the road engineering the effect of the cyclic loading can be observed as the rutting, and in foundation engineering it exists in the form of foundation settlement (Cuéllar et al. 2014; Kokkali et al. 2014; Soares et al. 2014). In the case of quasi-static loading the damage is not caused by the bearing capacity lost, but due to the plastic strain accumulation.

Among many types of soil, cohesive soil is merely examined by the researches; one of the reasons is that this type of soil is not often used as the bearing material. Nevertheless, the cohesive soil should be investigated because even the low stress level can cause high deformation which leads to weaken the above bearing layers. Therefore, engineers seek for the new design procedures which take into account the cyclic loading phenomena in cohesive soil (Soares et al. 2014; Lu et al. 2014).

Shakedown theory in soil mechanics is derived from the observations of the soil response to cyclic loading in various range of stresses (Goldscheider 1978). Shakedown theory has been used to explain the behaviour of engineering structures loaded with repetitive force. The pressure vessels under thermal cyclic loading problem were the first application of this concept, and later it was applied to the rolling on metal surfaces problem (König and Maier 1981). When it comes to geomechanics the shakedown theory was used in the road structures (Werkmeister et al. 2001; Werkmeister 2006; Tao et al. 2010; Nega et al. 2015; Sharp and Booker 1984; Boulbibane and Collins 2015).

The fundamental concept of shakedown theory is a division of the behaviour of the soil into five categories (König and Maier 1981):

- Purely elastic, where load is small enough to cause only the elastic strains. No plastic strains occur and response to cyclic loading is purely elastic (0).
- Elastic shakedown, where repeated loading causes plastic strains in the first cycles. After this phase, no further plastic strains occur and the material behave purely elastic. The maximum stress level when this phenomena occurs is called elastic shakedown limit (1).
- Plastic shakedown, where plastic strains occur, but after a few cycles material achieves steady hysteretic response. Nevertheless, the hardly noticeable small plastic strains can be observed. The maximum stress level when this phenomena occurs is called the plastic shakedown limit (2).
- Plastic creep shakedown, where after a few cycles the material hardens and the plastic strains occur (3).
- Incremental collapse, where the stress level causes accumulation of extensive plastic strains. The plastic strains in this stage cause cracks and the material degradation (4).

The shakedown concept typically defines the appropriate limit stress to prevent from excessive plastic strain (Fadaee et al. 2008). Settlement of subgrade soils is adverse because it mostly leads to road damage. The excessive settlement of subgrade soils is typically caused by the accumulation effect from traffic load. The repeated stress and strain will not completely dissipate in the unloading state and will accumulate and transfer along road structure to subgrade layer (Puppala 2009).

Conducted tests with shakedown application mostly concern unbound granular material (UGM) testing. The shakedown concept is utilised as a method of permanent strain behaviour analysis. The shakedown theory is now utilised in permanent strain evaluation for UGM and the procedure of shakedown analysis for this materials is standardised and presented in European Standards: EN 13286–7 (2004; Cerni et al. 2012). The shakedown concept in applications in transportation geotechnics mostly concerns cyclic compression (Soliman and Shalaby 2015). Shakedown concept was utilised for study of fine addition to UGM. The resistance of material to permanent deformation for UGM and non-cohesive soils was linked to shear strength of material. The slip between particles under traffic loading was recognised as most sensitive phenomenon which triggers permanent deformation occurrence (Soliman and Shalaby 2015).

Under cyclic loading the same response can be observed in cohesive as well as non-cohesive soil. A mechanism is similar (rolling and sliding of grains develop plastic strain), but in addition in cohesive soil the cohesion forces counteract the development of the plastic strains (Chen et al. 2015; Karg et al. 2010). Identification of this mechanism is very complicated and time-consuming, so the simplified methods are usually employed for analysis. One of the branches of the simplified methods is energy-based method which is utilised in many fields of applied mechanics, especially when it comes to cyclic and dynamic loading (Liang et al. 2015; Seo et al. 2015). This method is based on the first law of thermodynamics (assuming the negligibility of kinetic energy) and tells that all external forces transform into internal energy and are dissipated in the form of the plastic strains. Using energy-based method we were able to categorise a response of a material to one of the shakedown categories. In this paper a new proposition of the energy-based method application is proposed.

Studies under clayey subsoil traffic-load-influenced depths employed shakedown concept to distinguish three depths of cyclic traffic load influence: the threshold depth beyond which traffic loads becomes negligible, the plastic shakedown limit depth where subgrade experiences continuous deformation and the critical failure depth where subsoil will fail due to excessive strain (Tang et al. 2015).

The studies on effects of cyclic confining pressure on the deformation characteristics of natural soil have found that the shapes of all hysteresis loops looks similar in general. But with the number of cycles, the increment of residual axial strain caused by a single cycle is reducing gradually (Sun et al. 2015), which may be classified as plastic shakedown.

Other studies concerning low frequent cyclic loading, cyclic loading in consolidometer or in variable confining pressure mention the decrease on plastic strain rate during cyclic loading which again may be related to shakedown concept (Cai et al. 2013; Kalinowska and Jastrzębska 2014; Li et al. 2011; Gu et al. 2012). Therefore, simple criterion of distinguishing the permanent strain accumulation phenomena in cohesive soil is needed for proper classification of plastic strain development.

Materials and methods

Materials

In this study the cohesive soil material obtained from 2 m deep earthwork was analysed. The samples of this material were remoulded and tested physically and mechanically. Table 1 presents results of mechanical and physical properties of material used in this study.

Based on sieve test (PKN-CEN ISO/TS 17892-4:2009) with respect to Polish standards (PN-EN ISO 14688-2:2006), tested soil was recognised as sandy clay (saCl).

Soil material before main tests was prepared with respect to the Proctor method. This procedure leads to establish constant test conditions (constant moisture and maximum dry density). The undisturbed samples were first dried and ground to powder. The material was next compacted at optimum moisture content with respect to the Proctor energy of compaction. The data about optimum moisture and maximal dry density were gathered by previous Proctor's test with respect to (PN-EN 13286-2:2010/AC). The abovementioned conditions are as follows: optimum moisture content, wopt. 10.5%, volume density maximum dry density ρ_s 2.15 g/cm^3.

The compaction of the samples was performed by rescaling the Proctor hammer with diameter of 3.5 cm and mass equal to 500 g. The sample dimensions were as follows: 7 cm diameter and 14 cm high. The clay powder was mixed with water with respect to optimum moisture content.

Table 1 Physical and mechanical properties of sandy clay in this study

Properties	Symbol	Value
Specific density of soil	ρ_s (g cm^{-3})	2.66
Maximum dry density	ρ_d (g cm^{-3})	2.15
Natural moisture	w_n (%)	12.82
Optimal moisture	w_{opt} (%)	10.5
Liquid limit	w_l (%)	37.4
Plasticity limit	w_p (%)	12.3
Plasticity index	I_p (–)	25.1
Void ratio	e_0 (–)	0.41

Methods

The tests of uniaxial cyclic loading were performed with Instron's loading frame where axial stress and displacement were registered. One series of ten tests was performed. Table 2 presents the parameters of each test in constant frequency equal to 0.1 Hz.

Uniaxial cyclic loading tests were performed using one-way loading test method without reversion of its direction. A sample was placed on a base and a rigid cap was placed on a top. The sample was covered by a rubber membrane to prevent moisture loss during a test. In Fig. 1a a schema of the uniaxial cyclic test is presented.

The one-dimensional tests are widely used in field of geotechnical engineering where comparative or preliminary studies can highlight material properties or non-traditional materials have not been tested (Koseki et al. 2014). The soil samples with "A" index represent high amplitude of cyclic loading. The samples with "B" index denote collapsed specimens and "C" index presents samples loaded by repeating stress with small amplitude. The cyclic stress can be characterised by five parameters which are maximal stress σ_{max}, minimal stress σ_{min}, stress amplitude σ_a, stress median σ_m and stress difference $\Delta\sigma$. Those stress values can fully describe repeated loading of any material (Jain et al. 2015). Figure 1b presents schematic stress–time curve with highlighter above-mentioned stress values. Table 3 presents detailed stress values for each test.

Energy density calculations

The results of the tests were later analysed by calculating the area of plastic energy density (dE^P) and elastic energy density (dE^E). The calculations were performed by

Table 2 Parameters of uniaxial cyclic loading test for tested samples

Sample	σ_{max} (kPa)	CSR (–)	No. of cycles
A.1	39.0	0.17	500
A.2	116.9	0.50	1000
A.3	142.9	0.61	2000
A.4	153.1	0.65	100
B.1	194.9	0.83	11
B.2	168.9	0.72	8
C.1	118.2	0.53	10,000
C.2	134.4	0.61	10,000
C.3	134.4	0.61	10,000
C.4	156.9	0.70	10,000

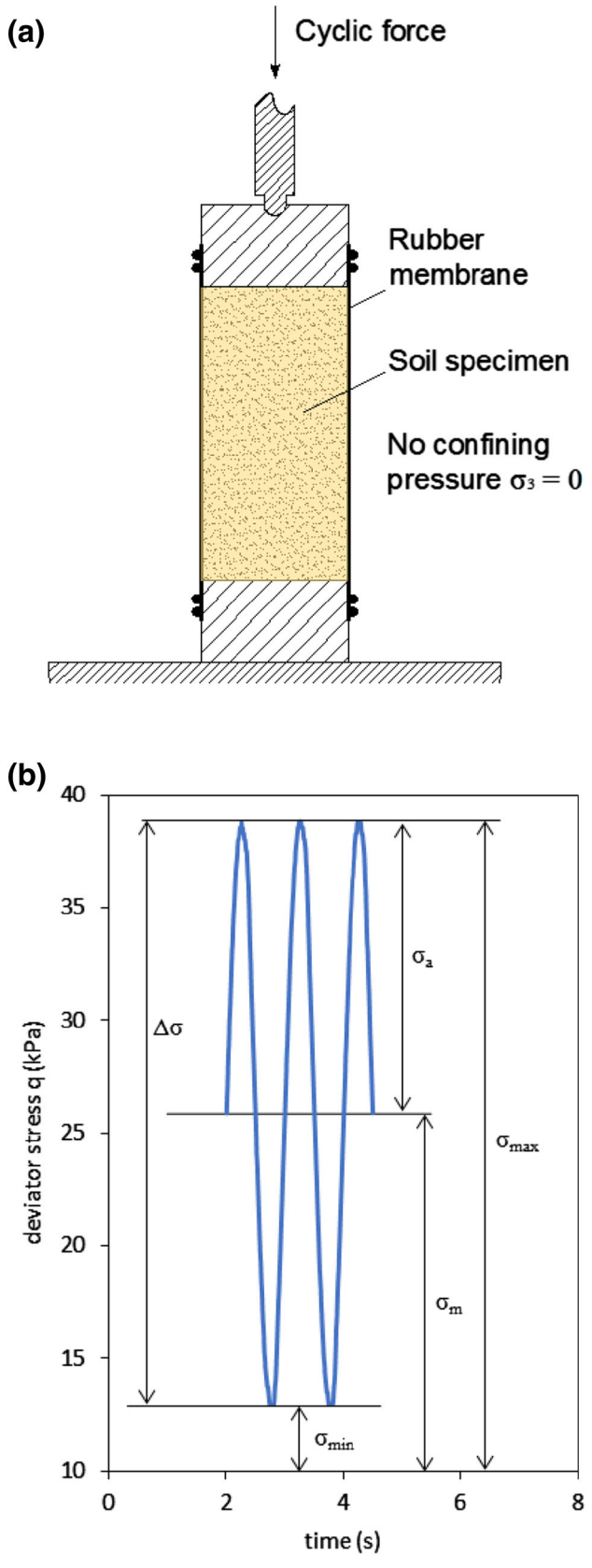


Fig. 1 The methods in this study: **a** schema of the uniaxial cyclic test, **b** schema of the stress characteristic values

Table 3 Stress characteristic values for uniaxial cyclic loading in this study

Sample	σ_{min}	σ_{max}	σ_a	σ_m	$\Delta\sigma$
A.1	13.0	38.8	12.9	25.9	25.8
A.2	13.2	65.4	26.1	39.3	52.2
A.3	16.5	145.9	64.7	81.2	129.5
A.4	16.7	150.1	66.7	83.4	133.4
B.1	16.5	194.9	89.2	105.7	178.4
B.2	16.5	168.9	76.2	92.7	152.4
C.1	94.3	118.2	11.9	106.3	23.8
C.2	121.5	134.4	6.4	128.0	12.9
C.3	121.5	134.4	6.4	128.0	12.9
C.4	134.5	156.9	11.2	145.7	22.4

Mathematica software where the self-made algorithm of the area calculation was employed.

The energy concept has been utilised in many fields in theory of plasticity and elasticity. Many constitutive laws and energy principles are based on energy concept (Desai and Siriwardane 1984; Pasik et al. 2015; Panoskaltsis and Bahuguna 1996). The energy-based methods can be divided into two categories: stress methods and strain methods. In the stress methods, the amount of strain energy is calculated from recorded stress and strain data during cyclic uniaxial loading. A hysteresis loop can be plotted from stress–strain data (Daum 2008). Figure 2 presents a typical hysteresis loops obtained from the cyclic compression loading from the stress-controlled cyclic uniaxial test.

The plastic strain energy density is equal to the area inside of hysteresis loop, and elastic energy density is equal to area under unloading curve (Ostadan et al. 1996; Green et al. 2000). In this study the plastic strain accumulation and elastic strain evolution are changed to equivalent plastic and elastic energy density to study changes of aforementioned strains during cyclic compression test. Plastic strain accumulation would be not observed between two next cycles, which was previously confirmed (Karg et al. 2010). The plastic energy density is, therefore, more sensitive to changes of soil sample behaviour during repeated loading. In Fig. 2, dE^P decreases rapidly and dE^E decrease is steady. The decline value of dE^P is represented by area of hysteresis loop reduction. Reduction of dE^E value with number of cycles represents the reduction of inclination of the hysteresis loop towards strain axis. The phenomena of dE^E decreasing indicate that the stiffness of the specimens tends to decrease. The dE^E descent has also impact on plastic strain accumulation and needs to be taken into consideration of plastic strain development.

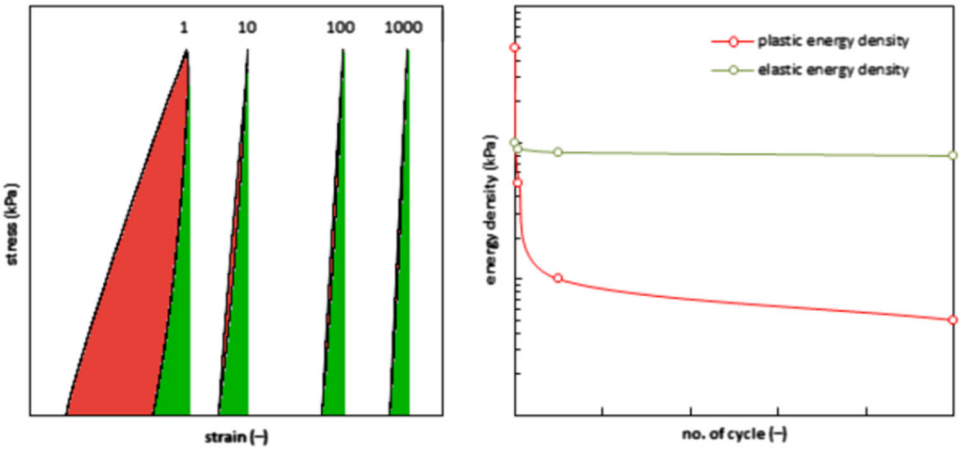


Fig. 2 Schema of the energy calculation for cyclic loading in this study

Results and discussion

Uniaxial cyclic test results

The results of the experiment were focused on the change of the energy densities during uniaxial cyclic compression loading. The results of uniaxial cyclic compression loading tests are presented in Figs. 3, 4 and 5. In Fig. 3 the results of the four tests from A.1 to A.4 are presented with stress–strain axis to study the shakedown concept. The plot presents hysteretic loops among different cycle numbers in test with the same stress conditions. Figures 3 and 5 show that with number of cycles the hysteresis loop inclination towards the X-axis tends to increase.

During cyclic loading, each sample was loaded to different amplitude of stress. For sample A.1 where axial of maximal stress σ_{max} was equal to 39.0 kPa plastic strain increment $\Delta\varepsilon_p$ after 500 cycles was equal to 0.0001034 (0.01034%).

For samples A.2, A.3 and A.4 the axial maximal stress was equal to 116.9, 142.9 and 153.1 kPa, respectively, and the plastic strain increment $\Delta\varepsilon_p$ was equal to 0.000416, 0.000401 and 0.000675, respectively.

Under higher maximal stress values (168.9 for B.2 and 194.9 kPa for B.1), the incremental collapse occurred. Standard uniaxial compression test has shown maximal strength of material σ_{max} 223.8 kPa which is 0.87 and 0.75 of maximum stress applied in tests B.1 and B.2, respectively (Fig. 4).

The samples B.1 and B.2 were destroyed during the tests. It was noticed that the last cycle before the collapsing was different from previous ones, and the crossing of the loading–unloading curves occurred in the lower stress level than in the previous cycles which may indicate the softening behaviour of the material.

The stress values characteristic parameters differed for samples "A" and "C". The amplitude of loading and stress median change results in another response of soil to repeated loading. For these samples, the decrease of axial stress was noted. The stiffness reduction was caused by characteristic stress conditions which differ from the test conditions in "A" phase of tests. This means that higher axial stress median is the higher stiffness reduction will occur (see Fig. 5).

In Fig. 6 a plot of the plastic strain increment during cyclic loading is presented, where various responses to cyclic loading can be distinguished. In a sample A.1 the plastic shakedown response to repeated loading can be noticed. In samples A.2 and A.3 at first the plastic shakedown response can be noticed as well, but after 100 repetitions the plastic strain tends to increase and, therefore, the plastic shakedown creep response occurs. The response of sample A.4 can be categorized as plastic creep shakedown. The characteristic of the curve from Fig. 4 is similar to the incremental collapse response, but the material was not destroyed. It may be caused by insufficient number of cycles that was programmed.

Samples B.1 and B.2 after a few cycles of loading collapsed due to accumulation of excessive plastic strains caused by strain softening. The response of the material can be categorised as incremental collapse.

Samples C.1 and C.3 follow the same pattern as samples A.2 and A.3. The plastic strain tends to stabilise and plastic shakedown occurs.

The plastic strain rate versus number of cycle for samples A.1–A.4 changes with logarithm of cycles. The decrease of plastic strain can be distinguished for samples from A.1 to A.4; nevertheless, the rate of this decrease in terms of shakedown criterion proposed by Werkmeister (2003) cannot be included in plastic shakedown response. The Werkmeister proposition consists of the accumulated permanent deformation analysis between 3000 and 5000 cycles. The proposition of such permanent strain accumulation analysis was established based on testing program of UGM (EN 13286-7:2004).

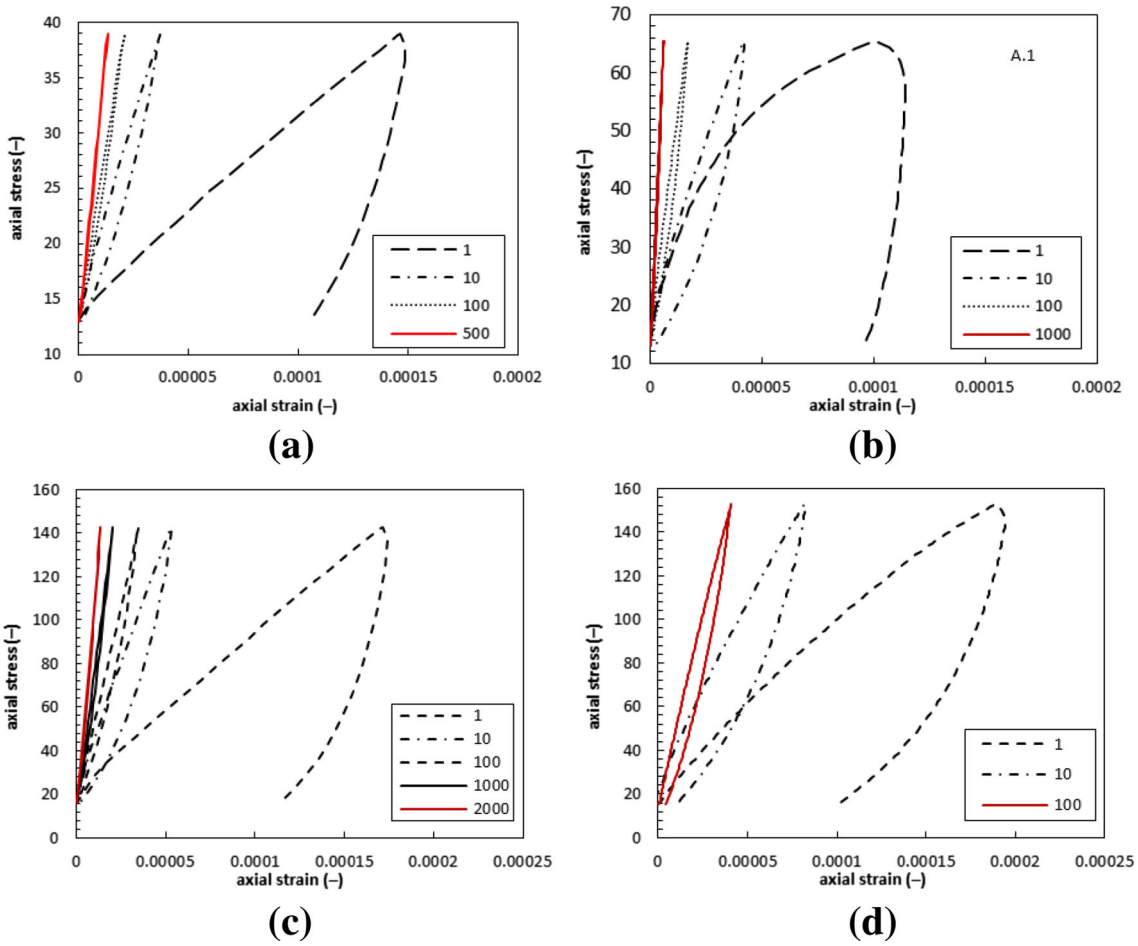


Fig. 3 Plot of the axial stress–strain for uniaxial cyclic loading of sandy clay in various stress levels—shakedown concept test (**a** A.1, **b** A.2, **c** A.3, **d** A.4), comparison of hysteretic loops among different cycle numbers

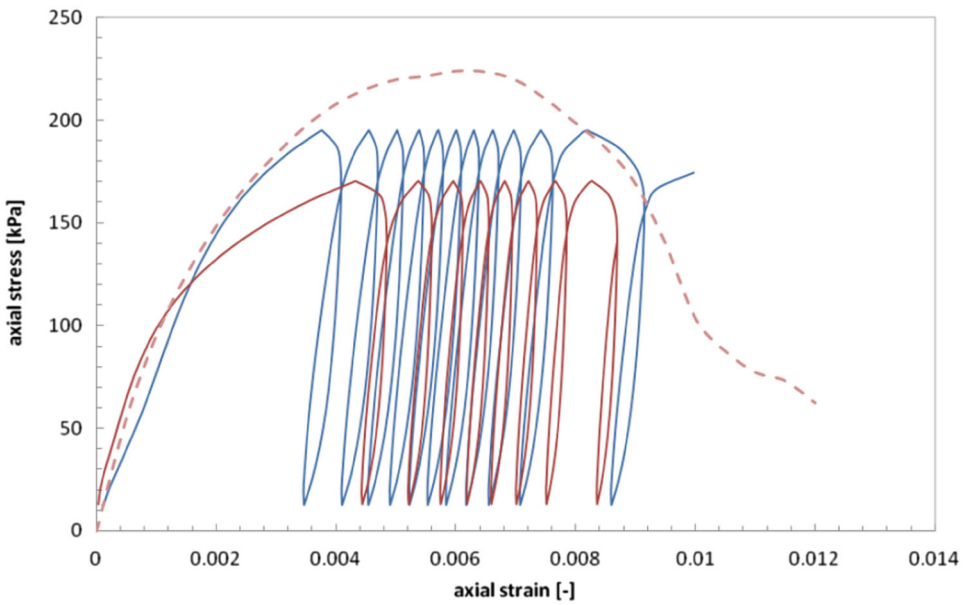


Fig. 4 Plot of the axial stress–strain for uniaxial cyclic loading of sandy clay in various stress levels—incremental collapse tests (*red* B.1, *blue* B.2, *dashed line* standard uniaxial compression test result)

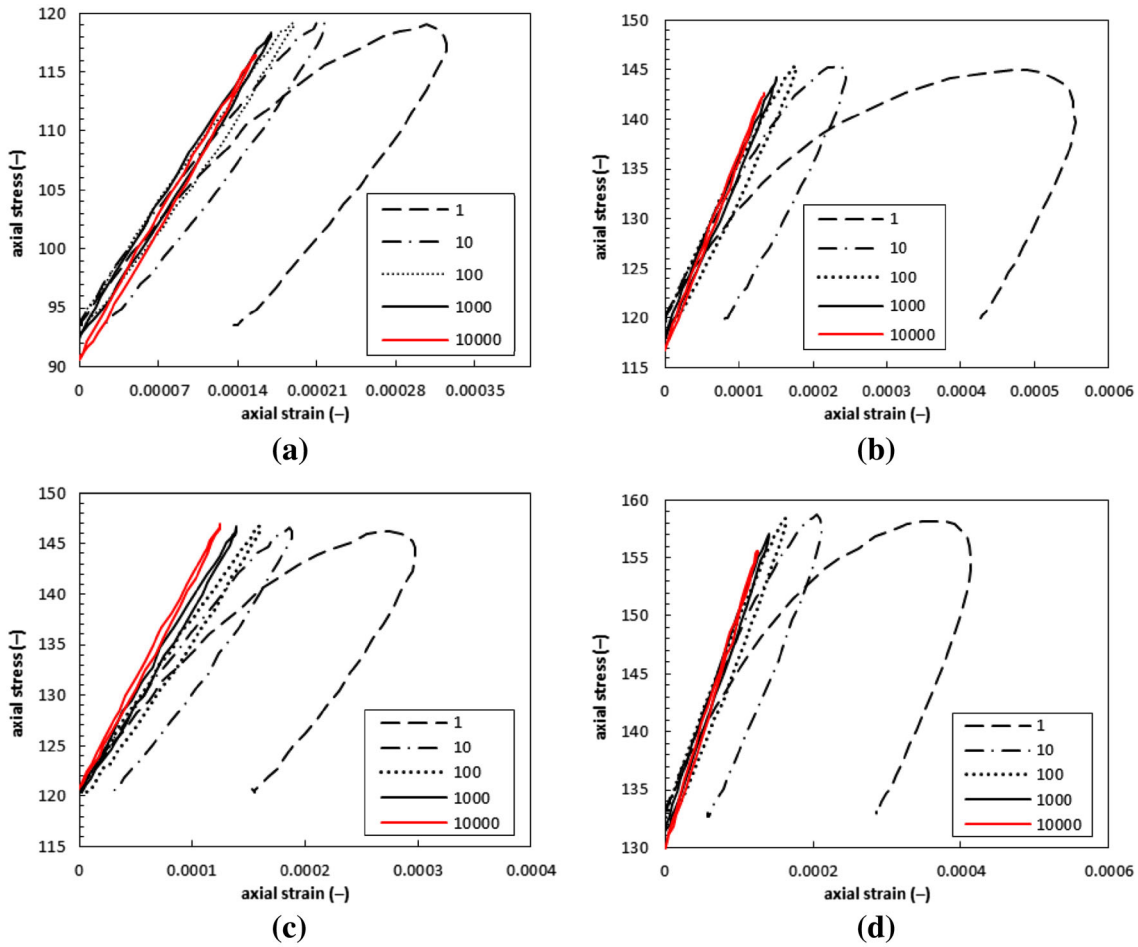


Fig. 5 Plot of the axial stress–strain for uniaxial cyclic loading of sandy clay in various stress levels—plastic creep concept (**a** C.1, **b** C.2, **c** C.3, **d** C.4), comparison of hysteretic loops among different cycle numbers

Aforementioned plastic shakedown limit is presented by Eq. (1):

$$\varepsilon_{p5000} - \varepsilon_{p3000} < 0.4 < 10^{-3} \quad (1)$$

If the difference between 5000 and 3000 cycle is less than $0.4 \cdot 10^{-3}$, then plastic shakedown occurs; if the difference crosses this limit, plastic shakedown creep occurs.

Plastic strain rate analysis presented in Fig. 6 for "C" indexed samples leads to estimate the shakedown response based on the Werkmeister proposition. The C.1 specimen difference between 5000th cycle and 3000th cycle is equal to $0.2 \cdot 10^{-3}$. The samples C.2, C.3 and C.4 were classified to plastic creep shakedown response and the difference was equal to $0.43 \cdot 10^{-3}$, $0.54 \cdot 10^{-3}$, $0.494 \cdot 10^{-3}$.

Energy calculation results

Figure 7a presents the energy calculation for the results of cyclic loading for A.1 sample. During cyclic loading drop of the dE^{P} (plastic energy density) can be observed. First loading caused the plastic strain equal to over 40% of all registered plastic strains in this study. It can be noticed that the dE^{E} does not change so drastically as the dE^{P}.

Figure 7b presents the energy calculation for the results of A.2 sample. In this plot it can be observed that in the beginning of the test the dE^{P} is close to A.1 sample (0.108 for A.1 and 0.0303 for A.2 sample). After 205 repetitions for A.1 sample dE^{P} was lower than 0.0001. In case of A.2 sample the same dE^{P} value occurred after 957 cycles. The differences between dE^{P} evolution means that the dE^{P} in sample A.2 decrease faster than in sample A.1 which can be understood as the huge decrease of the plastic strain rate.

Sample A.3 (Fig. 7c), where 2000 cycles were applied, behaves similar to A.2. In the specimen A.3 the initial value of dE^{P} is lower than in A.2 sample. It can be seen that dE^{E} is greater than dE^{P} from the beginning. This behaviour was different from A.2 and A.3 where at the beginning dE^{P} is greater than dE^{E}. The dE^{P} decreases steadily and the

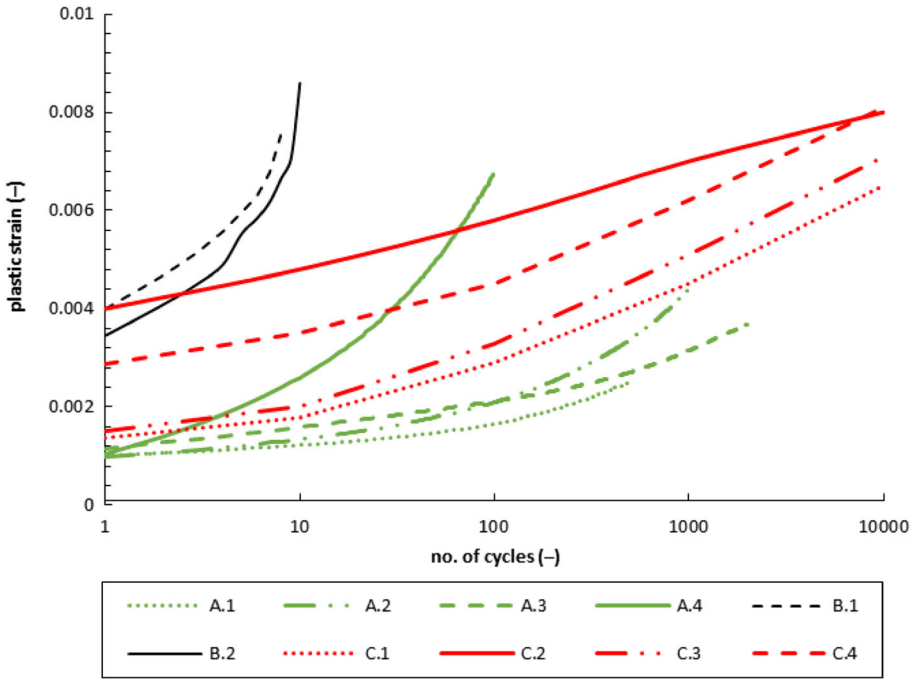


Fig. 6 Plot of the plastic strain from uniaxial cyclic loading of sandy clay vs. logarithm of number of cycles. (*Green* "A" indexed tests, *black* "B" indexed tests, *red* "C" indexed tests)

instant drop of this energy density was not noticed as it was in the previous samples.

Figure 7d presents the results of cyclic loading for sample A.4. The amplitude of the axial stress was equal to 153.1 kPa which was 0.68 of the maximum stress from the static uniaxial compression strength test. The beginning of the test was similar to A.3 sample. The dE^{P} decreases in the same way as in the sample A.3. The dE^{P} in the first two cycles was greater than the dE^{E}.

Figure 7e, f presents the results of the tests of B.1 and B.2 samples, respectively. For both of the specimens incremental collapse occurs after eight and 12 cycles, respectively. In both tests, the dE^{P} density is greater than the dE^{E} during the whole test. Before the collapse, the samples start to experience more plastic strain and the softening phenomena can be observed.

The samples from C.1 to C.4 present similar respond to cyclic loading (Fig. 7g–j). The stress amplitude and stress difference for these samples were close with the values. The plastic energy density drops after few first cycles.

Similar result was observed in the case of sample A.1. The A.1 sample characterizes with stress amplitude equal to 12.9 kPa, and the σ_a value for samples C.1 and C.4 was equal to 11.9 kPa and 11.2 kPa, respectively. Figure 8 presents comparison between A.1 and C.1 sample energy calculation. The energy density change during cyclic loading for both cases is very similar. Figure 10b presents comparison of plastic and elastic energy density between A.1 and C.1 specimens. The linear function was fitted to these relationships. The coefficient of determination R^2 for plastic energy density function was equal to 0.9388, and for elastic energy density the R^2 value was equal to 0.9042. The conclusion can be drawn that for close stress amplitude and stress difference values the energy density change is similar.

Results of energy calculation discussion

Accumulated plastic energy density

Figure 9 presents the change of accumulated dE^{P} during cyclic loading in various axial stress amplitude σ_{max}. It can be seen that after some value of axial stress the accumulated dE^{P} function changes its rate from the small rate in the beginning to higher rate after exceeding some σ value. Accumulation of the dE^{P} change can be explained by the fact of elastic and dE^{P} quotient. From the previous Figs. (7a–j, 8a–b), a conclusion was drawn that when the dE^{P} is closer to the dE^{E}, the bigger plastic strain occurs. From Fig. 9 it can be seen that after exceeding some σ value the accumulated dE^{P} density changes its rate. The higher the axial stress, the higher the accumulation of dE^{P}. This change of behaviour corresponds to a limit between plastic shakedown and plastic creep shakedown. Nevertheless, this limit is hard to estimate due to a lack of clear deflection point and its development in following cycles.

Determination of shakedown response from energy densities

Werkmeister et al. (2001) conducted studies on the responses of unbound aggregates subjected to cyclic

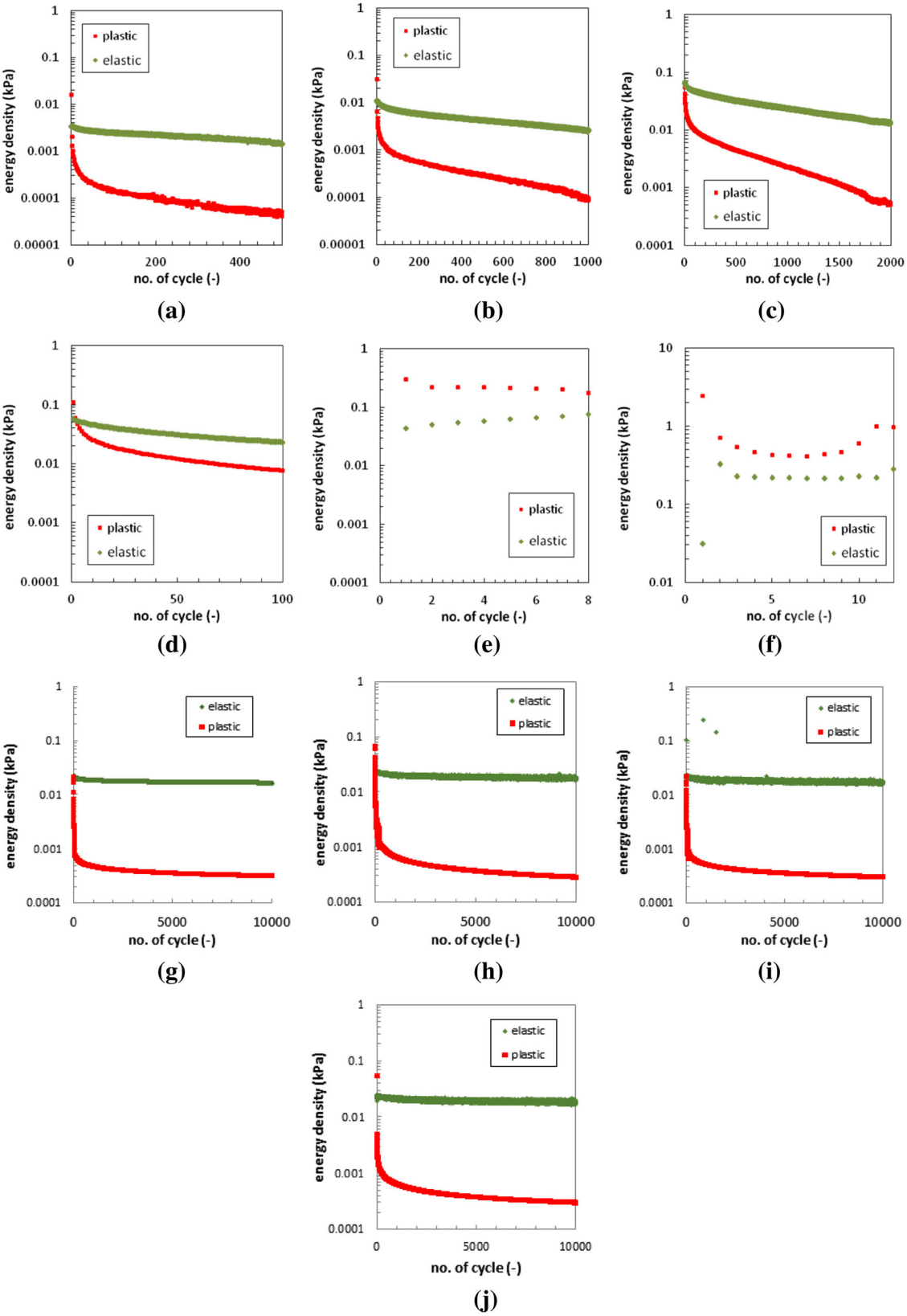


Fig. 7 Plot of the energy density versus number of cycles uniaxial cyclic loading of sandy clay in various stress levels (**a** A.1, **b** A.2, **c** A.3, **d** A.4, **e** B.1, **f** B.2, **g** C.1, **h** C.2, **i** C.3, **j** C.4)

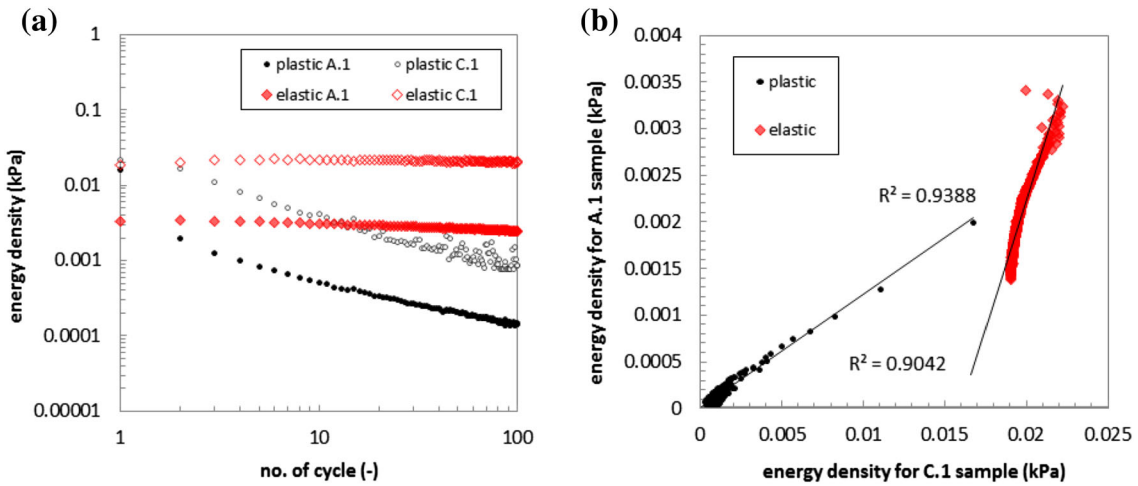


Fig. 8 Plot of the comparison of plastic and elastic energy density for samples A.1 and C.1

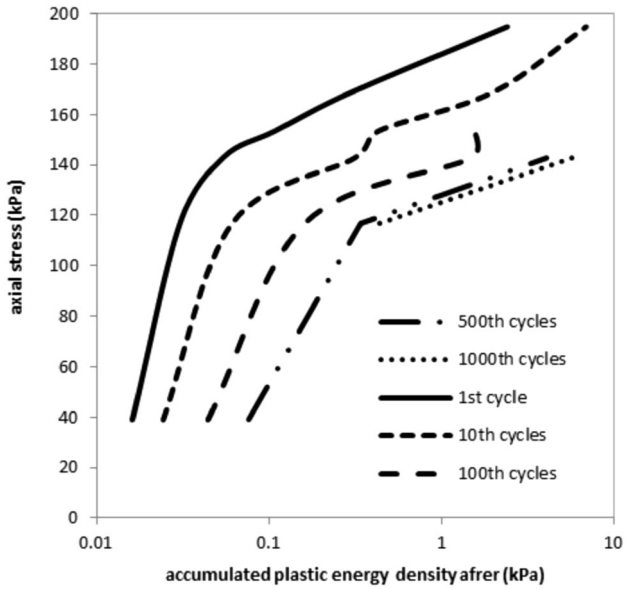


Fig. 9 Plot of accumulated plastic energy after first, tenth, 100th, 500th and 1000th cycle versus axial stress from uniaxial cyclic loading of cohesive soil

loading, which resulted in the proposition of the three ranges of response:

- Range A—plastic shakedown category; the response is plastic only for a first cycles of load and becomes resilient after the post–compaction. The accumulation of permanent strain decreases rapidly to a very small level,
- Range B—plastic creep shakedown category; the level of the plastic strain rate decreases to a low and nearly constant level during the first loading cycles,
- Range C—incremental collapse category; the accumulation of the plastic strain decreases slowly or does not.

Figure 10 presents an adjustment of shakedown criterion concept for cohesive soil based on proposition by Werkmeister (2006). The plot of normalized number of permanent strain versus number of cycles was calculated from the plastic strain rate analysis presented in Fig. 6. The exponential function's coefficient of determination of R^2 for samples A.1–A.4 was equal to 0.97 or higher. For cycle 3000 and 5000 plastic strains' difference based on Eq. 1 was calculated. Samples A.1 and A.3 were included in plastic shakedown range. Samples A.2 and A.4 were included in plastic creep shakedown range.

The proposition of the quotient of plastic and elastic energy dE^P/dE^E can help to understand the shakedown phenomena and plastic strain development in cohesive soil. In cohesive soil, limits of those three ranges are not clear as in the case of non-cohesive soil. The limit between the plastic shakedown and plastic creep shakedown ranges is not sharp and, therefore, the response of cohesive soil to cyclic loading is more fluent. The decrease of the plastic strain in low stress conditions is characterized by low dE^P in first cycles and by the fast decrease in further cycles. When stress level is greater, the initial dE^P is also bigger but a drop in dE^P value is lower. The plastic shakedown creep occurs when the abovementioned drop of dE^P is not presented; in other words, the decrease of the dE^P is constant. This causes greater plastic strain accumulation in the first few cycles and the creep phenomena in the next stage. When dE^P is big enough like in the case of high stress level, it may lead to incremental collapse even if the drop of its rate is observed.

The stress characteristic parameters such as stress amplitude and stress difference have their impact on the energy density change and, therefore, on strain development. The problem of cyclic stress impact on behaviour of

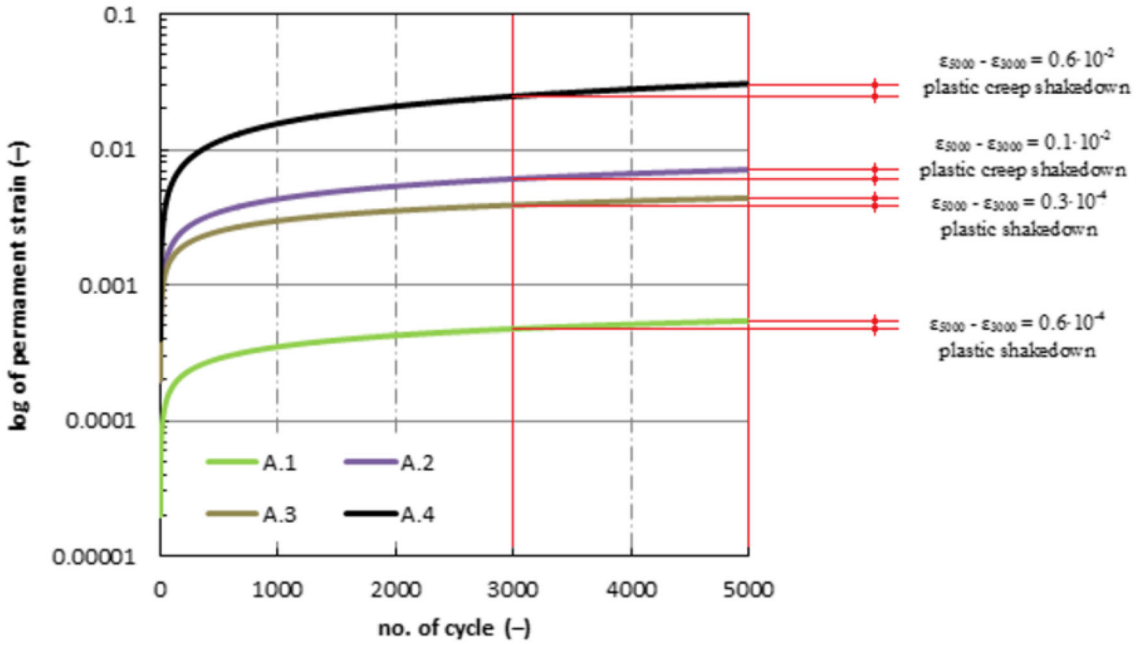


Fig. 10 Plot of shakedown criterion based on the Werkmeister proposition

soil can be resolved by including the stress parameters in energy calculations.

The plastic strain development during cyclic loading can be described by proposed in following Eq. (2) SE factor. The SE factor takes into account the shakedown concept and stress parameters: stress amplitude and stress median recognised as the cause of soil stiffness loss:

$$\left(\frac{\sigma_a}{\sigma_m}\right) \cdot \log\left(\frac{dE^P}{dE^E}\right) = S_E \tag{2}$$

The S_E factor value versus number of cycles is presented in Fig. 11. The cohesive soil can behave in plastic creep shakedown manner at the beginning of the test and later the response shifts to plastic shakedown. In range A, plastic strain can occur, but after numerous repetitions. In other words, the plastic strain could not be observed between two cycles but between 50 and 100 cycles. If the amplitude of axial stress increases, the dE^P approaches the dE^E and more plastic strain can be observed. If the amplitude of axial stress is big enough or in other words, the dE^P is nearly the same as the dE^E, the plastic strains occur in every cycle and plastic creep shakedown can be recognized. If the amount of dE^P is greater than dE^E, the plastic strain begins to increase in fast rate and incremental collapse can be observed.

In Fig. 11 it can be seen that incremental collapse occurs when the S_F is greater than 0. This corresponds to the range C from the Werkmeister proposition. Plastic creep shakedown occurs when value of S_F is between 0 and −0.25. When $S_F < -0.25$ plastic shakedown occurs.

Similar proposition of energy calculation was developed for hot-mix asphalt (HMA). The principle of this mechanism was describing of the fracture properties of HMA. In this framework, upper and lower thresholds, namely dissipated creep strain energy (DCSE) and fracture energy (FE), were distinguished. The DCSE limit is bound with continuous repeated loading and FE limit is associated with fracture of the HMA with a single application (Zhang et al. 2001; Birgisson et al. 2007; Tasdemir et al. 2010). When the energy threshold is exceeded, non-healable macrocracks develop and propagate along the mixture. Under this energy threshold, the rate of damage is governed by the creep properties of the mixture.

Conclusions

In this paper, cyclic uniaxial tests on cohesive soil were conducted to characterise shakedown phenomena. The test results were later analysed by application of the energy-based method. Based on the experimental results the following conclusions can be drawn:

1. During the uniaxial cyclic loading tests, the three ranges of the shakedown behaviour were observed which depended on the stress amplitude level. For the plastic shakedown response, rapid decrease of the dE^P and steady decrease of the dE^E were recognised. This behaviour results in small or in the lack of plastic strain

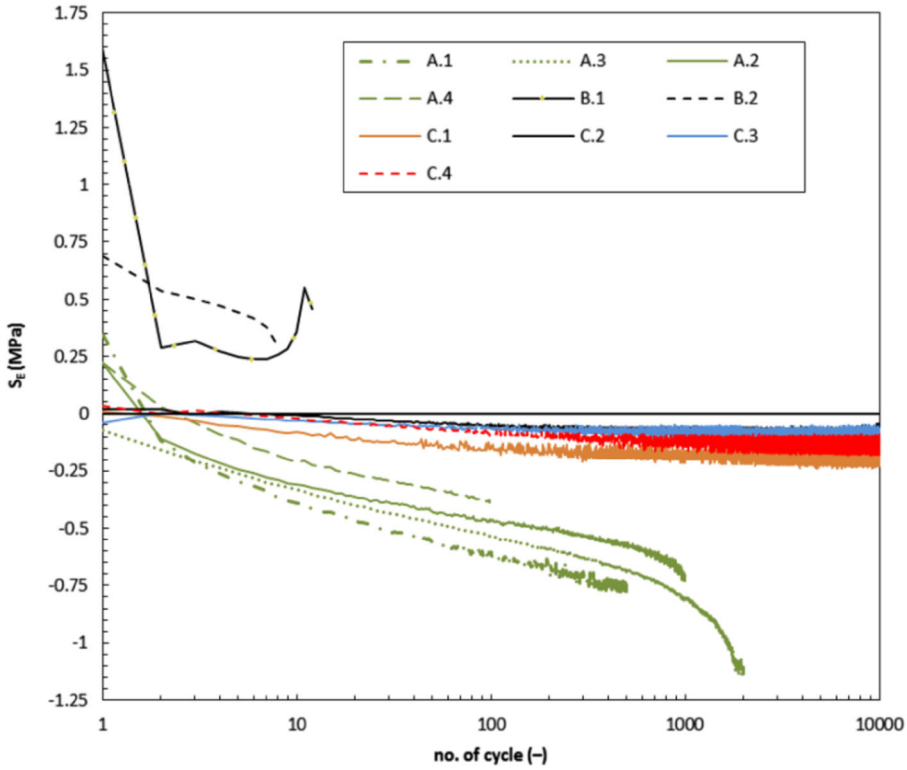


Fig. 11 Plot of S_E factor versus number of cycles

occurrence. For plastic creep shakedown a range of the dE^P decrease was less rapid, and the dE^E decreased more than that in plastic shakedown case. When the value of the dE^P decrease is near the dE^E, plastic creep shakedown may occur. Incremental collapse occurs when the dE^P is greater than the dE^E in all cycles which leads to failure.

2. Accumulated dE^P depends on the axial stress amplitude. This relation in non-linear and some inflection areas were recognised. These areas represent possible change of behaviour. Cohesive soil can behave as the plastic creep shakedown, but after numerous repetitions it may change to plastic shakedown.
3. A new proposition of distinguishing shakedown ranges for cohesive soil was proposed. The presented method bases on the ratio of plastic to dE^E and helps to recognise shakedown categories.
4. Energy-based method leads to the identification of smooth limit between plastic shakedown and plastic creep shakedown. The cohesive soil can behave in both ways which depend on the stress amplitude levels and the plastic and dE^E during cyclic loading.
5. Shakedown limit determination was performed with the calculation of the value of S_F. When the $S_F \geq 0$ the incremental collapse occurs. When $-0.25 \leq S_F < 0$ cohesive soil behaves as the plastic creep shakedown. When $S_F < -1$ the plastic shakedown may occur.
6. Cohesive soil subjected to cyclic loading behaves differently comparing to non-cohesive soil. The limit between the plastic shakedown and the plastic creep shakedown range is smooth and changes during cyclic loading. It may by possible that after post-compaction stage, the response can change due to fatigue of the material.
7. The practical application of the common logarithm of energy density quotient can lead to estimate maximal amplitude of cyclic loading which can improve pavement and foundation design process.

References

Birgisson B, Sangpetngam B, Roque R, Wang J (2007) Numerical implementation of a strain energy-based fracture model for a HMA materials. Int J Road Mat Pavement Des 8(1):7–45. doi:10.1080/14680629.2007.9690065

Boulbibane M, Collins IF (2015) Development Of A pavement rutting model using shakedown theory. Int J Pavement Eng Asph Technol 16(1):55–65. doi:10.1515/ijpeat-2015-0003

Cai Y, Gu C, Wang J, Juang CH, Xu C, Hu X (2013) One-way cyclic triaxial behavior of saturated clay: comparison between constant and variable confining pressure. J Geotech Geoenviron Eng 139(5):797–809. doi:10.1061/(ASCE)GT.1943-5606.0000760

Cai Y, Sun Q, Guo L, Juang CH, Wang J (2015) Permanent deformation characteristics of saturated sand under cyclic loading. Can Geotech J 52(6):1–13. doi:10.1139/cgj-2014-0341

Cerni G, Cardone F, Virgili A, Camilli S (2012) Characterisation of permanent deformation behaviour of unbound granular materials under repeated triaxial loading. Constr Build Mater 28(1):79–87. doi:10.1016/j.conbuildmat.2011.07.066

Chen C, Indraratna B, McDowell G, Rujikiatkamjorn C (2015) Discrete element modelling of lateral displacement of a granular assembly under cyclic loading. Comput Geotech 69:474–484. doi:10.1016/j.compgeo.2015.06.006

Cuéllar P, Mira P, Pastor M, Merodo JAF, Baeßler M, Rücker W (2014) A numerical model for the transient analysis of offshore foundations under cyclic loading. Comput Geotech 59:75–86. doi:10.1016/j.compgeo.2014.02.005

Danne S, Hettler A (2015) Experimental strain response-envelopes of granular materials for monotonous and low-cycle loading processes. In: Triantafylidis TH (ed) Holistic simulation of geotechnical installation processes. Springer International Publishing, Berlin-Heidelberg, pp 229–250

Daum M (2008) Simplified presentation of the stress-energy method for general commercial use. J Test Eval 36(1):100–102. doi:10.1520/JTE101202

Desai CS, Siriwardane HJ (1984) Constitutive laws for engineering materials with emphasis on geologic materials. Prentice-Hall, New Jersey

EN 13286-7 (2004) European Committee for Standardization. Unbound and hydraulically bound mixtures – Part 7: cyclic load triaxial test for unbound mixtures

Fadaee MJ, Saffari H, Tabatabaei R (2008) Shear effects in shakedown analysis of offshore structures. J Ocean Univ China 7(1):77–83. doi:10.1007/s11802-008-0177-z

Feng J, Wu XY, Zhu BL, Yang QX (2015) Analytical solution to one-dimensional consolidation in unsaturated soils under sinusoidal cyclic loading. J Cent South Univy 22:646–653. doi:10.1007/s11771-015-2566-y

Goldscheider M. (1978) Shakedown and incremental collapse of structures in dry sand bodies. In: Proceedings of dynamical methods in soil and rock mechanics, plastic and long-term effects in soils, Balkema, Rotterdam, 2

Green RA, Mitchell JK, Polito CP (2000) An energy-based excess pore pressure generation model for cohesionless soils. In: Proceedings of the Developments in Theoretical Geomechanics—The John Booker Memorial Symposium, Australia, 1–9

Gu C, Wang J, Cai Y, Yang Z, Gao Y (2012) Undrained cyclic triaxial behavior of saturated clays under variable confining pressure. Soil Dyn Earthq Eng 40:118–128. doi:10.1016/j.soildyn.2012.03.011

Jain A, Veas JM, Straesser S, Van Paepegem W, Verpoest I, Lomov SV (2015) The master sn curve approach–a hybrid multi-scale fatigue simulation of short fiber reinforced composites. Compos Part A. doi:10.1016/j.compositesa.2015.11.038

Kalinowska M, Jastrzębska M (2014) Behaviour of cohesive soil subjected to low-frequency cyclic loading in strain-controlled tests. Studia Getechnica et Mechanica 36(3):21–35. doi:10.2478/sgem-2014-0024

Karg C, François S, Haegeman W, Degrande G (2010) Elasto-plastic long–term behavior of granular soils: modelling and experimental validation. Soil Dyn Earthquake Eng 30(8):635–646. doi:10.1016/j.soildyn.2010.02.006

Kokkali P, Anastasopoulos I, Abdoun T, Gazetas G (2014) Static and cyclic rocking on sand: centrifuge versus reduced–scale 1 g experiments. Géotechnique 64(11):865–880. doi:10.1680/geot.14.P.064

Kokusho, T., and Y. Kaneko (2014), Dissipated & Strain Energies in Undrained Cyclic Loading Tests for Liquefaction Potential Evaluations, In: Proceedings of the tenth National Conference in Earthquake Engineering, Earthquake Engineering Research Institute, Anchorage, 1–10

Kokusho T, Ito F, Nagao Y, Green RA (2012) Influence of non/low–plastic fines and associated aging effects on liquefaction resistance. J Geotech Geoenviron Eng ASCE 138(6):747–756. doi:10.1061/(ASCE)GT.1943-5606.0000632

König JA, Maier G (1981) Shakedown analysis of elastoplastic structures: a review of recent developments. Nucl Eng Des 66(1):81–95. doi:10.1016/0029-5493(81)90183-7

Koseki J, Mikami T, Sato T (2014) Deformation characteristics of granular materials in cyclic one-dimensional loading tests. Transport Infrastruct Geotechnol 1(1):54–67. doi:10.1007/s40515-014-0002-7

Li LL, Dan HB, Wang LZ (2011) Undrained behavior of natural marine clay under cyclic loading. Ocean Eng 38(16):1792–1805. doi:10.1016/j.oceaneng.2011.09.004

Liang C, Liu T, Xiao J, Zou D, Yang Q (2015) Effect of stress amplitude on the damping of recycled aggregate concrete. Materials 8(8):5298–5312. doi:10.3390/ma8085242

Lu Z, Yao HL, Liu J, Hu Z (2014) Experimental evaluation and theoretical analysis of multi–layered road cumulative deformation under dynamic loads. Road Mat Pavement Des 15(1):35–54. doi:10.1080/14680629.2013.852609

Nega A, Nikraz H, Al-Qadi IL (2015) Simulation of shakedown behavior for flexible pavement's unbound granular layer. In: Harvey J, Chou KF (eds.) Airfield and Highway Pavements, pp 801–812. doi: 10.1061/9780784479216.071

Ostadan F, Deng N, Arango I (1996) Energy-based method for liquefaction potential evaluation, phase I Feasibility study U.S. Department of Commerce, Technology Administration, National Institute of Standards and Technology, Building and Fire Research Laboratory, San Francisco, pp 96–701

Panoskaltsis VP, Bahuguna S (1996) Micro and macromechanical aspects of the behavior of concrete materials with special emphasis on energy dissipation and on cyclic creep. J Mech Behav Mat 6(2):119–134. doi:10.1515/JMBM.1996.6.2.119

Pasik T, Chalecki M, Koda E (2015) Analysis of embedded retaining wall using the subgrade reaction method. Studia Geotechnica et Mechanica 37(1):59–73. doi:10.1515/sgem-2015-0008

PN–CEN ISO/TS 17892-4:2009, Polish committee for standardization. geotechnical investigations—Soil Laboratory Testing—Part 4: Sieve Analysis, Polish Committee for Standardization: Warsaw, Poland, 2009

PN–EN 13286:2010/AC, Polish Committee for Standardization. Unbound and hydraulically bound mixtures—Part 2: Testing methods for laboratory reference density and water content—Proctor compaction; Polish Committee for Standardization: Warsaw, Poland, 2010

PN–EN ISO 14688-2:2006, Polish Committee for Standardization. Geotechnical Investigations—Soil Classification—Part 2: Classification Rules, Polish Committee for Standardization: Warsaw, Poland, 2006

Puppala AJ, Saride S, Chomtid S (2009) Experimental and modeling studies of permament strains of subgrade soils. J Geotech Geoenviron Eng 135(10):1379–1389. doi:10.1061/(ASCE)GT.1943-5606.0000163

Sas W, Głuchowski A, Szymański A (2014) Impact of the stabilization of compacted cohesive soil–sandy clay on field criterion improvement. Ann Warsaw Univ Life Sci 46(2):139–151. doi:10.2478/sggw-2014-0012

Sas W, Głuchowski A, Radziemska M, Dzięcioł J, Szymański A (2015) Environmental and geotechnical assessment of the steel slags as a material for road structure. Materials 8(8):4857–4875. doi:10.3390/ma8084857

Seo J, Kim YC, Hu JW (2015) Pilot Study for investigating the cyclic behavior of slit damper systems with recentering shape memory alloy (SMA) bending bars used for seismic restrainers. Appl Sci 5(5):187–208. doi:10.3390/app5030187

Sharp RW, Booker JR (1984) Shakedown of pavements under moving surface loads. J Transport Eng 110:1–14. doi:10.1061/(ASCE)0733-947X(1984)110:1(1)

Soares R, Allen D, Little DN Berthelot CA (2014) Multi–scale computational mechanics model for predicting rutting in asphaltic pavement subjected to cyclic mechanical loading. In: Transportation Research Board 93rd Annual Meeting DC, USA. Accessed 12–16 Jan 2014 pp 1–17

Soliman H, Shalaby A (2015) Permanent deformation behavior of unbound granular base materials with varying moisture and fines content. Transport Geotech 4:1–12. doi:10.1016/j.trgeo.2015.06.001

Sun L, Gu Ch, Wang P (2015) Effects of cyclic confining pressure on the deformation characteristics of natural soft clay. Soil Dyn Earthquake Eng 78:99–109. doi:10.1016/j.soildyn.2015.07.010

Tang LS, Chen HK, Sang HT, Zhang SY, Zhang JY (2015) Determination of traffic-load-influenced depths in clayey subsoil based on the shakedown concept. Soil Dyn Earthquake Eng 77:182–191. doi:10.1016/j.soildyn.2015.05.009

Tao M, Mohammad LN, Nazzal MD, Zhang Z, Wu Z (2010) Application of shakedown theory in characterizing traditional and recycled pavement base materials. J Transport Eng 136(3):214–222. doi:10.1061/(ASCE)0733-947X(2010)136:3(214)

Tasdemir Y, Das PK, Brigisson B (2010) Determination of mixture fracture performance with the help of fracture mechanics. In: Proceedings of Ninth International Congress on Advances in Civil Engineering, Sept 2010, Turkey, 1–7

Werkmeister S (2003) Permanent deformation behavior of unbound granular materials. Ph.D. dissertation. University of Technology, Dresden, Germany

Werkmeister S (2006) Shakedown analysis of unbound granular materials using accelerated pavement test results from New Zealand's CAPTIF facility. Pavement Mech Perf 154:220–228. doi:10.1061/40866(198)28

Werkmeister S, Dawson A, Wellner F (2001) Permanent deformation behavior of granular materials and the shakedown concept. Transport Res Record 1757:75–81. doi:10.3141/1757-09

Wichtmann T (2005) Explicit accumulation model for non-cohesive soils under cyclic loading. Ph.D. Thesis, Ruhr-University Bochum

Zhang Z, Roque R, Birgisson B, Sangpetngam B (2001) Identification and verification of a suitable crack growth law. J Assoc Asphalt Paving Technol 70:206–241

Zhou J, Gong XN (2001) Strain degradation of saturated clay under cyclic loading. Can Geotech J 38(1):208–212. doi:10.1139/cgj-38-1-208

Acta Geophys. (2017) 65:345–351
DOI 10.1007/s11600-017-0027-8

RESEARCH ARTICLE

Evaluation of geological conditions for coalbed methane occurrence based on 3D seismic information: a case study in Fowa region, Xinjing coal mine, China

Juanjuan Li[1,2,3] · Fanjia Li[4] · Mingshun Hu[3] · Wei Zhang[1,2] · Dongming Pan[3]

Received: 28 February 2017 / Accepted: 4 March 2017 / Published online: 10 March 2017

Abstract The research on geological conditions of coalbed methane (CBM) occurrence is of great significance for predicting the high abundance CBM rich region and gas outburst risk area pre-warning. The No. 3 coal seam, in Yangquan coalfield of Qinshui basin, is the research target studied by 3D seismic exploration technique. The geological factors which affect CBM occurrence are interpreted based on the 3D seismic information. First, the geological structure (faults, folds, and collapse columns) is found out by the 3D seismic structural interpretation and the information of buried depth and thickness of the coal seam is calculated by the seismic horizons. Second, 3D elastic impedance (EI) and natural gamma attribute volumes are generated by prestack EI inversion and multi-attribute probabilistic neural network (PNN) inversion techniques which reflect the information of coal structure types and lithology of the roof and floor. Then, the information of metamorphic degree of seam and hydrogeology conditions can be obtained by the geological data. Consequently, geological conditions of CBM occurrence in No. 3 coal seam are evaluated which will provide scientific reference for high abundance CBM rich region prediction and gas outburst risk area pre-warning.

Keywords Coalbed methane occurrence · 3D seismic information · Geological conditions · Geological factors

Introduction

The coalbed methane (CBM) is considered to be self-generation and self-accumulation unconventional gas resource which has great significance in clean energy utilization, coal resources mining safety, and environmental protection. The reservoir formation conditions of CBM resources are quite complex, though the gas content is very rich in China. In addition, the reservoir characterized by strong heterogeneity, low gas-bearing saturation, poor permeability, and low formation pressure has a complicated occurrence law. It is the vital issue for CBM resource development how we search for its enrichment areas and it is also the key problem for gas accidents prevention how we predict gas outburst risk area. The research on geological conditions of CBM occurrence is of great importance for solving the above issues.

Experts and scholars at home and abroad have presented related research works on this subject from the following aspects: geological law of CBM occurrence, analysis of effects of geological factors on CBM, geophysical detection techniques of coal seam reservoir, and coal and gas outburst risk area. Li et al. (2013a) studied the occurrence characteristics of CBM and showed that its enrichment was mainly controlled by the regional tectonic subsidence and weak structural deformation on coal reservoirs after coal formation. The impacts of faults on the occurrence of CBM were analyzed in Huaibei coalfield (Guo et al. 2014). Li

✉ Juanjuan Li
lijuanjuan19850110@163.com

[1] IoT Perception Mine Research Center, China University of Mining and Technology, Xuzhou, China

[2] The National and Local Joint Engineering Laboratory of Internet Application Technology on Mine, China University of Mining and Technology, Xuzhou, China

[3] School of Resources and Geosciences, China University of Mining and Technology, Xuzhou, China

[4] Jiangsu Province Xuzhou Technician Institute, Xuzhou, China

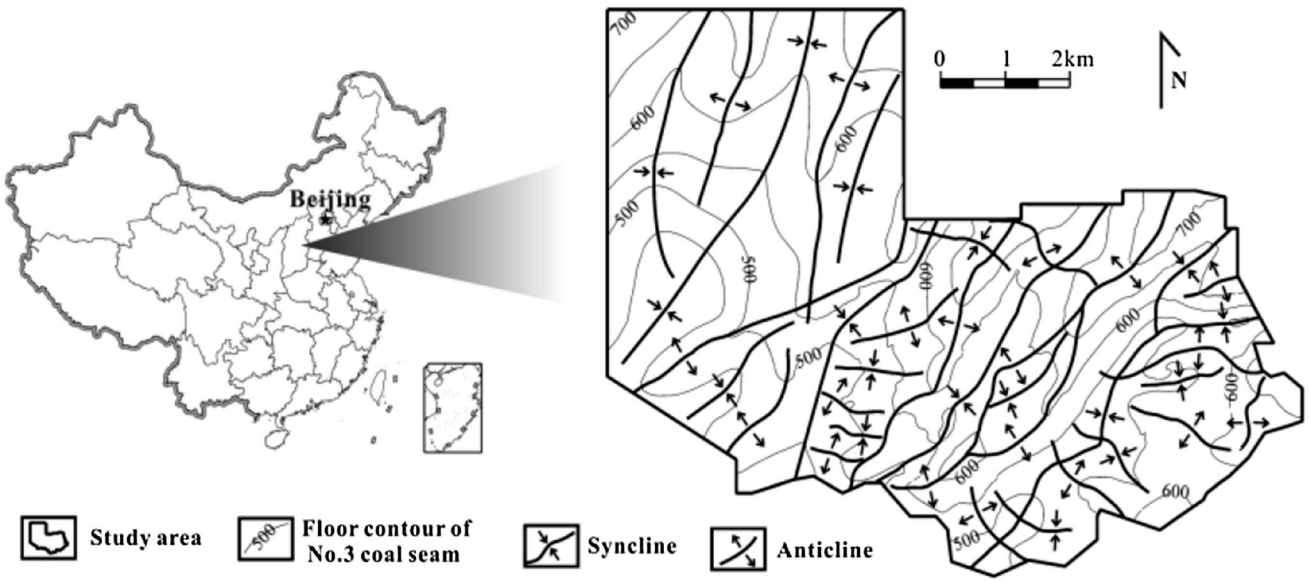


Fig. 1 Regional tectonic location and structure outline map of the study area (Liu et al. 2015)

et al. (2015) studied the formation history by numerical simulation in Shuigonghe Syncline and proposed tectonic evolution history as the most important factor for determining the generation–preservation–dissipation process of CBM. The fracture opening–closing degree parameter Δ and the fracture development degree parameter ξ were discussed for quantitative study of the effective migration system of CBM reservoir in southern Qinshui Basin (Wu et al. 2014). The mechanism of how the flowing water affected the CBM content was studied and it was proposed that it was the water solution that diminished the CBM content, rather than the water-driven action or the gas dissipation through cap rocks (Qin et al. 2005). Multi-component 3D seismic data were required and analyzed for the detailed characterization of CBM reservoirs in the Cedar Hill Field (Shuck et al. 1996). A new approach of integration of 3D, three-component seismic data, and AVO analysis were applied to detect fracture for characterizing CBM reservoirs (Ramos and Davis 1997). Hou et al. (2012) suggested that coal and gas outburst in the tectonically deformed coal, especially in the mylonitic coals, may be not only controlled by geological structure, but also influenced by the tectonic stress degradation of ductile deformation. A new method to predict coal and gas outburst was applied based on prestack seismic inversion techniques (Li et al. 2013b, 2016). Analysis indicated that geological conditions have significant effects on CBM occurrence for which the effective detection method should be applied.

Taking Fowa region in Xingjing coal mine as an example, the mainly mineable No. 3 coal seam as the research target, the authors of this paper attempted to evaluate the geological conditions by 3D seismic exploration technique. Based on the seismic data and well logs data, the structure and lithology information are interpreted, by which the geological conditions and distribution laws are studied in the reservoir, combined with the regional geological data and gas geological data.

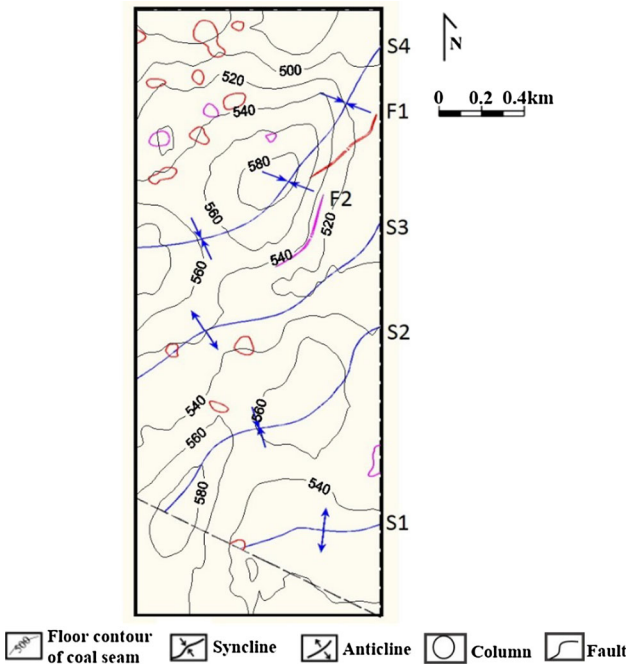


Fig. 2 Spatial distribution of geological structure in No. 3 coal seam [S1 and S3 represent syncline, S2 and S4 the anticline, F1 the reliable fault, and F2 the relatively reliable fault; the red circles (*semicircles*) represent the reliable collapse columns, and the *purple circles* (*semicircles*) represent the relatively reliable collapse columns]

Geological factors on CBM occurrence

The occurrence of CBM varies greatly from coalfield to coalfield and even the gas-bearing properties in the same coalfield have significant difference in regions. Generally speaking, CBM occurrence is mainly controlled by the geological factors: coal seam conditions (metamorphic degree, buried depth, thickness, coal structure types, etc.), geological structure conditions, lithology of roof and floor, and the hydrogeological conditions.

As the source and reservoir bed, physical conditions of coal seam directly influence the occurrence of CBM (specially its content). CBM content usually increases with the metamorphic degree, which is quite higher in meagre and anthracite coal, as well as with the fragmentation degree of coal structure. The preservation potency of the seam becomes stronger with an increasing thickness; thus, there is a positive correlationship between gas content and seam thickness in some regions.

The geological structures, folds and faults, are closely related with the CBM occurrence. Syncline has various gas contents in different parts, while anticline has low gas content in the top where the fissure develops well. In addition, the normal fault causes the CBM dissipation, while the footwall of reversed fault generally forms favorable sealing conditions.

Lithology conditions of the roof and floor affect the sealing ability of surrounding rocks directly. It is usually unfavorable for CBM preservation if the sand rock and carbonatite consist the roof and floor. However, the mud rock, an impermeable formation, has good sealing ability which will be enhanced with the increasing mud content. In addition, the greater the thickness, the more stable the continuity, and the stronger the sealing ability.

Hydrogeological conditions have significant impact on the CBM preservation and migration which have dual effects—gas dissipation and gas accumulation. In addition, the hydraulic migration escape effect causes the gas migration and dissipation in the areas where the faults develop with strong water conductivity.

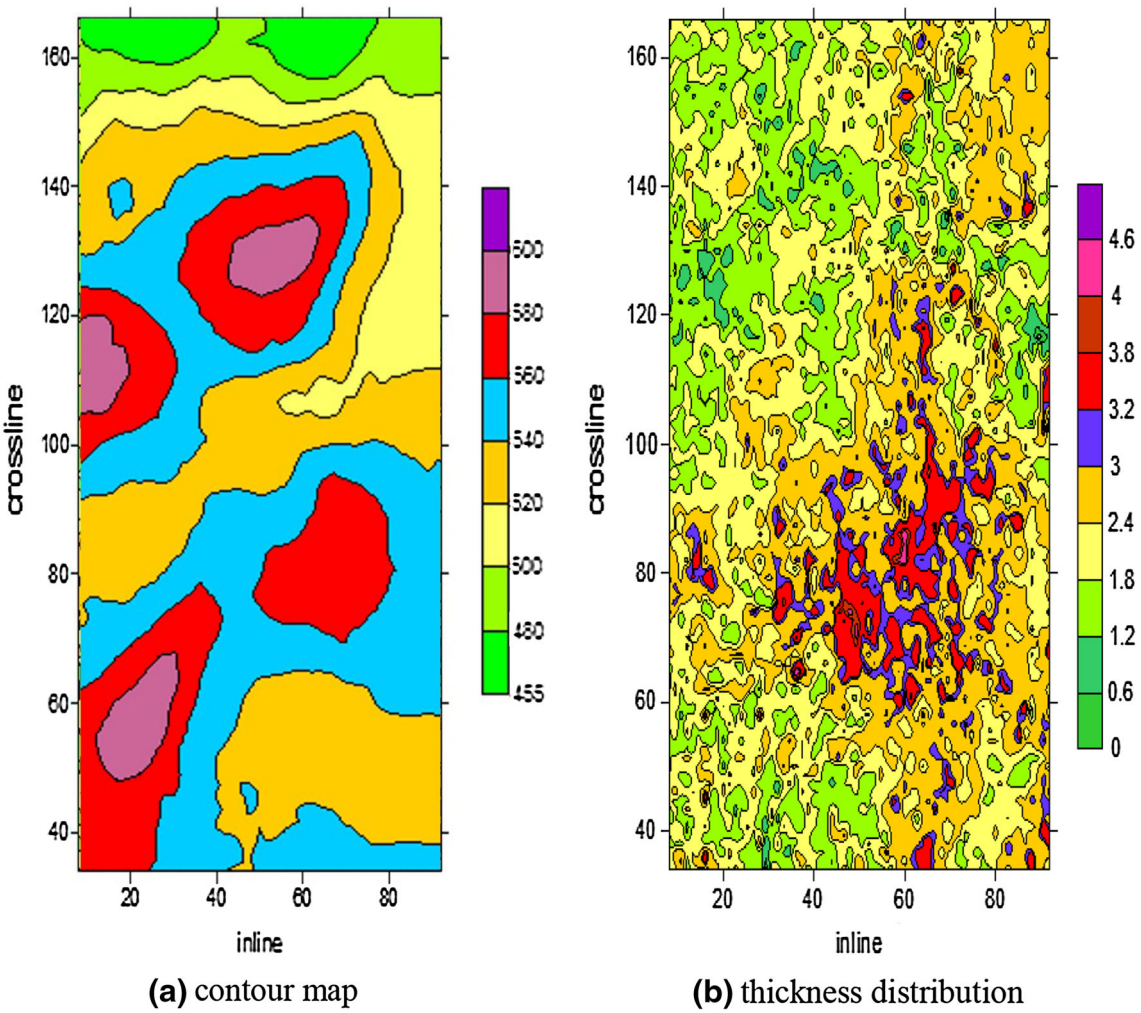


Fig. 3 Contour map of the floor and thickness distribution of the No. 3 coal seam

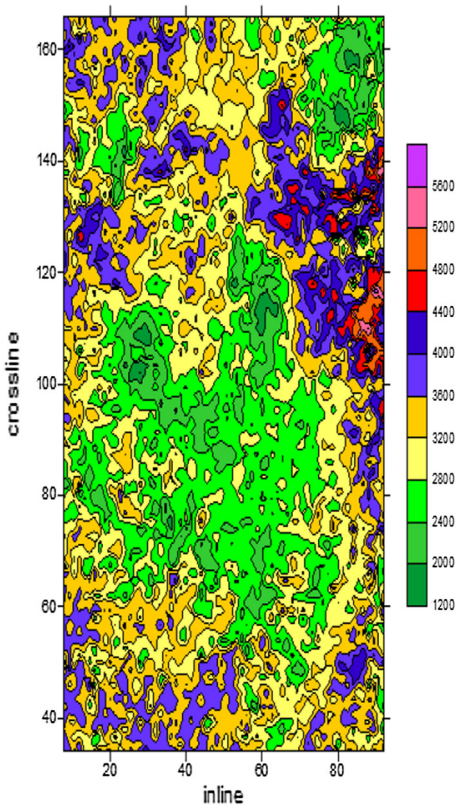


Fig. 4 EI slice along the No. 3 coal seam

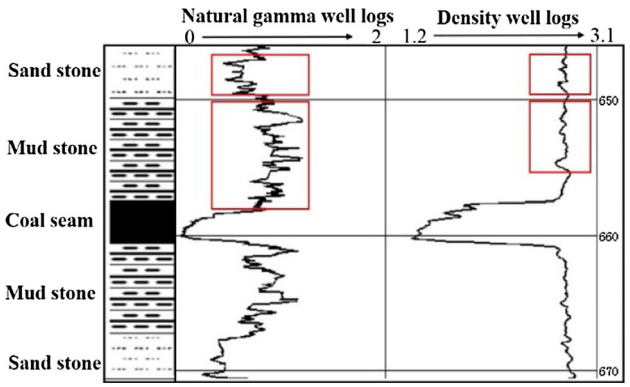


Fig. 5 Natural gamma and density well logs in No. 3-166 borehole

Therefore, the factors mentioned above have different influences in CBM occurrence. The following authors will attempt to evaluate geological conditions of CBM occurrence based on geological factors which can be reflected by 3D seismic information.

Case studies

The Xingjing coalmine, located in western part of Yangquan coalfield, northeastern part of Qinshui basin (Fig. 1), is one of the favorable blocks for CBM exploitation because of its high gas content in coal seams and Fowa region locates in its southwest corner (with an area of 2.88 km^2). However, it also encounters coal and gas outburst hazard during mining. The coal-bearing strata are the Upper Carboniferous–Lower Permian Taiyuan formation (C2-P1t) and the Lower Permian Shanxi formation (P1s). The stable No. 3 coal seam (P1s) and No. 15 coal seam (C2-P1t) are the mainly minable seams, with an average thickness of 2.33 and 6.14 m, respectively.

In this paper, No. 3 coal seam is considered as the study target zone which had numerous coal and gas outburst accidents during the excavation and recovery process. To predict high abundance CBM rich region or gas outburst risk area, the geological conditions evaluation of CBM occurrence will be done based on the 3D seismic data as well as the geological data and well logs data.

Geological structure conditions

The geological structure of this region is relatively simple and the No. 3 coal seam is a monoclinal structure which has secondary folds. Based on 3D seismic data, structures in the area have been explored by structural interpretation in the seam: undulate shape of folds has been proven and the spatial positions have been modified; the 22 collapse columns are exposed and there are 16 columns reliably interpreted; the 10 faults are detected; and 2 normal faults are reliably interpreted. The distribution of geological structure in No. 3 coal seam is shown in Fig. 2.

S1, the symmetric anticline with broad and gentle wings, has poor gas-bearing properties in its axis but good in other parts, while S2, the symmetric syncline with broad and gentle wings, is favorable for CBM preservation. IN addition, S3, the symmetric anticline, has poor preservation condition in its northwestern wing where the normal fault develops, while S4, the symmetric syncline, has poor preservation condition in its southeastern wing where the normal fault develops. Moreover, collapse columns would impact the integrity of the seam and destroy the gas sealing condition which cause the gas desorption near the columns in the northwestern part of the seam.

Buried depth and thickness of the coal seam

Prestack seismic inversion has been done after seismic data amplitude-preserving processing. When the horizon of the seam floor has been picked, time-depth conversion would be done. Then, the contour map of the seam floor can be obtained which represents the buried depth information. In addition, we can also get the time-thickness information of the seam by the horizons of roof and floor and the thickness information will be obtained combining average velocity which is calculated by coal samples laboratory testing. See Fig. 3.

The coal seam is buried rather deeper (mostly 500 m more) (Fig. 3a) and the deep burial depth is favorable for CBM preservation. Figure 3b shows that the thickness of the seam has increasing tendency from northwest to southeast direction; hence, the southeastern part is more favorable for CBM preservation.

Coal structure type division

There are significant differences between tectonically deformed coal (TDC) and primary coal. TDC, also called the soft coal, is characterized by low density, small velocity, and large porosity. It has close relationship with CBM rich area and is also considered as the necessary condition for coal and gas outburst.

As is known, EI has higher resolution in lithology identification and fluid detection than acoustic impedance (Connolly 1999); therefore, division of coal structure types will be done by EI information. The angle gathers (19°) have been extracted after seismic data amplitude-preserving processing and EI volume is obtained by inversion method. Green areas shown in Fig. 4 represent the low value of EI which indicates the existence of TDC.

Lithology of the roof and floor

It is the lithology of the roof and floor in the coal seam as key factor that impacts sealing ability of the surrounding rocks and the sealing ability will be enhanced with an increasing mud content. Therefore, evaluation of mudstone content can be regarded as an important index for determining the sealing ability of the surrounding rocks. The lithology difference of the roof can be distinguished from the natural gamma well logs but not the density well logs, which is presented in the red rectangles of Fig. 5; based on this, the lithology identification

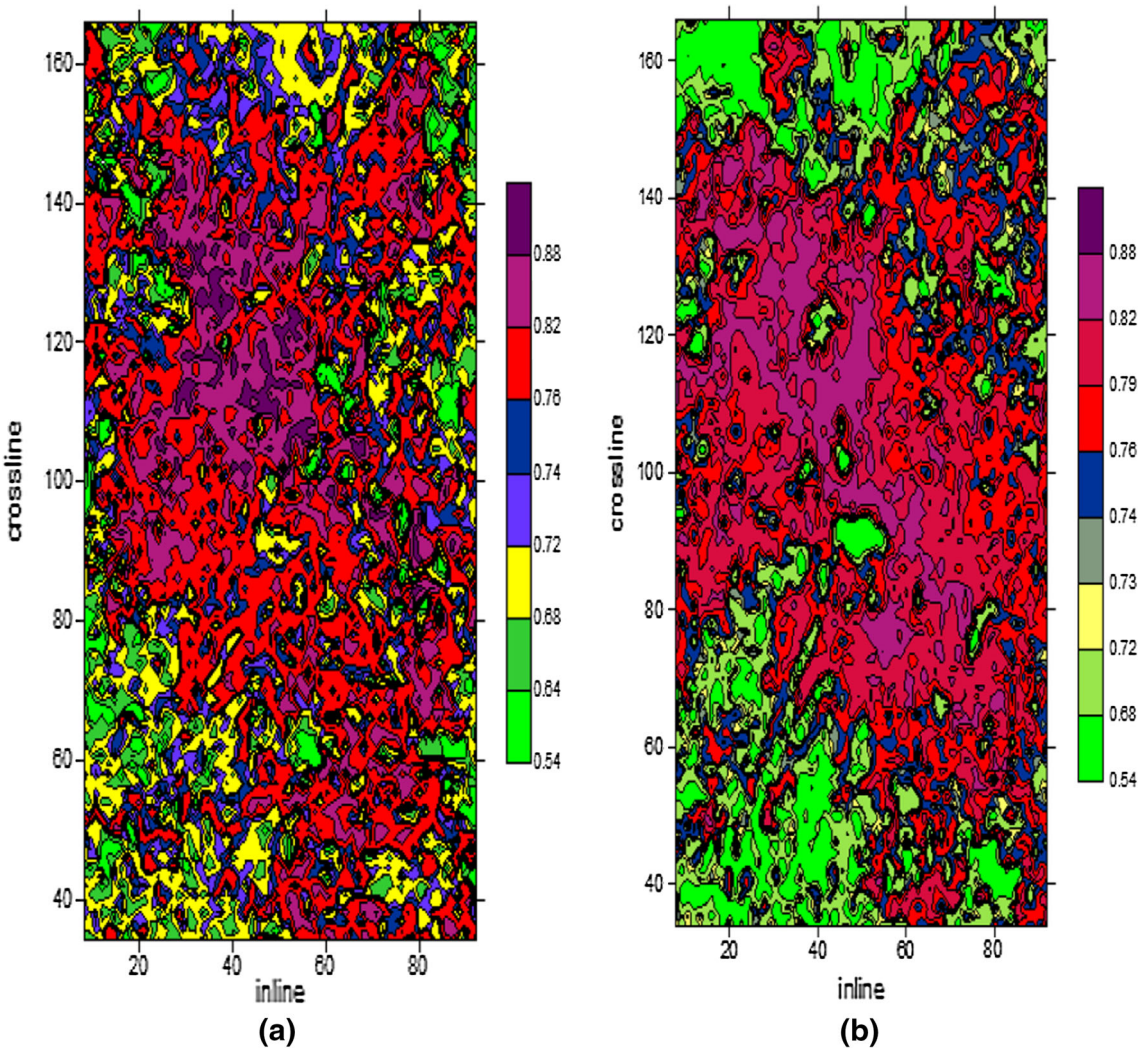


Fig. 6 Natural gamma slices along the roof and floor of the No. 3 coal seam

in layers can be performed by the natural gamma response.

The natural gamma value (from well logs) considered as the output and seismic stack data & EI data as the input, PNN technique is applied to train the network, and then the natural gamma volume is generated (Hampson et al. 2000). The slices along the roof are presented in Fig. 6a, where the area with natural gamma value greater than 0.7API accounts for 80%; the slices along the floor are presented in Fig. 6b where the area with natural gamma value greater than 0.75API reaches 85%. According to the above, we predict that the mud content of the strata is relatively high and the lithology is dominated by mudstone or sandy mudstone. Therefore, lithology of the roof and floor in the coal seam is favorable for CBM preservation.

Metamorphic grades of coal and hydrogeological condition

According to the regional geological data, Fowa region is in the high metamorphic zone and the grade of coal is anthracite. The migration condition of the seam is poor due to its permeability of 0.000375 md.

The formation strata have three aquifers with feeble water bearing. In addition, the aquifers have little hydraulic connection with each other as well as the surface water; therefore, the strata have poor recharge condition. Consequently, the hydraulic dissipation effect is quite weak and the hydraulic sealing effect is not significant as well as hydraulic plugging effect, which is generally favorable for CBM preservation.

Comprehensive evaluation of CBM enrichment area

The favorable areas for CBM enrichment, interpreted by geological factors (in sections "Geological structure conditions", "Buried depth and thickness of the coal seam", "Coal structure types division", "Lithology of the roof and floor", and "Metamorphic grades of coal and hydrogeological condition"), are not inconsistent and the influences of geological factors discussed above are also different on the CBM occurrence. According to Fig. 7, northern part of the research target is of better gas-bearing properties than the southern part, especially the northwestern part with the highest gas content. It is also proved that CBM occurrence analysis by a single geological factor has limitations.

Conclusions

The 3D seismic exploration technique has successfully been applied for evaluating geological conditions of CBM occurrence in the Fowa region, Xingjing coal mine, and

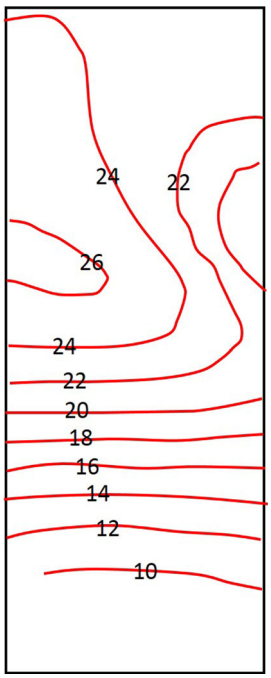


Fig. 7 Distribution of CBM content prediction in No. 3 coal seam (unit: m^3/t)

geological factors which contain geological structure, buried depth, coal seam thickness, coal structure types, lithology of the roof and floor, etc., are interpreted based on the 3D seismic information in the No. 3 coal seam. The findings can be summarized as follows:

1. The occurrence of CBM is influenced by variable geological factors and the prediction of CBM enrichment area by a single factor has limitations.
2. The following geological conditions are favorable for CBM occurrence in the research target: anthracite with high gas content in the No. 3 coal seam, aquifers with feeble water bearing and little hydraulic connection, the deep burial depth, and the lithology of the roof and floor with high mud content.
3. Geological structure distribution, division of coal structure type, and thickness of the seam cause uneven distribution of CBM.

Acknowledgements The authors acknowledge the financial support from Fundamental Research Funds for the Central Universities (2015QNB19) and the National Natural Science Foundation of China (41430317). Also special thanks to reviewers and editors' hard work.

References

Connolly P (1999) Elastic impedance the leading edge 18(4):438–452. doi:10.1190/1.1438307

Guo P, Cheng Y, Jin K, Liu Y (2014) The impact of faults on the occurrence of coal bed methane in Renlou coal mine, Huaibei coalfield, China. J Nat Gas Sci Eng 17:151–158. doi:10.1016/j.jngse.2013.12.003

Hampson D, Todorov T, Russell B (2000) Using multi-attribute transforms to predict log properties from seismic data. Explor. Geophys 31(3):481–487. doi:10.1071/EG00481

Hou Q, Li H, Fan J, Ju Y, Wang T, Li X, Wu Y (2012) Structure and coalbed methane occurrence in tectonically deformed coals. Sci China Earth Sci 55(11):1755–1763. doi:10.1007/s11430-012-4493-1

Li J, Cui R, Pan D, Chen S, Hu M, Zhao X (2013a) Prediction of gas outburst in a coal mine based on simultaneous prestack seismic inversion: a case study in China. J. Seismic Explor. 22(5):463–475

Li M, Jiang B, Lin S, Fe Lan, Wang J (2013b) Structural controls on coalbed methane reservoirs in Faer coal mine. Southwest China J Earth Sci 24(3):437–448. doi:10.1007/s12583-013-0340-3

Li L, Wei C, Qi Y, Cao J, Wang K, Bao Y (2015) Coalbed methane reservoir formation history and its geological control at the Shuigonghe Syncline. Arab J Geosci 8(2):619–630. doi:10.1007/s12517-013-1246-1

Li J, Pan D, Cui R, Ding E, Zhang W, Hu M (2016) Prediction of tectonically deformed coal based on lithologic seismic information. J Geophys Eng 13(1):116–122. doi:10.1088/1742-2132/13/1/116

Liu J, Jiang B, Li M, Qu Z, Wang L, Li L (2015) Structural control on pore-fracture characteristics of coals from Xinjing coal mine, northeastern Qinshui basin. China Arab J Geosci 8(7):4421–4431. doi:10.1007/s12517-014-1551-3

Qin S, Song Y, Tang X, Fu G (2005) The mechanism of the flowing ground water impacting on coalbed gas content. Chin Sci Bull 50(Suppl. 1):118–123. doi:10.1007/BF03184093

Ramos ACB, Davis TL (1997) 3-D AVO analysis and modeling applied to fracture detection in coalbed methane reservoirs. Geophysics 62(6):1683–1695. doi:10.1190/1.1444268

Shuck EL, Davis TL, Benson RD (1996) Multicomponent 3-D characterization of a coalbed methane reservoir. Geophysics 61(2):315–330. doi:10.1190/1.1443961

Wu C, Qin Y, Zhou L (2014) Effective migration system of coalbed methane reservoirs in the southern Qinshui Basin. Sci China Earth Sci 57(12):2978–2984. doi:10.1007/s11430-014-4988-z

Acta Geophys. (2017) 65:353–361
DOI 10.1007/s11600-017-0035-8

RESEARCH ARTICLE

Solution of one-dimensional space- and time-fractional advection–dispersion equation by homotopy perturbation method

Mritunjay Kumar Singh[1] · **Ayan Chatterjee**[1]

Received: 9 March 2017 / Accepted: 28 March 2017 / Published online: 4 April 2017

Abstract This study develops solution of one-dimensional space–time fractional advection–dispersion equation (FADE). Various forms of dispersion and velocity profiles (i.e. space dependent and both space–time dependent) are considered throughout the study. Homotopy perturbation method (HPM) is used to solve the problem semi-analytically. The advantage of HPM is that it does not require much information about the boundary of the aquifer. The initial condition may be measured for an aquifer, but sometimes it is very difficult to specify the boundary conditions. The FADE is employed for modeling the fate of contaminants in both homogeneous and heterogeneous porous formations subject to an increasing spatially dependent source condition. It is found that the contaminant concentration changes with the order of FADE as fractional-order derivative contains the memory of the system, i.e. how the system changes from one integer order to another integer order. FADEs are used to model the non-local system, hence this study helps understand the physical meaning of parameters involved in the velocity and dispersion.

Keywords Space–time dependent FADE · Dispersion · Velocity · HPM

✉ Mritunjay Kumar Singh
drmks29@rediffmail.com

Ayan Chatterjee
ayan@am.ism.ac.in

[1] Department of Applied Mathematics, Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand 826004, India

List of symbols

x and t	Distance (L) and time (T) for aquifer
$D(x,t)$	Dispersion (L^2T^{-1}) along positive x direction
$V(x,t)$	Velocity profile (L^2T^{-1}) for groundwater
$p, q, \alpha, \beta, \gamma$	Fractional orders
$C(x,t)$	Contaminant concentration (ML^{-3})
$G(x,t)$	Additional source or sink term in ADE
F and s	Dispersive flux and mean velocity for aquifer
$g(x)$	Initial contaminant concentration for aquifer

Introduction

Industrial effluents, municipal garbage, cemeteries, mine spoils, etc., are amongst the sources that generate contaminants that seep down from the soil surface to groundwater through percolation (Batu 2006). The fate of contaminants in groundwater is often described by the advection–dispersion equation (ADE). This equation is originally non-linear in nature and solved numerically, using finite difference or finite element methods. Otherwise, the equation is linearized by means of transformations and then solved analytically using either Laplace transform, Fourier transform, Hankel transform or Green's function methods. A range of initial and boundary value ADEs are solved either analytically or numerically or even stochastically. A short review of analytical techniques or solutions is provided below.

Sander and Braddock (2005) derived an analytical solution for transient, unsaturated transport of water and contaminants through horizontal porous media and discussed a range of analytical solutions. They derived a similarity solution using scale- and time-dependent dispersivities for contaminant transport in unsaturated flow and found that the solute did not disperse relative to water.

Smedt (2006) derived an analytical solution for transport of decaying solute in rivers with transient storage and first order decay for non-conservative solute. Singh et al. (2008) obtained analytical solution for one-dimensional ADEs with variable coefficients. Srinivasan and Clement (2008) derived an analytical solution for sequentially coupled reactive solute transport, such as radioactive wastes or chlorinated solvent carried nitrogenous species. Chen and Liu (2011) developed a generalized analytical solution in a finite spatial domain subject to arbitrary inlet boundary condition.

From last decade, space- or time-fractional advection–dispersion equations are modeled and solved by researchers and hydrologists to know how this contaminant concentration changes with the fractional order of the equations or to carry forward the change of the system with memory. Actually one can say that fractional order is used because it preserves the history, but general derivative does not preserve history. Fractional order (non-local differential operator) is used to model the natural system because integer-order (local operator) differential operator is not able to describe the situation properly. So the fractional-order derivative becomes popular to model the critical non-local systems. Many space-dependent or time-dependent FADEs are solved numerically or analytically independently, whereas very few researchers considered both time- and space-fractional advection–dispersion equation. So, we considered both space–time dependent FADE in the present work with space and space–time dependent velocity and dispersion.

Zhang et al. (2007) observed local variation of transport speed with space-dependent coefficients in homogeneous space-fractional advection–dispersion equation using generalized Fick's law. Benson et al. (2000) applied FADE with dispersion parameter independent of time or distance. Scaling behavior of plumes that underwent Levy motion was accounted by the fractional derivative and solved using the Green's function. Meerschaert and Tadjeran (2004) presented finite difference approximations of the FADE for modeling one-dimensional variable coefficient transport of passive tracers in a porous medium. The FADE was illustrated for capturing the arrival of tracer in the beginning in a radial flow. Roop (2008) solved one-dimensional space FADE with boundary layer using a finite element method where he derived a theoretical estimate to show that the under-resolved case corresponding to $\varepsilon < h^{\alpha-1}$, where α was the order of the diffusion operator. Huang et al. (2008) developed an unconditionally stable finite element (FEM) approach to solve the one-dimensional FADE based on the Caputo definition of the fractional derivative with singularity at the boundaries. The simulation results can be improved by using the third kind of boundary with a fractional-order derivative as the inlet boundary condition. A note on the finite element method for the space FADE was reported by Zheng et al. (2010). Li et al. (2011) provided a numerical approximation of non-linear fractional differential equations with sub-diffusion and super-diffusion. Murio (2008) developed a semi-analytical finite element method for a class of time FADEs in which time-fractional diffusion equations were used when transport processes were described with long memory where the rate of diffusion was inconsistent with the classical Brownian motion. Schumer et al. (2001) described that a fractional advection–dispersion equation was a generalization of the classical ADE in which the second-order derivative was replaced with a fractional-order derivative. Schumer et al. (2009) described a FADE can arise for non-integer-order derivative on time or space terms. Space-fractional and time-fractional ADEs arised with velocity variations and power law particle residence time distributions with memory in time, respectively. Liu et al. (2004) derived numerical solution of the space-fractional Fokker–Planck equation. Meerschaert et al. (2006) described finite difference methods for two-dimensional fractional dispersion equation. In that paper they discussed an implicit method to solve a class of two-dimensional initial-boundary value fractional partial differential equations with variable coefficients in a finite domain. Marinca et al. (2009) used an optimal homotopy asymptotic method applied to the steady flow of a fourth-grade fluid past through a porous plate. Pandey et al. (2012) described analytical solution for the space–time fractional advection–dispersion equation using the optimal homotopy asymptotic method in which they explored numeric examples on space–time fractional equation.

Now from the review of literature, it is found that very few problems were considered on both space–time fractional advection–dispersion equations. Most of them incorporated fractional order directly on the classical advection–dispersion equation with space- and time-dependent dispersion and velocity. As an example, Pandey et al. (2011) generalized the classical ADE to a fractional ADE by replacing the first-order time and first- and second-order space derivatives by Caputo derivatives of orders $0 < \alpha \leq 1$, $0 < \beta \leq 1$ and $0 < \gamma \leq 1$, respectively. If the equation is considered like this, it does not always follow the fractional flux condition and it is possible only when the dispersion coefficient and velocity coefficients are constant or time-dependent.

In this paper, a general type of advection–dispersion equation is considered and fractional derivative in Caputo sense on space and time is used with the orders $0 < \alpha \leq 1$, $1 < \beta \leq 2$ and $0 < \gamma \leq 1$, respectively, to incorporate the effect of space or both space–time dependent velocity and

dispersion coefficients. Space–time fractional advection–dispersion equation is considered for two types of dispersion and velocity profiles. HPM is used to solve the modeled problem semi-analytically and from this solution all types of FADEs, i.e. only space-dependent and only time-dependent and general ADE are found. The graphical representations for all the cases are shown.

Model of the advection–dispersion equation

According to the generalized Fick's law (Schumer et al. 2001; Paradisi et al. 2001) the dispersive flux is proportional to a fractional derivative of solute concentration. So the flux was considered in the following form:

$$F = -D\frac{1+s}{2}\frac{\partial^p C}{\partial x^p} - D\frac{1-s}{2}\frac{\partial^p C}{\partial(-x)^p}, \tag{1}$$

where F is dispersive flux, D is dispersion coefficient, s is mean velocity, p is fractional order, C is contaminant concentration and x is space variable. Considering $p = 1$, Eq. (1) reduces to the classical Fick's law. Using the mean velocity $s = 1$ (Schumer et al. 2001; Zhang et al. 2007) in (1) the dispersive flux reduces to:

$$F = -D\frac{\partial^p C}{\partial x^p}. \tag{2}$$

The dispersive flux is considered as space-dependent function. So the space-fractional advection–dispersion equation can be written as

$$\frac{\partial C}{\partial t} = -\frac{\partial}{\partial x}\left[V(x,t)C - D(x,t)\frac{\partial^p C}{\partial x^p}\right] \tag{3}$$

Now to consider both space- and time-fractional advection–dispersion equation, we can write Eq. (3) as

$$\frac{\partial^q C}{\partial t^q} = -\frac{\partial^q}{\partial x^q}\left[V(x,t)C - D(x,t)\frac{\partial^p C}{\partial x^p}\right], \tag{4}$$

where $V(x, t)$ and $D(x, t)$ are velocity and dispersion of the groundwater, respectively, and p and q are fractional orders. According to our consideration, velocity and dispersion are function of space, but it may be possible that dispersion and velocity are both space- and time-dependent. If dispersion and velocity are dependent only on time or constant then Eq. (4) becomes:

$$\frac{\partial^q C}{\partial t^q} = -V\frac{\partial^q C}{\partial x^q} + D\frac{\partial^{p+q} C}{\partial x^{p+q}}. \tag{5}$$

So we see that the application of fractional order on the classical ADE holds only when the dispersion and velocity profiles are constant or time-dependent.

The source term is incorporated in the FADE and initial condition is given by $C(x, 0) = 0$. The velocity and dispersion coefficients are considered as space and time-dependent or constant. So the mathematical formulation of the space- and time-fractional ADE is given as

$$\frac{\partial^q C}{\partial t^q} = -\frac{\partial^q}{\partial x^q}\left[V(x,t)C - D(x,t)\frac{\partial^p C}{\partial x^p}\right] + G(x,t). \tag{6}$$

Equation (6) can be written as follows:

$$\frac{\partial^\alpha C}{\partial t^\alpha} = D_1(x,t)\frac{\partial^\beta C}{\partial x^\beta} - V_1(x,t)\frac{\partial^\gamma C}{\partial x^\gamma} + V_2(x,t)C + G(x,t), \tag{7}$$

where $0 < \alpha \le 1$, $1 < \beta \le 2$, $0 < \gamma \le 1$ and $G(x, t)$ is an additional source or sink term incorporated in the fractional ADE. The initial condition is given by $C(x, 0) = 0$.

Analytical solution

Momani and Odibat (2007) developed the HPM for a non-linear partial differential equation (PDE) with initial condition. Here, D^m is the usual integer differential operator of order m, $m - 1 < \alpha \le m$, and J^α is the Riemann–Liouville integral operator of order $\alpha > 0$, defined as:

$$J^\alpha f = \frac{1}{\Gamma(\mu)}\int_0^x (x-t)^{\alpha-1} f(t)\mathrm{d}t, \quad x > 0, \tag{8a}$$

where μ is the parameter, and J^α is an operator satisfying the following properties for $f \in C_\mu, \mu > -1, \alpha > 0, \beta > 0$:

1. $$J^\alpha t^\gamma = \frac{\Gamma(\gamma+1)}{\Gamma(\alpha+\gamma+1)}t^{\alpha+\gamma} \tag{8b}$$

2. $$J^\alpha J^\beta f(t) = J^{\alpha+\beta} f(t) \tag{8c}$$

3. $$J^\alpha J^\beta f(t) = J^\beta J^\alpha f(t). \tag{8d}$$

The PDE can be expressed in a general form as

$$D^\beta_{*t} C(x,t) = L(C, C_x, C_{xx}) + N(C, C_x, C_{xx}) + f(x,t) \quad t > 0, \tag{9}$$

where L is a linear operator, N is a non-linear operator, f is a known function, and D^β_* is the Caputo fractional derivative of order β. The initial condition is given as follows:

$$C^k(x,0) = g_k(x), \quad k = 0, 1, 2, \ldots m-1. \tag{10}$$

Following He (2005), we now construct the homotopy for Eq. (9) as:

$$\frac{\partial C^m}{\partial t^m} - f(x,t) = p_1\left[\frac{\partial C^m}{\partial t^m} + L(C, C_x, C_{xx}) + N(C, C_x, C_{xx}) - D^\beta_{*t} C(x,t)\right], \tag{11}$$

where $p_1 \in [0,1]$ is the homotopy parameter, called the embedding parameter, that varies from zero to unity. If $p_1 = 0$, then Eq. (11) becomes linearized:

$$\frac{\partial C^m}{\partial t^m} = f(x,t). \tag{12}$$

Solution of Eq. (12) yields the initial approximation. When $p_1 = 1$ Eq. (11) becomes the original Eq. (9) whose solution is the exact solution. Thus, the change in p_1 from zero to unity corresponds to the initial approximation to the exact solution. In topology this process is referred to as deformation. The basic assumption is that the solution of Eq. (9) can be expressed as a power series in embedding parameter p_1. Following Momani and Odibat (2007) approximation, we approximate the solution of Eq. (11) as follows:

$$C(x,t) = \sum_{n=0}^{\infty} C_n(x,t). \tag{13}$$

Here, $C_n(x,t)$ is the nth order coefficient of the power series generated by p_1.

Using the truncated series the solution can be written as:

$$C_N(x,t) = \sum_{n=0}^{N-1} C_n(x,t). \tag{14}$$

Therefore, $C(x,t) \approx C_N(x,t)$. (15)

Obibat used this method for the time-fractional case and we want to use this for time as well as space-fractional case, so we construct the homotopy in the following way:

$$\frac{\partial C}{\partial t} - G(x,t) = p_1\left[\frac{\partial C}{\partial t} + D_1(x,t)\frac{\partial^\beta C}{\partial x^\beta} - V_1(x,t)\frac{\partial^\gamma C}{\partial x^\gamma} + V_2(x,t)C - D^\alpha_{*t}C(x,t)\right], \tag{16}$$

where $0 < \alpha \le 1$, $1 < \beta \le 2$, $0 < \gamma \le 1$ and p_1 is homotopy parameter. If p_1 is zero then we get that initial guess but when $p_1 = 1$ then we get the original equation back, so if p_1 varies from zero to one then the solution also varies from initial guess to the exact solution. In Eq. (16) the series approximation in terms of p_1 and c_n is used and written as follows:

$$C(x,t) = C_0 + C_1 p_1 + C_2 p_1^2 + C_3 p_1^3 + \cdots$$

Therefore, the truncated exact solution is given by:

$$C(x,t) = \sum_{n=0}^{N-1} C_n(x,t) \tag{17}$$

Case-I: analytical solution with space-dependent dispersion and velocity

The source term is considered linear in spatial domain and the aquifer is initially contamination free. So we consider that $G(x,t) = x$ and dispersion and velocity depends on time. Then $V_1(x,t) = x$, $V_2(x,t) = 1$ and $D_1(x,t) = aV$.

$$\frac{\partial^\alpha C}{\partial t^\alpha} = D_1(x,t)\frac{\partial^\beta C}{\partial x^\beta} - V_1(x,t)\frac{\partial^\gamma C}{\partial x^\gamma} + V_2(x,t)C + G(x,t) \tag{18}$$

with $C(x,0) = 0$.

Using homotopy, Eq. (18) can be rewritten as

$$\frac{\partial C}{\partial t} - G(x,t) = p_1\left[\frac{\partial C}{\partial t} + D_1(x,t)\frac{\partial^\beta C}{\partial x^\beta} - V_1(x,t)\frac{\partial^\gamma C}{\partial x^\gamma} + V_2(x,t)C - D^\alpha_{*t}C(x,t)\right]. \tag{19}$$

We approximate the contaminant concentration in terms of p_1 and therefore we write as follows:

$$C(x,t) = C_0 + C_1 p_1 + C_2 p_1^2 + C_3 p_1^3 + \cdots \tag{20}$$

Putting the value of C in Eq. (19) yields:

$$\frac{\partial C_0}{\partial t} = x; \quad C_0(x,0) = 0. \tag{21}$$

Equating the equal powers of p we get that

$$\frac{\partial C_1}{\partial t} = \left[\frac{\partial C_0}{\partial t} + D_1(x,t)\frac{\partial^\beta C_0}{\partial x^\beta} - V_1(x,t)\frac{\partial^\gamma C_0}{\partial x^\gamma} + V_2(x,t)C_0 - D^\alpha_{*t}C_0(x,t)\right]; \quad C_1(x,0) = 0 \tag{22}$$

$$\frac{\partial C_2}{\partial t} = \left[\frac{\partial C_1}{\partial t} + D_1(x,t)\frac{\partial^\beta C_1}{\partial x^\beta} - V_1(x,t)\frac{\partial^\gamma C_1}{\partial x^\gamma} + V_2(x,t)C_1 - D^\alpha_{*t}C_1(x,t)\right]; \quad C_2(x,0) = 0 \tag{23}$$

and so on.

Solving Eq. (21), we get that: $\frac{\partial C_0}{\partial t} = x, \Rightarrow C_0 = xt + k_1$ using the boundary condition $C_0(x,0) = 0$ we get that $k_1 = 0$ then

$$C_0 = xt. \tag{24}$$

Using Eq. (24) on Eq. (22) we get that

$$\frac{\partial C_1}{\partial t} = \left[\frac{\partial (xt)}{\partial t} + ax\frac{\partial^\beta (xt)}{\partial x^\beta} - x\frac{\partial^\gamma (xt)}{\partial x^\gamma} + xt - D^\alpha_{*t}(xt)\right]$$

$$\Rightarrow \frac{\partial C_1}{\partial t} = x - \frac{tx^{2-\gamma}}{\Gamma(2-\gamma)} + xt - \frac{xt^{1-\alpha}}{\Gamma(2-\alpha)}$$

$$\Rightarrow C_1 = xt - \frac{t^2 x^{2-\gamma}}{2\Gamma(2-\gamma)} + \frac{xt^2}{2} - \frac{xt^{2-\alpha}}{(2-\alpha)\Gamma(2-\alpha)} + k_2.$$

Using the condition $C_1(x,0) = 0$ we get that $k_2 = 0$ and therefore

$$\Rightarrow C_1 = xt - \frac{t^2x^{2-\gamma}}{2\Gamma(2-\gamma)} + \frac{xt^2}{2} - \frac{xt^{2-\alpha}}{(2-\alpha)\Gamma(2-\alpha)} \tag{25}$$

Now using Eq. (25) on Eq. (23) we get that:

$$\frac{\partial C_2}{\partial t} = \left[\frac{\partial C_1}{\partial t} + D_1(x,t)\frac{\partial^\beta C_1}{\partial x^\beta} - V_1(x,t)\frac{\partial^\gamma C_1}{\partial x^\gamma} + V_2(x,t)C_1 - D^\alpha_{*t}C_1(x,t)\right]$$

$$\begin{aligned}\Rightarrow \frac{\partial C_2}{\partial t} = {} & \frac{\partial C_1}{\partial t} + \frac{at^2(1-\gamma)(2-\gamma)\Gamma(1-\alpha)}{2\Gamma(2-\gamma)\Gamma(3-\gamma-\beta)}x^{3-\gamma-\beta} \\ & - \frac{tx^{2-\gamma}}{\Gamma(2-\gamma)} + \frac{(2-\gamma)t^2x^{3-2\gamma}}{2\Gamma(3-2\gamma)} \\ & - \frac{t^2x^{2-\gamma}}{2\Gamma(2-\gamma)} + \frac{t^{2-\alpha}x^{2-\gamma}}{(2-\alpha)\Gamma(2-\alpha)\Gamma(2-\gamma)} \\ & - \frac{xt^{1-\alpha}}{\Gamma(2-\alpha)} + \frac{x^{2-\gamma}t^{2-\alpha}}{\Gamma(2-\gamma)\Gamma(3-\alpha)} \\ & + \frac{xt^{2-2\alpha}}{\Gamma(3-2\alpha)} - \frac{xt^{2-\alpha}}{\Gamma(3-\alpha)} + xt - \frac{t^2x^{2-\gamma}}{2\Gamma(2-\gamma)} \\ & + \frac{xt^2}{2} - \frac{xt^{2-\alpha}}{(2-\alpha)\Gamma(2-\alpha)}\end{aligned}$$

$$\begin{aligned}\Rightarrow C_2 = {} & xt - \frac{t^2x^{2-\gamma}}{2\Gamma(2-\gamma)} + \frac{xt^2}{2} - \frac{xt^{2-\alpha}}{(2-\alpha)\Gamma(2-\alpha)} \\ & + \frac{at^3(1-\gamma)(2-\gamma)\Gamma(1-\alpha)}{6\Gamma(2-\gamma)\Gamma(3-\gamma-\beta)}x^{3-\gamma-\beta} \\ & - \frac{t^2x^{2-\gamma}}{2\Gamma(2-\gamma)} + \frac{(2-\gamma)t^3x^{3-2\gamma}}{6\Gamma(3-2\gamma)} - \frac{t^3x^{2-\gamma}}{6\Gamma(2-\gamma)} \\ & + \frac{t^{3-\alpha}x^{2-\gamma}}{(2-\alpha)(3-\alpha)\Gamma(2-\alpha)\Gamma(2-\gamma)} \\ & + \frac{xt^{3-2\alpha}}{(3-2\alpha)\Gamma(3-2\alpha)} - \frac{xt^{3-\alpha}}{(3-\alpha)\Gamma(3-\alpha)} \\ & - \frac{xt^{2-\alpha}}{(2-\alpha)\Gamma(2-\alpha)} \\ & + \frac{x^{2-\gamma}t^{3-\alpha}}{(3-\alpha)\Gamma(2-\gamma)\Gamma(3-\alpha)} + \frac{xt^2}{2} - \frac{t^3x^{2-\gamma}}{6\Gamma(2-\gamma)} \\ & + \frac{xt^3}{6} - \frac{xt^{3-\alpha}}{(2-\alpha)(3-\alpha)\Gamma(2-\alpha)} + k_3\end{aligned} \tag{26}$$

Using the initial condition, we get that $k_3 = 0$, so the approximate analytical solution is as follows: $C = C_0 + C_1 + C_2$, i.e.

$$\begin{aligned}\Rightarrow C(x,t) = {} & xt + 2\left(xt - \frac{t^2x^{2-\gamma}}{2\Gamma(2-\gamma)} + \frac{xt^2}{2} - \frac{xt^{2-\alpha}}{(2-\alpha)\Gamma(2-\alpha)}\right) \\ & + \frac{at^3(1-\gamma)(2-\gamma)\Gamma(1-\alpha)}{6\Gamma(2-\gamma)\Gamma(3-\gamma-\beta)}x^{3-\gamma-\beta} \\ & - \frac{t^2x^{2-\gamma}}{2\Gamma(2-\gamma)} + \frac{(2-\gamma)t^3x^{3-2\gamma}}{6\Gamma(3-2\gamma)} - \frac{t^3x^{2-\gamma}}{6\Gamma(2-\gamma)} \\ & + \frac{t^{3-\alpha}x^{2-\gamma}}{(2-\alpha)(3-\alpha)\Gamma(2-\alpha)\Gamma(2-\gamma)} \\ & + \frac{xt^{3-2\alpha}}{(3-2\alpha)\Gamma(3-2\alpha)} - \frac{xt^{3-\alpha}}{(3-\alpha)\Gamma(3-\alpha)} \\ & - \frac{xt^{2-\alpha}}{(2-\alpha)\Gamma(2-\alpha)} \\ & + \frac{x^{2-\gamma}t^{3-\alpha}}{(3-\alpha)\Gamma(2-\gamma)\Gamma(3-\alpha)} + \frac{xt^2}{2} - \frac{t^3x^{2-\gamma}}{6\Gamma(2-\gamma)} \\ & + \frac{xt^3}{6} - \frac{xt^{3-\alpha}}{(2-\alpha)(3-\alpha)\Gamma(2-\alpha)} + k_3\end{aligned} \tag{27}$$

Case-II: analytical solution with space- and time-dependent dispersion and velocity

In the second case, we considered that velocity and dispersion are dependent on both spatial and temporal domain. Dispersion is directly proportional to the velocity, so it is considered that $V_1(x,t) = xt$, $V_2(x,t) = t$ and $D_1(x,t) = aV_1$. A linear space-dependent source term is considered with the condition that the aquifer is initially contamination free. The modeled equation becomes:

$$\frac{\partial^\alpha C}{\partial t^\alpha} = D_1(x,t)\frac{\partial^\beta C}{\partial x^\beta} - V_1(x,t)\frac{\partial^\gamma C}{\partial x^\gamma} + V_2(x,t)C + G(x,t) \tag{28}$$

with $C(x,0) = 0$.

Using homotopy on Eq. (28) one can write:

$$\frac{\partial C}{\partial t} - G(x,t) = p_1\left[\frac{\partial C}{\partial t} + D_1(x,t)\frac{\partial^\beta C}{\partial x^\beta} - V_1(x,t)\frac{\partial^\gamma C}{\partial x^\gamma} + V_2(x,t)C - D^\alpha_{*t}C(x,t)\right]. \tag{29}$$

We approximate the concentration profile in terms of p_1 then:

$$C(x,t) = C_0 + C_1p_1 + C_2p_1^2 + C_3p_1^3 + \cdots \tag{30}$$

Using the value of C in (29) and equating the equal powers of p_1 we get:

$$\frac{\partial C_0}{\partial t} = x; \quad C_0(x,0) = 0 \tag{31}$$

$$\frac{\partial C_1}{\partial t} = \left[\frac{\partial C_0}{\partial t} + D_1(x,t)\frac{\partial^\beta C_0}{\partial x^\beta} - V_1(x,t)\frac{\partial^\gamma C_0}{\partial x^\gamma} + V_2(x,t)C_0 - D^\alpha_{*t}C_0(x,t)\right]; \quad C_1(x,0) = 0 \tag{32}$$

$$\frac{\partial C_2}{\partial t} = \left[\frac{\partial C_1}{\partial t} + D_1(x,t)\frac{\partial^\beta C_1}{\partial x^\beta} - V_1(x,t)\frac{\partial^\gamma C_1}{\partial x^\gamma} + V_2(x,t)C_1 - D^\alpha_{*t}C_1(x,t)\right]; \quad C_2(x,0) = 0 \tag{33}$$

and so on.

Solving (31), (32) and (33) we get that

$$C_0 = xt \tag{34}$$

$$C_1 = xt - \frac{t^3x^{2-\gamma}}{3\Gamma(2-\gamma)} + \frac{xt^3}{3} - \frac{xt^{2-\alpha}}{(2-\alpha)\Gamma(2-\alpha)} \tag{35}$$

$$\begin{aligned} C_2 = {} & xt - \frac{t^3x^{2-\gamma}}{3\Gamma(2-\gamma)} + \frac{xt^3}{3} - \frac{xt^{2-\alpha}}{(2-\alpha)\Gamma(2-\alpha)} \\ & + \frac{at^4(1-\gamma)(2-\gamma)\Gamma(1-\alpha)}{12\Gamma(2-\gamma)\Gamma(3-\gamma-\beta)}x^{3-\gamma-\beta} \\ & - \frac{t^3x^{2-\gamma}}{3\Gamma(2-\gamma)} + \frac{(2-\gamma)t^5x^{3-2\gamma}}{15\Gamma(3-2\gamma)} \\ & - \frac{t^5x^{2-\gamma}}{15\Gamma(2-\gamma)} + \frac{t^{4-\alpha}x^{2-\gamma}}{(2-\alpha)(4-\alpha)\Gamma(2-\alpha)\Gamma(2-\gamma)} \\ & + \frac{xt^{3-2\alpha}}{(3-2\alpha)\Gamma(3-2\alpha)} - \frac{xt^{2-\alpha}}{(2-\alpha)\Gamma(2-\alpha)} \\ & - \frac{2xt^{4-\alpha}}{(4-\alpha)\Gamma(2-\alpha)} + \frac{x^{2-\gamma}t^{4-\alpha}}{(4-\alpha)\Gamma(2-\gamma)\Gamma(3-\alpha)} \\ & + \frac{xt^3}{3} - \frac{t^5x^{2-\gamma}}{15\Gamma(2-\gamma)} + \frac{xt^5}{15} - \frac{xt^{4-\alpha}}{(2-\alpha)(4-\alpha)\Gamma(2-\alpha)} \end{aligned} \tag{36}$$

and the required solution is as follows:

$$C(x,t) = C_0 + C_1 + C_2,$$

i.e.

$$\begin{aligned} C(x,t) = {} & xt + 2\left(xt - \frac{t^3x^{2-\gamma}}{3\Gamma(2-\gamma)} + \frac{xt^3}{3} - \frac{xt^{2-\alpha}}{(2-\alpha)\Gamma(2-\alpha)}\right) \\ & + \frac{at^4(1-\gamma)(2-\gamma)\Gamma(1-\alpha)}{12\Gamma(2-\gamma)\Gamma(3-\gamma-\beta)}x^{3-\gamma-\beta} \\ & - \frac{t^3x^{2-\gamma}}{3\Gamma(2-\gamma)} + \frac{(2-\gamma)t^5x^{3-2\gamma}}{15\Gamma(3-2\gamma)} \\ & - \frac{t^5x^{2-\gamma}}{15\Gamma(2-\gamma)} + \frac{t^{4-\alpha}x^{2-\gamma}}{(2-\alpha)(4-\alpha)\Gamma(2-\alpha)\Gamma(2-\gamma)} \\ & + \frac{xt^{3-2\alpha}}{(3-2\alpha)\Gamma(3-2\alpha)} - \frac{xt^{2-\alpha}}{(2-\alpha)\Gamma(2-\alpha)} \\ & - \frac{2xt^{4-\alpha}}{(4-\alpha)\Gamma(2-\alpha)} + \frac{x^{2-\gamma}t^{4-\alpha}}{(4-\alpha)\Gamma(2-\gamma)\Gamma(3-\alpha)} \\ & + \frac{xt^3}{3} - \frac{t^5x^{2-\gamma}}{15\Gamma(2-\gamma)} + \frac{xt^5}{15} - \frac{xt^{4-\alpha}}{(2-\alpha)(4-\alpha)\Gamma(2-\alpha)}. \end{aligned} \tag{37}$$

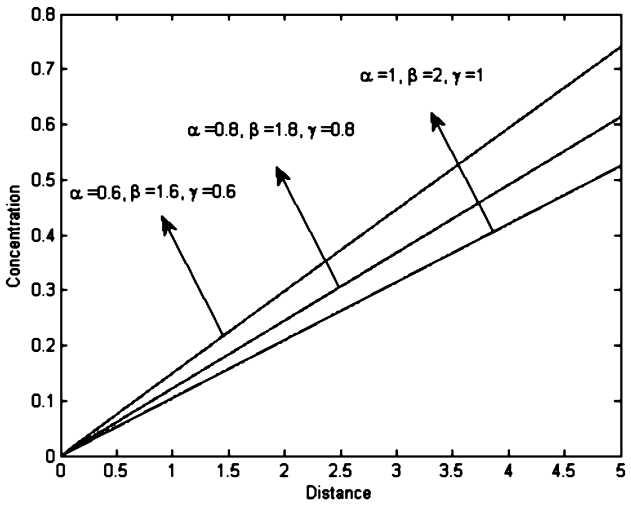


Fig. 1 Concentration with distance with varying fractional-order with the space-dependent velocity and dispersion at a fixed time $t = 0.1$ year

Results and discussion

In this study, the aquifer was assumed initially contamination free. A one-dimensional space–time fractional advection–dispersion equation was used to model solute transport, where the source of contamination was incorporated in the equation. Both dispersion coefficient and velocity were temporally and spatially dependent to reflect the aquifer heterogeneity. The fractional derivative retains the memory and that is why the time-fractional advection–dispersion equation was used to reflect the changes in flow and solute migration. To solve the non-linear ADE, HPM was used to obtain an approximate analytical solution. The different values of fractional order were considered to examine how the analytical solution behaves with the variation of fractional order.

For investigating the effect of order and the kind of memory preserve on the solution, we considered two types of velocity and dispersion profiles: (i) depends on spatial domain and (ii) depends on spatial and temporal domains. From Fig. 1, we observe that the contaminant concentration increased with distance. But the concentration values at each of the positions were decreased with increased fractional order α, β, γ at a fixed time $t = 0.1$ year. Here, the concentration profile with distance shows linear type nature for space-dependent velocity and dispersion. From Fig. 2, we observe that the contaminant concentrations

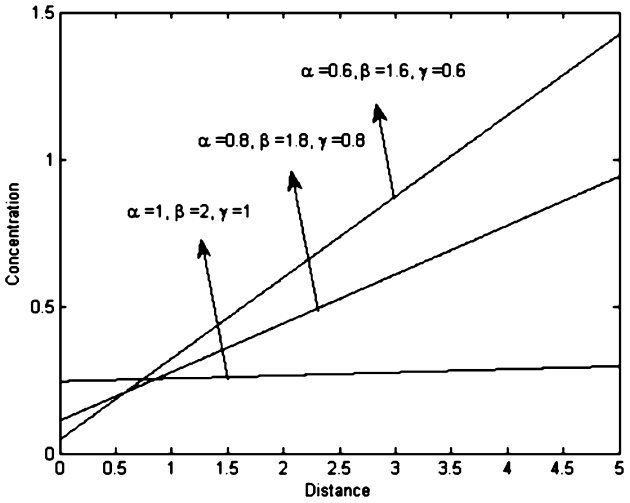


Fig. 2 Concentration with distance with varying fractional order with the space–time dependent velocity and dispersion at a fixed time $t = 0.1$ year

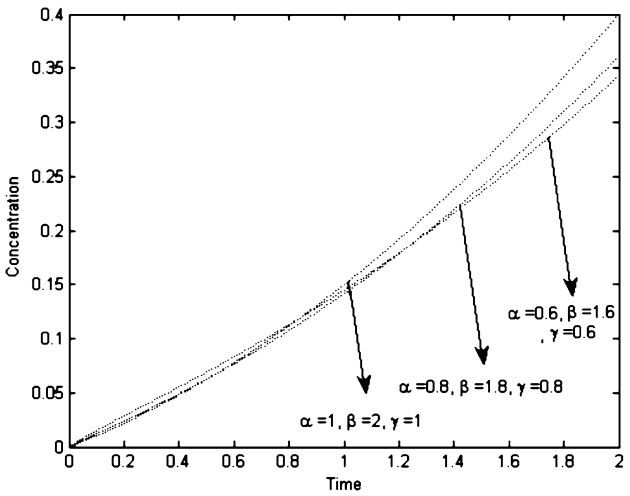


Fig. 3 Concentration with time with varying fractional order with the space-dependent velocity and dispersion at fixed distance $x = 0.2$ km

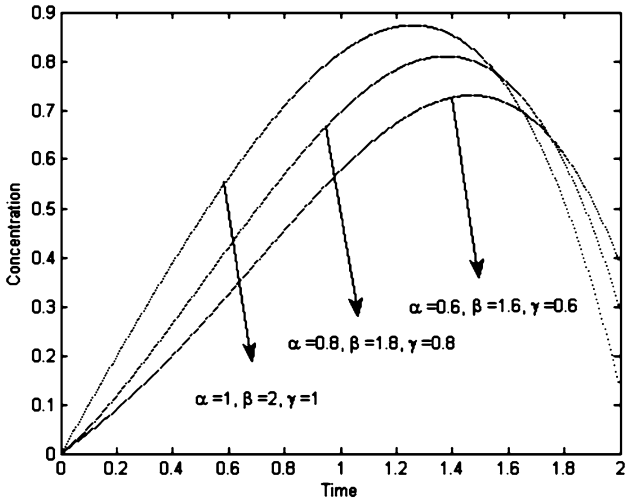


Fig. 4 Concentration with time with varying fractional order with both space- and time-dependent velocity and dispersion at fixed distance $x = 0.2$ km

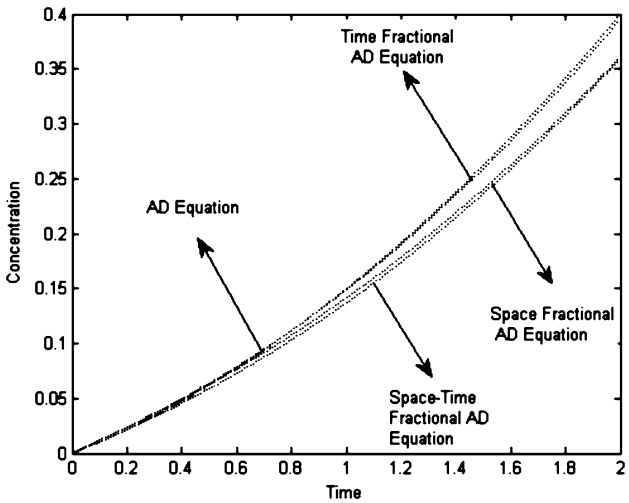


Fig. 5 Concentration with time for various advection–dispersion equations for both space-dependent velocity and dispersion at fixed distance $x = 0.2$ km

increased with increased fractional order α, β, γ at a fixed time $t = 0.1$ year approximately up to distance $x = 1$ km. However, the patterns remain same as in Fig. 1 beyond the distance $x = 1$ km. This may happen due to space–time dependent velocity and dispersion taken into consideration.

From Fig. 3, we observe that the contaminant concentration was increased with time for a fixed distance $x = 0.2$ km. It was also observed that contaminant concentration increased with increasing fractional order. From Fig. 4 it was observed that contaminant concentration initially increases with time and after $t = 1.4$ year, the concentration starts decreasing with time for a fixed distance $x = 0.2$ km. It was also observed that the contaminant concentration values increases with increase in fractional order up to certain time $t = 1.4$ year, but the reverse trend was found beyond the time $t = 1.4$ years.

By solving both space- and time-fractional advection–dispersion equation, we deduced solutions of any type of fractional ADE equation or general ADE, i.e. we can explore only time-fractional advection–dispersion equation by considering space-fractional term as one, space-fractional advection–dispersion equation considering time fractional as one, and general advection–dispersion equation by considering both space and time fractional as one.

In Figs. 5, 6, 7 and 8, we pictorially denote the solution for time fractional, space fractional, space–time fractional and general advection–dispersion equation. From Fig. 5, it was observed that AD equation and time-fractional AD

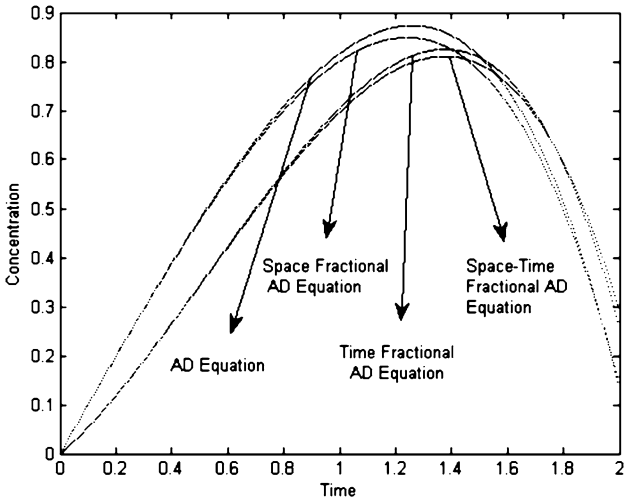


Fig. 6 Concentration with time for various advection–dispersion equations for both space–time dependent velocity and dispersion at fixed distance $x = 0.2$ km

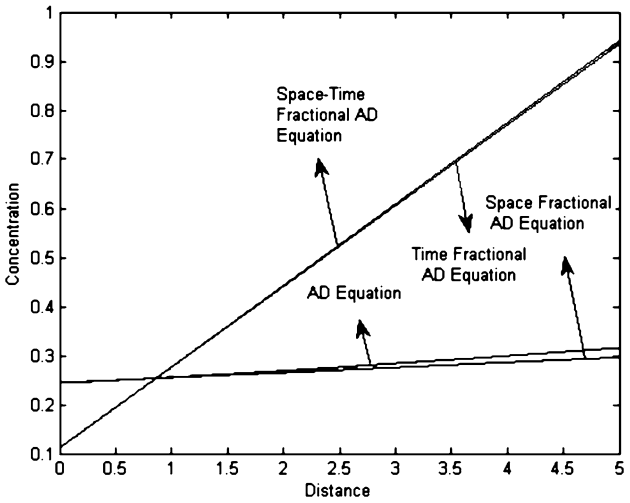


Fig. 8 Concentration with distance for various advection–dispersion equations for both space–time dependent velocity and dispersion at a fixed time $t = 0.1$ year

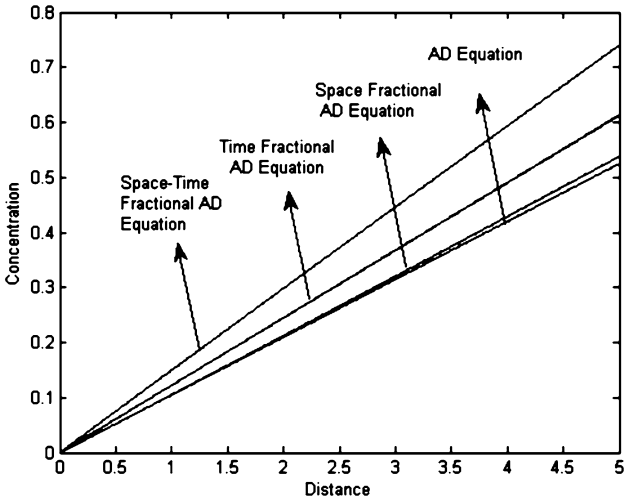


Fig. 7 Concentration with distance for various advection–dispersion equations for both space-dependent velocity and dispersion at a fixed time $t = 0.1$ year

equation show similar impact but space-fractional AD equation and space–time fractional AD equation show similar nature when velocity and dispersion both depend on space. From Fig. 6, it was observed that AD equation and space-fractional AD equation show similar impact, but time-fractional AD equation and space–time fractional AD equation show similar nature when velocity and dispersion both depend on space. From both Figs. 5 and 6 our general observation was that the space–time fractional AD equation shows minimum impact with time for a fixed distance $x = 0.2$ km. From both Figs. 7 and 8 our general observation was that the space-fractional AD equation showed the maximum impact with space for a fixed time $t = 0.1$ km.

Conclusions

The following conclusions are drawn from this study:

1. The HPM yields an approximate analytical solution of the one-dimensional non-linear space–time fractional advection–dispersion equation. The solution can be used to model contaminant transport in a heterogeneous aquifer even without having an idea about the boundary. It may be the most general case as we are not bothered about the real shape and the formation of boundary.
2. The dispersion coefficient and velocity are taken as functions of space and both time and space which may be the most general case for contaminant transport modeling in heterogeneous aquifers.
3. The contaminant concentration with respect to distance increases as the source term is considered as increasing function with distance.
4. The HPM can be made more accurate, depending upon the number of terms considered in the approximate analytical solution. Limitation of this method is that it cannot hold for a large value of time, but we can overcome this problem by substituting time with another variable.
5. Using the fractional-order advection–dispersion equation we may be able to model more complex systems where the domain is differentiable up to some fractional order.
6. Solving this problem we may be able to get any result, such as only space-fractional ADE, only time-fractional ADE and simple ADE. Hence this is a powerful technique for solving this kind of situation.

Acknowledgements The authors are thankful to Indian Institute of Technology (Indian School of Mines), Dhanbad, India, for providing financial support for Ph.D. studies. This work is partially supported by the DST (SERB) Project no. EMR/2016/001628.

References

Batu V (2006) Applied flow and solute transport modeling in aquifers: Fundamental principles and analytical and numerical methods. CRC, Boca Raton, FL

Benson DA, Wheatcraft SW, Meerschaert MM (2000) Application of a fractional advection–dispersion equation. Water Resour Res 36(6):1403–1412

Chen JS, Liu CW (2011) Generalized analytical solution for advection–dispersion equation in finite spatial domain with arbitrary time-dependent inlet boundary condition. Hydrol Earth Syst Sci 15:2471–2479

He JH (2005) Application of homotopy perturbation method to nonlinear wave equations. Chaos Soliton Fractals 26:695–700

Huang Q, Huang G, Zhan H (2008) A finite element solution for the fractional advection dispersion equation. Adv Water Resour 31:1578–1589

Li C, Zhaoa Z, Chenb Y (2011) Numerical approximation of nonlinear fractional differential equations with subdiffusion and superdiffusion. Comput Math Appl 62:855–875

Liu F, Anh V, Turner I (2004) Numerical solution of the space fractional Fokker–Planck equation. J Comput Appl Math 166(1):209–219

Marinca V, Herişanu N, Bota C, Marinca B (2009) An optimal homotopy asymptotic method applied to the steady flow of a fourth-grade fluid past a porous plate. Appl Math Lett 22(2):245–251

Meerschaert MM, Tadjeran C (2004) Finite difference approximations for fractional advection dispersion flow equations. J Comput Appl Math 172:65–77

Meerschaert MM, Scheffler HP, Tadjeran C (2006) Finite difference methods for two-dimensional fractional dispersion equation. J Comput Phys 211(1):249–261

Momani S, Odibat Z (2007) Homotopy perturbation method for non linear partial differential equations of fractional order. Phys Lett A 365:345–350

Murio DA (2008) Implicit finite difference approximation for time fractional diffusion equations. Comput Math Appl 56:1138–1145

Pandey RK, Singh OP, Baranwal VK (2011) An analytic algorithm for the space–time fractional advection–dispersion equation. Comput Phys Commun 182(5):1134–1144

Pandey RK, Singh OP, Baranwal VK, Tripathi MP (2012) An analytic solution for the space–time fractional advection–dispersion equation using the optimal homotopy asymptotic method. Comput Phys Commun 183(10):2098–2106

Paradisi P, Cesari R, Mainardi F, Tampieri F (2001) The fractional Fick's law for non-local transport processes. Phys A 293(1):130–142

Roop JP (2008) Numerical approximation of a one-dimensional space fractional advection dispersion equation with boundary layer. Comput Math Appl 56:1808–1819

Sander GC, Braddock RD (2005) Analytical solutions to the transient, unsaturated transport of water and contaminants through horizontal porous media. Adv Water Resour 28:1102–1111

Schumer R, Benson DA, Meerschaert MM, Wheatcraft SW (2001) Eulerian derivation of the fractional advection–dispersion equation. J Contam Hydrol 48(1):69–88

Schumer R, Meerschaert MM, Baeumer B (2009) Fractional advection–dispersion equations for modeling transport at the earth surface. J Geophys Res Earth Surf 114:F00A07, 1–15. doi:10.1029/2008JF001246

Singh MK, Mahato NK, Singh P (2008) Longitudinal dispersion with time dependent source concentration in semi-infinite aquifer. J Earth Syst Sci 117(6):945–949

Smedt FD (2006) Analytical solution for transport of decaying solutes in rivers with transient storage. J Hydrol 330(3–4):672–680

Srinivasan V, Clement TP (2008) An analytical solution for sequentially coupled one-dimensional reactive transport problems Part-I: mathematical derivations. Water Resour Res 31(2):203–218

Zhang Y, Benson DA, Meerschaert MM, LaBolle EM (2007) Space-fractional advection–dispersion equations with variable parameters: diverse formulas, numerical solutions, and application to the macrodispersion experiment site data. Water Resour Res 43(5):W05439, 1–16. doi:10.1029/2006WR004912

Zheng YY, Li CP, Zhoa ZG (2010) A note on the finite element method for the space fractional advection dispersion equation. Comput Math Appl 59(5):1718–1726

Acta Geophys. (2017) 65:363–375
DOI 10.1007/s11600-017-0033-x

RESEARCH ARTICLE

Direct runoff assessment using modified SME method in catchments in the Upper Vistula River Basin

A. Wałęga[1] · A. Rutkowska[2] · M. Grzebinoga[3]

Received: 8 March 2017 / Accepted: 16 March 2017 / Published online: 24 March 2017

Abstract Correct determination of direct runoff is crucial for proper and safe dimensioning of hydroengineering structures. It is commonly assessed using SCS-CN method developed in the United States. However, due to deficiencies of this method, many improvements and modifications have been proposed. In this paper, a modified Sahu–Mishra–Eldo (SME) method was introduced and tested for three catchments located in the upper Vistula basin. Modification of SME method involved a determination of maximum potential retention S based on CN parameter derived from SCS-CN method. The modified SME method yielded direct runoff values very similar to those observed in the investigated catchments. Moreover, it generated significantly smaller errors in the direct runoff estimation as compared with SCS-CN and SME methods in the analyzed catchments. This approach may be used for estimating the runoff in uncontrolled catchments.

Keywords Direct runoff · SCS-CN method · SME method · Potential retention · Flexibility

✉ A. Wałęga
a.walega@ur.krakow.pl

[1] Department of Sanitary Engineering and Water Management, University of Agriculture in Kraków, ul. Mickiewicza 24/28, 30-059 Kraków, Poland

[2] Department of Applied Mathematics, University of Agriculture in Kraków, ul. Balicka 253C, 30-198 Kraków, Poland

[3] MGGP S.A., ul. J. Lea 112, 30-133 Kraków, Poland

Introduction

Direct runoff is one of the key parameters used at the design stage of hydraulic engineering projects associated with flood protection. The direct runoff, perceived as a surface runoff formed after allowing for precipitation interception, infiltration, surface retention and evapotranspiration, is also an input for the majority of hydrological models (Banasik et al. 2014a; Bedient et al. 2013). Therefore, its proper estimation is extremely important. Its size depends on the course of a precipitation event, evapotranspiration, substrate permeability, soil moisture, etc. (Banasik et al. 2014b; Darvishan et al. 2015; Deshmukh et al. 2013; Ponce and Hawkins 1996; Rallison and Miller 1981; Rutkowska et al. 2015). In catchment hydrology, the soil conservation service-curve number (SCS-CN) method is one of the most popular methods for computing the volume of surface runoff for a given rainfall event from small agricultural watersheds (Mishra et al. 2005). Given its perceived advantages (simplicity, predictability and stability), this method has been widely used in many countries and successfully applied to runoff calculation, to land use change assessment and to comprehensive hydrological/water quality simulations (Hong and Adler 2007). The main reason for its widespread use is an adoption of the basic characteristics of a catchment, i.e., soil type, land use, and antecedent moisture conditions as the primary determinants of the runoff (Mishra and Singh 2002, 2003; USDA Natural Resources Conservation Service 2004). The main equations are:

$$Q = \frac{(P - I_a)^2}{P - I_a + S} \quad if \quad P > I_a \text{ and } Q = 0 \text{ otherwise} \tag{1}$$

$$S = \frac{25,400}{\mathrm{CN}} - 254, \tag{2}$$

where Q, P, S, and CN are the direct runoff, the rainfall depth, the retention parameter, and the curve number, respectively, and $I_a = \lambda S$ is the initial abstraction ratio (usually λ is assumed as 0.2). Q, P and S are in [mm], while CN is dimensionless. In the original SCS-CN method, constant values of S and CN are assumed, being functions of the soil type and land use. The tabulated value, CN_{tab}, is determined on the basis of the National Engineering Handbook, "Methods" (NEH4) tables. On the other hand, the use of the SCS-CN method results in the abrupt changes of the calculated direct runoff, depending on the assumed moisture content for the catchment (AMC I, AMC II or AMC III) (Hawkins et al. 2009). This may be regarded as a limitation of the method since the runoff continuity with changes in the moisture content for the catchment is not preserved. Therefore, the method is not used for predicting freshets or in continuous long-term series simulations. Its ability to simulate only single rainfall–runoff events may result in a higher uncertainty of the results of calculations. As reported by Calver et al. (2009), the emerging continuous simulation approaches show a considerable potential to offer better performance for peaks and flow time series than event simulation. According to Hoes and Nelen (2005), the use of a continuous rainfall–runoff simulation still results in an uncertainty. However, many assumptions and simplifications of the event-based approach can be excluded because they are the direct or indirect components of the precipitation input series. Another feature of the SCS-CN method is that the time variable is not taken into account in formula (1). This limits the method's applicability in runoff variability simulations for rainfall events, as it only enables the calculation of the direct runoff layer volume for the specified rainfall depth. Unfortunately, the original formula is often applied during a precipitation event. Furthermore, this method has some additional limitations such as a lack of a clear guidance on how to vary AMC, and a lack of an explicit relation between the initial abstraction and the antecedent moisture (Sahu et al. 2012). For this reason, some modifications of SCS-CN method have been introduced. Soulis and Valiantzas (2012) and Wałęga et al. (2015) used a two-CN heterogeneous system for calculating a catchment runoff. Singh et al. (2015) proposed a continuous function of changes in the substrate moisture content as a modification of the original method. Petroselli et al. (2013) and Grimaldi et al. (2013a, b) modified the SCS-CN model by taking the Green and Ampt infiltration method into account, to get a correct distribution of the direct runoff during a precipitation event. Particularly interesting reports were published by Sahu et al. (2010, 2012) and Sahu et al. (2007), who allowed for antecedent moisture in their calculations of the direct runoff. In this model (called Sahu-Mishra-Eldo SME method, after the authors), λ is the initial abstraction, S_0 is the absolute maximum retention capacity and M is the antecedent moisture that is assumed to be a fraction (β) of the amount of moisture/water infiltrated during the antecedent period and to depend on the 5-day antecedent rainfall amount P_s. Such an assumption is based on the fact that only a fraction, in general, of the moisture/water added to the soil profile will contribute to M because of evapotranspiration, drainage, and other losses in the antecedent 5-day period. Additionally, $M = 0$ for $P_s < \lambda S_0$. As S_0 is the potential maximum retention in completely dry condition in SME method, it is independent of the antecedent moisture and fully depends on catchment characteristics (Sahu et al. 2010). A very good performance of the model was shown (Sahu et al. 2010). SME method was verified by Wałęga and Rutkowska (2015) in a mountainous catchment of the upper Vistula basin.

The use of the original SME method in moderate climatic conditions is limited by the need to calibrate as many as three parameters, S_0, λ and β, whereas only a single CN parameter that depends on the type of soil and the structure of catchment land use is applied in the original SCS-CN method. This may limit the use of the SME method, particularly in uncontrolled catchments. Given those limitations, the present authors chose to modify the original SME method relative to the S_0 parameter of which the value was based on the parameter CN according to the NEH4 tables. Moisture content in the catchments in moderate climate conditions may, on very rare occasions, be extremely low and the absolute maximum retention, S_0, may then be a rather unreliable parameter. A novelty in the present study is that, unlike in the approach reported by Sahu et al. (2010), the value of the parameter S_0 as S was adopted in the SCS-CN method for mountain and sub-mountain catchments in the moderate climate. As mentioned earlier in this paper, the approach will limit the number of parameters of the SME method to two items which need to be calibrated, thus reducing the risk of uncertainty of the model. This is of particular importance in conditions of high uncertainty of measurement data, especially flows, based on which parameters are calibrated (Di Baldassarre et al. 2010; Domeneghetti et al. 2013), and of which the uncertainty may, consequently, contribute to erroneous values of the model parameters. An argument for choosing the approach was earlier studies reported by Kowalik and Wałęga (2015) as well as Wałęga et al. (2015) and Banasik et al. (2014a), performed in agricultural catchments in Poland: those reports demonstrated that, as a rule, the empirical values of the parameter CN correspond to the values of CN as reported by USDA, especially for AMC II. This means that the parameter S in the SME method for mountain and sub-mountain catchments in the south of

Poland may be determined by the SCS-CN method. Moreover, studies on the original method were performed for small agricultural catchments not larger than some 70 ha in the United States (Sahu et al. 2010) or in agricultural-and-forested catchments occupying a maximum of 250 km^2 and located in the humid climate (Sahu et al. 2012). For that reason, the present study on agricultural-and-forested catchments of an area of 330 km^2 or less, located in the humid climate, is a unique contribution to studies on the method in question.

The aim of this work was to modify the SME method by replacing the S_0 parameter with the maximum potential retention S parameter based on original SCS-CN method and assessment of the sensitivity of the new model to changes in parameters in three rivers located in the upper Vistula basin.

Catchment description

All three rivers are located in the water region of the upper Vistula River. The localization and the hydrographic network are presented in Fig. 1.

The Kamienica is a right tributary of the Dunajec. The catchment is mountainous with its area down to Łabowa gauging station of 64.878 km^2. Average catchment exposure was determined as southern (181.3°). The catchment is covered mainly by forests (ca. 76.4%), predominantly coniferous ones. Arable lands take up ca. 23.2%, and the rest (ca. 0.4%) is an anthropogenic area.

The Stobnica is a right tributary of the Wisłok. The catchment has an area of 328.635 km^2 to the Godowa gauging station. Average catchment exposure was determined as eastern (91.16°). The Stobnica catchment is covered mainly by arable lands (ca. 62.3%). About 33.7% of the area is covered by forests, mainly mixed and coniferous forests. The rest (ca. 4%) is an anthropogenic area.

The Rudawa is a right tributary of the Vistula. This upland catchment has an area of 282.664 km^2 to the Balice gauging station. Average catchment exposure is southern (168.1°). Most area is covered by arable lands (ca. 65%). A large part (ca. 26.6%) is covered by forests, dominated by deciduous species. The rest (ca. 8.4%) is an anthropogenic area.

Basic catchment characteristics are presented in Table 1. Water resources are mainly surface water, like in the entire Polish Carpathians (Policht-Latawiec et al. 2016).

Material

Initial material for the analyses comprised the data on precipitation and runoff in the investigated catchments obtained from the Institute of Meteorology and Water Management National Research Institute in Warsaw. The data were collected for the multi-year period 1980–2012 for a time step of 24 h for both the precipitation and the flows. The flow size was registered at the water gauges

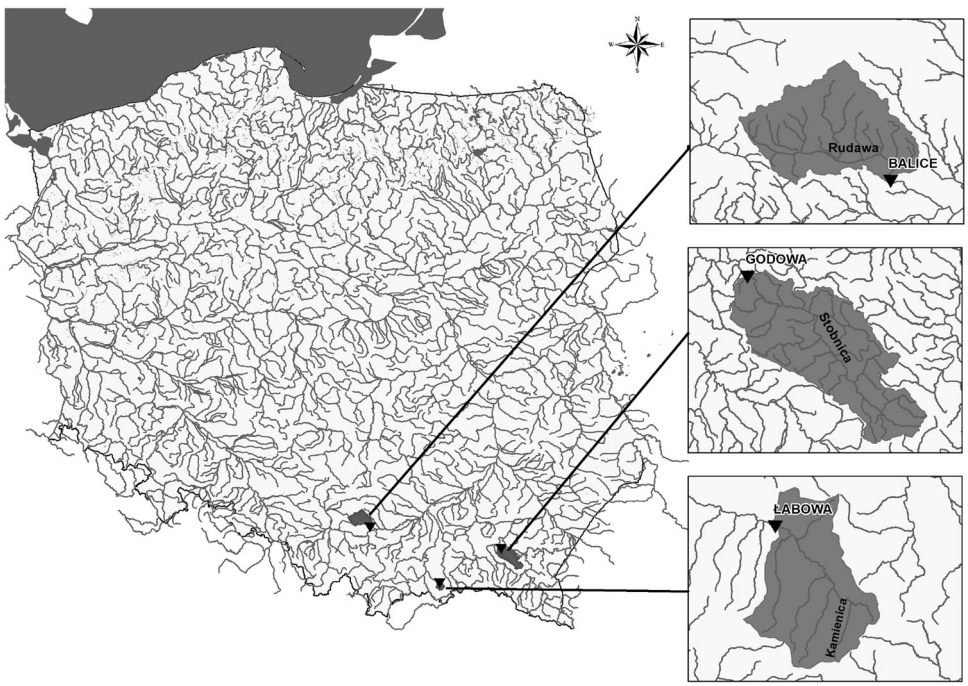


Fig. 1 Location of the investigated catchments

Table 1 Basic characteristics of the analyzed catchments

Characteristic	Unit	The Kamienica catchment	The Stobnica catchment	The Rudawa catchment
Area	km^2	64.878	328.635	282.664
Length of main river	km	13.359	43.072	26.006
Length of river network	km	42.149	209.958	137.157
Density of river network	km km^{-2}	0.650	0.639	0.485
Average altitude of catchment	m a.s.l.	695	335	351
Average slope of catchment	%	29.42	19.50	11.53

closing the investigated catchments, i.e., Łabowa for the Kamienica, Godowa for the Stobnica, and Balice for the Rudawa. The analyses included the rivers differing in their river network density, catchment slope and land use. These factors play an important role in shaping the catchment runoff and the time of its response to precipitation. The initial data on flow size were used to identify the floods constituting the basis for further analysis. The selected floods were characterized by peak flow that exceeded the multi-year mean flow for the summer half-year. An additional criterion for identifying the greatest floods was a condition that the lowest peak flow had to be by at least 10% higher than the mean flow in the summer half-year. Flood selection criteria were determined according to Ciepielowski (1987). The rainfall–runoff events investigated for the various catchments were: 30 in the catchment of Kamienica, 15 of Rudawa and 14 of Stobnica.

The events were visibly distinguishable in the records of river discharge and independent of other events. The base flow and the direct runoff were separated before the actual analysis. This was done by drawing a straight line on a hydrograph from the point where the flow starts to increase to the point on the descending part where the direct runoff ends. The second point was specified on the semilogarithmic plot of discharge (Dingman 2008). This procedure allowed us to determine the actual amount of the direct runoff layer for individual episodes. The present authors chose not to separate the direct runoff from the base flow by means of advanced methods such as recursive filtering (Serinaldi and Grimaldi 2011), since the methods involved subjectivity in adopting parameters, which means that it is not possible to establish the set of parameters leading to a satisfactory result. Hingray et al. (2014) reported that satisfactory criteria for the estimation of filter parameters are, in fact, highly subjective and the estimation of the parameters is consequently delicate. The precipitation data were spatially averaged using Le Thiessen method. The data preparation process resulted in a sample $(P_i \ldots Q_i)$, $i = 1, \ldots N$, where: P_i—the observed total amount of precipitation (rainfall depth) causing i—the flood [mm], Q_i—the observed direct runoff [mm]. Additionally, P_{5i}—the amount of the antecedent 5-day rainfall was measured for every rainfall–runoff event.

Methods

The modified SME method

The modified SME method consists of several stages. In the first stage, the tabulated CN values (CN_{tab}) are determined on the basis of NEH4 tables depending on soil type and land use (USDA 2004). These values, equal to CN(II), represent moderate antecedent runoff conditions. Additionally, CN(I) and CN(III) are calculated using the formulas of Hawkins for dry and wet runoff conditions, respectively (ASCE 2009). Next, using the original SCS-CN Eqs. (1), (2), the graph of Q (p) is drawn for P from P_{min} to P_{max} (P_{min} and P_{max} were the minimum and maximum values in the rainfall depth sample), while S is equal to S(I), S(II) or S(III) for dry, moderate and wet conditions, respectively, and $I_a = 0.2S$. Finally, three curves are obtained.

In the second stage, each (P_i, Q_i) pair is associated with a curve the most closely located to the point (P_i, Q_i) among the three curves obtained in the first stage, on a two-dimensional space in the sense of the Euclidean distance. The value of S corresponding to the curve obtained for most of the events is used in further calculations.

In the third stage, the direct runoff is estimated for every rainfall–runoff event using SME method (Sahu et al. 2010, 2012):

$$Q = \frac{(P - I_a) \cdot (P - I_a + M)}{P - I_a + S} \quad if \quad P > I_a \tag{3}$$

$$Q = 0 \; otherwise$$

$$I_a = \lambda \cdot (S - M) \tag{4}$$

where S is the maximum retention capacity obtained in the second stage. In the Eq. (3), the higher the antecedent moisture M, the lower the initial abstraction I_a, and vice versa. The antecedent moisture is given as:

$$M = \beta\left[\frac{(P_5 - \lambda S)\cdot S}{(P_5 - \lambda S) + S}\right] \quad for \quad P_5 > \lambda S \tag{5}$$

$$M = 0 for\, P_5 < \lambda S$$

where β is the parameter to be optimized and the parameter λ is equal to 0.20 (as in the first stage). The parameter β is a part of the 5-day rainfall that is intercepted. In this paper, the condition of the highest value of the Nash–Sutcliffe efficiency coefficient (8) is used for optimization. The input parameters β, ranging from 0.001 to 0.2, were considered in the optimization process.

The main difference between the original SME and the proposed modified SME method refers to the way the parameter S is calibrated. In the original SME method, absolute potential catchment retention, S_0, was assumed to be a new parameter which is functionally related to the catchment characteristics. As a result, the approach was entirely independent of the SCS-CN method and the authors were able to circumvent its limitations, for example, those connected with sudden jumps in the estimated direct runoff values relative to changes in the moisture content of the catchment. The parameter S is optimized with respect to the minimum of the least squares between the observed and estimated runoff depths in the original SME method. A different approach is adopted in this paper: the SCS-CN and SME method were combined: S_0 was substituted with the maximum potential retention, S, established from formula (2). It was assumed that, in moderate climate catchments, which are rather rich in water—such as those investigated and described in this paper—the absolute maximum retention values higher than S are not typical of catchments in the region of interest. Therefore, it was decided that the original parameter, S_0, would be substituted with S. Moreover, according to Sahu et al. (2010), the values of S_0 may equal or be higher than S; therefore, the authors adopted the less desirable variant, $S_0 = S$, in this work. This results in lower rainfall losses in the model and generates higher runoff values, in comparison with the original SME method. This is crucial, especially for the dimensioning of hydroengineering structures, designed to prevent floods. Moreover, the present authors' assumption regarding the parameter S allows for some degree of flexibility in the calculation of design runoff values, which enables moisture conditions in a given catchment to be differentiated depending on the assessed runoff having a specified exceedance probability [e.g., runoffs with a lower exceedance probability may occur in high soil moisture conditions before a freshet (Krzanowski et al. 2013)]. The applicability of the approach was described by De Paola et al. (2013).

The parameter β is to be calibrated in controlled catchments. Then, a regionalization of this parameter should be accounted for, to enable the use of the modified SME method in uncontrolled catchments, assuming that λ is 0.20 (USDA guidelines, USDA 2004).

Additionally, the direct runoff was estimated in this paper using two other methods, namely the original SCS-CN [Eqs. (1), (2)] and the original SME method [Eqs. (3), (4)]. Regarding the SCS-CN method, S equal to S(I), S(II) or S(III) was assumed, according to SCS guidelines, based on the amount of precipitation within 5 days prior to the analyzed flood.

Quality of the modified SME method in selected catchments

The root mean square error (RMSE), the relative bias (BIAS) and EF (Nash and Sutcliffe 1970) were used as goodness of fit measures to assess and compare the performance of the models, namely

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\cdot\sum_{i=1}^{N}\left(Q_{\mathrm{obs,i}} - Q_{\mathrm{calc,i}}\right)^2} \tag{6}$$

$$\mathrm{BIAS} = \frac{1}{N}\sum_{i=1}^{N}\frac{Q_i - Q_{\mathrm{calc,i}}}{Q_i}\cdot 100\% \tag{7}$$

$$\mathrm{EF} = 1 - \frac{\sum_{i=1}^{N}\left(Q_i - Q_{\mathrm{calc,i}}\right)^2}{\sum_{i=1}^{N}\left(Q_i - \overline{Q_{\mathrm{obs,i}}}\right)^2} \tag{8}$$

where Q_i is the observed direct runoff, $Q_{\mathrm{calc,i}}$ is the direct runoff calculated using formula (3) and $\overline{Q_{\mathrm{obs}}}$ is the mean value of the observed direct runoff values. These coefficients are often used in hydrological models to assess the discordance between the observed and estimated values. The higher the RMSE, the poorer the performance of a model, and vice versa. EF = 1 and RMSE = 0 show a perfect fit. EF has the advantage of being dimensionless that enables comparisons of various models and of being valid for nonlinear as well as linear models (McCuen 2003). Regarding BIAS, it is the mean relative difference between the observed and estimated values. High positive/low negative BIAS is a sign of a systematic overestimation/underestimation of the direct runoff by a model.

Sensitivity of modified SME method

For the assessment of the modified SME method properties, its sensitivity to input parameter changes was also derived using the coefficient of flexibility. The flexibility is an indicator of a relative change in output model properties resulting from a relative change in an input parameter. Theoretically, if $x_1, \ldots, x_n$ are model input variables and $y = y(x_1, \ldots, x_n)$ is model output, then the flexibility (relative sensitivity) of y with respect to ith variable at $(x_1, \ldots x_i^0, \ldots, x_n)$ is equal to

$$e(y) = \frac{\partial y}{\partial x_i}(x_1, \ldots x_i^0, \ldots, x_n) \cdot \frac{x_i^0}{y\,(x_1, \ldots x_i^0, \ldots, x_n)} \quad (9)$$

The absolute value of e equal to or greater than 1 is a sign of a substantial change of output value to the input value; otherwise, the model cannot be deemed flexible (Maidment and Hoogerwerf 2002; Wałęga et al. 2012, 2014).

Analysis of flexibility was performed in two stages. The impact of S, P_5, β, and λ on the antecedent moisture, M, and the impact of P, S, λ, M, and P_5 on the runoff value, Q, were assessed in the first and second stage, respectively. Regarding the flexibility of the modified SME method to parameters, the values of the other parameters were assumed to remain constant. The flexibility was calculated for the Stobnica catchment. The present authors chose to analyze the sensitivity of the model for the one catchment only. They assumed, by analogy, that similar reactions which are observed in the analyzed model will occur in other catchments as well. Since the sensitivity analysis covered, first of all, relationships between parameters in the model, and the calculations themselves were carried out for a broad range of values of the input and output variables—which were observed also in the other catchments investigated—the approach was believed to be appropriate.

Results and discussion

Results of the modified SME method

Tabular values of CN_{tab} for three antecedent moisture conditions were determined from SCS tables (USDA 2004). For the Kamienica and the Stobnica catchments, CN_{tab} was 81.0, and for the Rudawa catchment CN_{tab} was 72.6. Regarding the antecedent moisture conditions, the following values were obtained using the Hawkins formulas (ASCE 2009): $CN_I = 69.7$, $CN_{III} = 90.9$ (the Kamienica, the Stobnica) and $CN_I = 56.8$, $CN_{III} = 86.1$ (the Rudawa).

The basic statistical characteristics for the investigated rainfall–runoff events are shown in Table 2. The variability of parameters that characterize the rainfall–runoff events, reflected in the coefficient of variation, was somewhat higher in the upland catchment of Stobnica. The lowest values were obtained in the mountainous catchment of Kamienica. This is probably due to the extensive forest cover (over 76% of the catchment area) that increases the rainfall loss for interception, evapotranspiration or infiltration. Catchment land use affects its retention capacity and the formation of surface runoff, which is reflected in the values of CN_{obs}. It was assumed that the values of CN_{obs} were the empirical values of CN for the respective rainfall–runoff events. They can be seen in Table 2. The highest mean value of CN_{obs} (92.5) was recorded in the catchment Stobnica that is dominated by arable lands (over 62% of the area) and where forests cover nearly 34% of the area. Although the share of arable lands in the Rudawa catchment is similar to the Stobnica catchment, its geological structure, and particularly the presence of karst areas, increases water loss for infiltration and this causes a drop in the CN values. As mentioned before, even though the slopes of the Kamienica catchment are the steepest, the forest cover makes up for it and CN parameter is only slightly lower than in the upland Stobnica. The data displayed in Table 2 clearly indicate that the CN_{tab} values according to SCS were significantly lower than the observed ones. Similar conclusions were published by Banasik and Woodward (2010), Banasik et al. (2014b), Ebrahimian et al. (2012), Kowalik and Wałęga (2015), or Rutkowska et al. (2015). Summing up, estimating the amount of direct runoff in uncontrolled catchments based on the original SCS method may result in its underestimation and consequently adopting too low flow values during design of hydroengineering structures. This confirms the need to search for new or modifications of the existing methods for direct runoff estimation.

Table 3 shows the values of the parameters β and S for the modified SME method.

In the upland catchments of the Stobnica and the Rudawa, the values of β were higher than in the mountainous catchment of the Kamienica. As demonstrated by formula (5), this parameter affects the antecedent moisture M. Therefore, its higher value in the upland catchments may suggest slightly higher soil moisture content M in the period preceding the flood. This was obviously due to a slight catchment slope and higher substrate permeability that resulted in an increased retention of rainwater in the soil profile. Thus, the upland catchments would show a smaller initial abstraction as determined by the formula (4). Substantial changes in S parameter were observed in all the catchments, and S was the highest in the catchment of the Rudawa, which was undoubtedly due to its geological structure, upland nature, and highly anthropogenic character of some catchment areas.

Performance of the modified SME method

Table 4 presents the quality assessment of the modified SME method with regard to the observed runoff based on RMSE and EF metrics. Moreover, the runoff for S calculated from CN_{tab} (the USDA tables) was evaluated as per SME method. The results were compared with those obtained with the original SCS-CN method. The observed events clearly showed very high EF values close to 1.00 and low RMSE values. The only exception was the Rudawa catchment, where EF was the lowest and

Table 2 Characteristics of the selected rainfall–runoff events and CN_{obs}

Parametrer	Q [mm]	P [mm]	P_5 [mm]	CN_{obs} [−]
The Kamienica catchment				
Average	29.4	54.7	16.7	88.3
Median	27.7	54.8	9.0	88.5
Standard deviation	14.6	20.1	17.1	8.6
Minimum	2.8	20.8	0.0	64.0
Maximum	81.1	108.6	61.7	99.0
Coefficient of variability	0.5	0.4	1.0	0.1
The Stobnica catchment				
Average	25.9	41.9	15.4	92.5
Median	18.9	37.2	10.9	94.0
Standard deviation	19.0	20.3	18.5	6.6
Minimum	8.8	16.5	0.0	73.1
Maximum	72.4	80.4	66.3	98.3
Coefficient of variability	0.7	0.5	1.2	0.1
The Rudawa catchment				
Average	9.7	39.6	49.6	81.4
Median	7.9	40.5	30.9	80.8
Standard deviation	5.6	17.4	44.0	11.6
Minimum	3.4	11.6	1.6	61.8
Maximum	22.4	77.6	148.4	98.5
Coefficient of variability	0.6	0.4	0.9	0.1

Table 3 Parameters of the modified SME method in the investigated catchments

Catchment	β	$S_{ave}/S_{min}/S_{max}$
The Kamienica catchment	0.11	36.6/2.6/142.9
The Stobnica catchment	0.22	22.1/4.4/93.5
The Rudawa catchment	0.15	64.2/3.9/157.0

$S_{ave,min,max}$ – average, minimum and maximum value of S

amounted to 0.95. BIAS also indicated a very small error of the model for the Kamienica and the Stobnica catchments but significantly greater (−8.75%) for the Rudawa catchment. Underestimation of the runoff yielded from SME method as compared with the observed runoff was perceived in all the analyzed catchments. When S in the SME method was calculated from CN_{tab}, the quality of the model results markedly deteriorated. Despite this, the Kamienica and the Stobnica catchments were still characterized by relatively low RMSE and EF values. However, BIAS was clearly higher, and systematic overestimation of the calculations was observed for all cases. The worst results were obtained for the Rudawa catchment, where BIAS was 25.80%. Calculations performed according to the original SCS-CN method yielded the best outcomes for the Kamienica and the Stobnica catchments (EF above 0.70), and the worst for the Rudawa catchment. A comparison of the model performance for S based on CN_{tab} clearly showed less satisfactory results of the SCS-CN than that of the SME method. The possible cause behind better properties of the SME than of the SCS-CN method is that more parameters are used in the former that reflect the retention magnitude, for example the M parameter.

Assuming the criteria presented by Moriasi et al. (2007), the modified SME method is very reliable (EF > 0.75) for all catchments. The SME method with S derived from the

Table 4 Quality assessment of the modified SME and SCS-CN

Catchment	Modified SME method			SME method (S calculated for CN_{tab} from SCS-CN method)			SCS-CN model		
	RMSE [mm]	EF [−]	BIAS [%]	RMSE [mm]	EF [−]	BIAS [%]	RMSE [mm]	EF [−]	BIAS [%]
The Kamienica catchment	0.84	0.99	−1.64	6.95	0.72	9.81	7.63	0.72	10.64
The Stobnica catchment	1.00	0.98	−2.80	7.89	0.82	17.50	8.22	0.80	19.95
The Rudawa catchment	1.20	0.95	−8.75	4.74	0.24	25.80	4.92	0.18	32.59

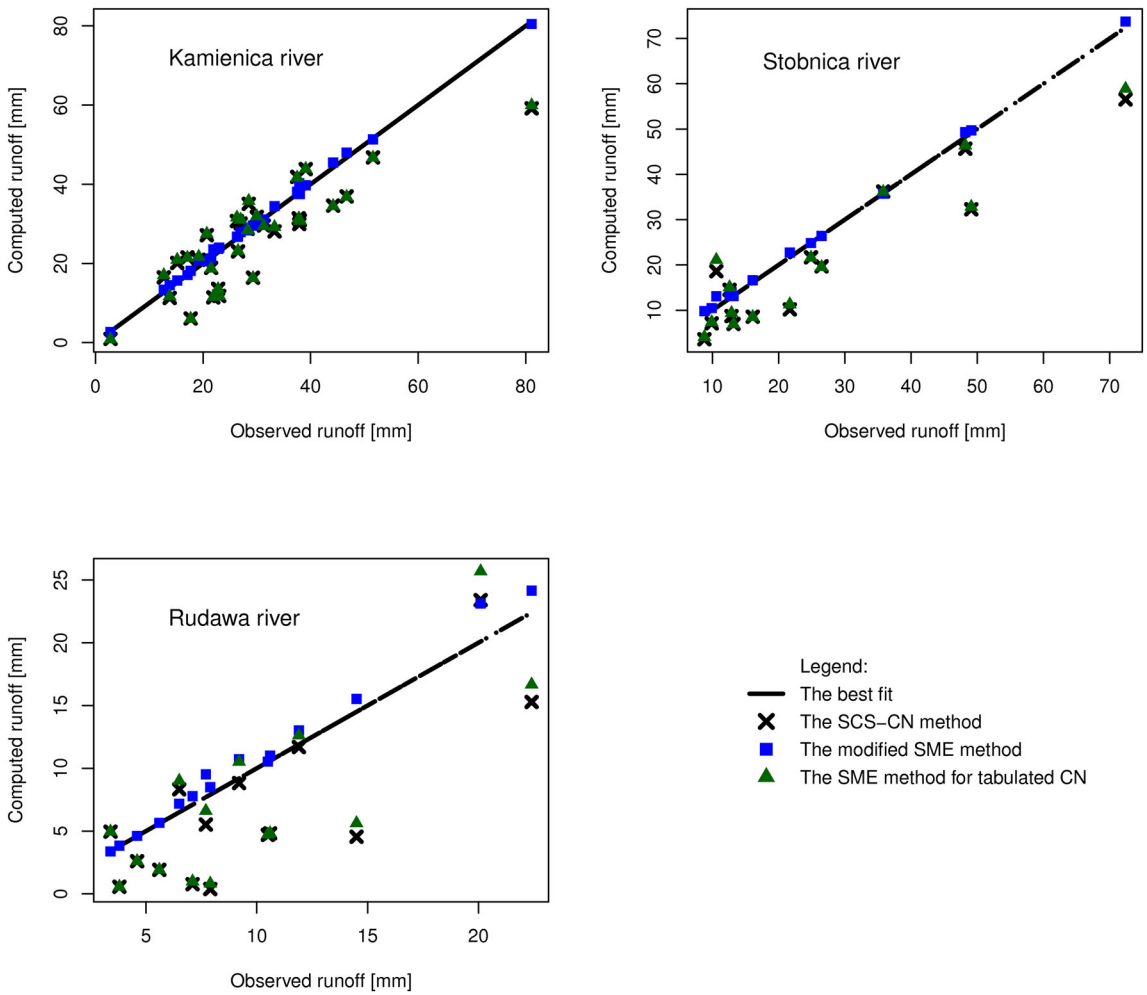


Fig. 2 Plots of the computed runoff versus the observed runoff. The use of the modified SME method, SME method based on CN_{tab} and the original SCS-CN model for three investigated catchments

CN_{tab} showed satisfactory results in two catchments. In the Kamienica catchment, it was found reliable (because $0.65 < EF \leq 0.75$) and in the Stobnica catchment it was very reliable. The results from the SME method were unsatisfactory ($EF < 0.50$) only for the Rudawa catchment. The original SCS-CN model is reliable according to the RMSE and EF metrics and unsatisfactory only for the Rudawa catchment. These results justify the use of the modified SME method and confirm the opportunity of using it for the evaluation of the surface runoff in uncontrolled catchments when supported by the components of the original SCS-CN method. In Fig. 2, the relationships between the estimated and observed direct runoff are presented. Very high compatibility was evidenced between the observed values and those computed with the modified SME method. Compatibility of SME method based on the CN_{tab} was slightly lower while the lowest value was obtained for the original SCS-CN method. This confirmed the best performance of the modified SME method out of the three investigated approaches. Figure 3 shows the proposed flowchart for determination of the direct runoff according to the modified SME method.

Sensitivity of modified SME method

To assess sensitivity of the modified SME method, flexibility of M and Q was computed with respect to S, P_5, λ, β, and S, λ, M, P, P_5, β, respectively, while derivatives of M and Q were calculated using differential calculus (see "Appendix"). Table 5 presents the values of flexibility for the modified SME method with the parameters determined for the event that caused the greatest runoff, namely $P = 80.4$, $P_5 = 40.5$, $\beta = 0.22$, $S = 7.31$, $P_5 = 40.5$, and $\lambda = 0.20$.

The parameter β was the only parameter for which M was flexible. It equaled to 1 because M depends linearly on β. High sensitivity of SME method to P was visibly reflected in high value of Q to P flexibility that was greater than 1. This was due to the fact that precipitation was the main factor controlling the surface runoff. The flexibility

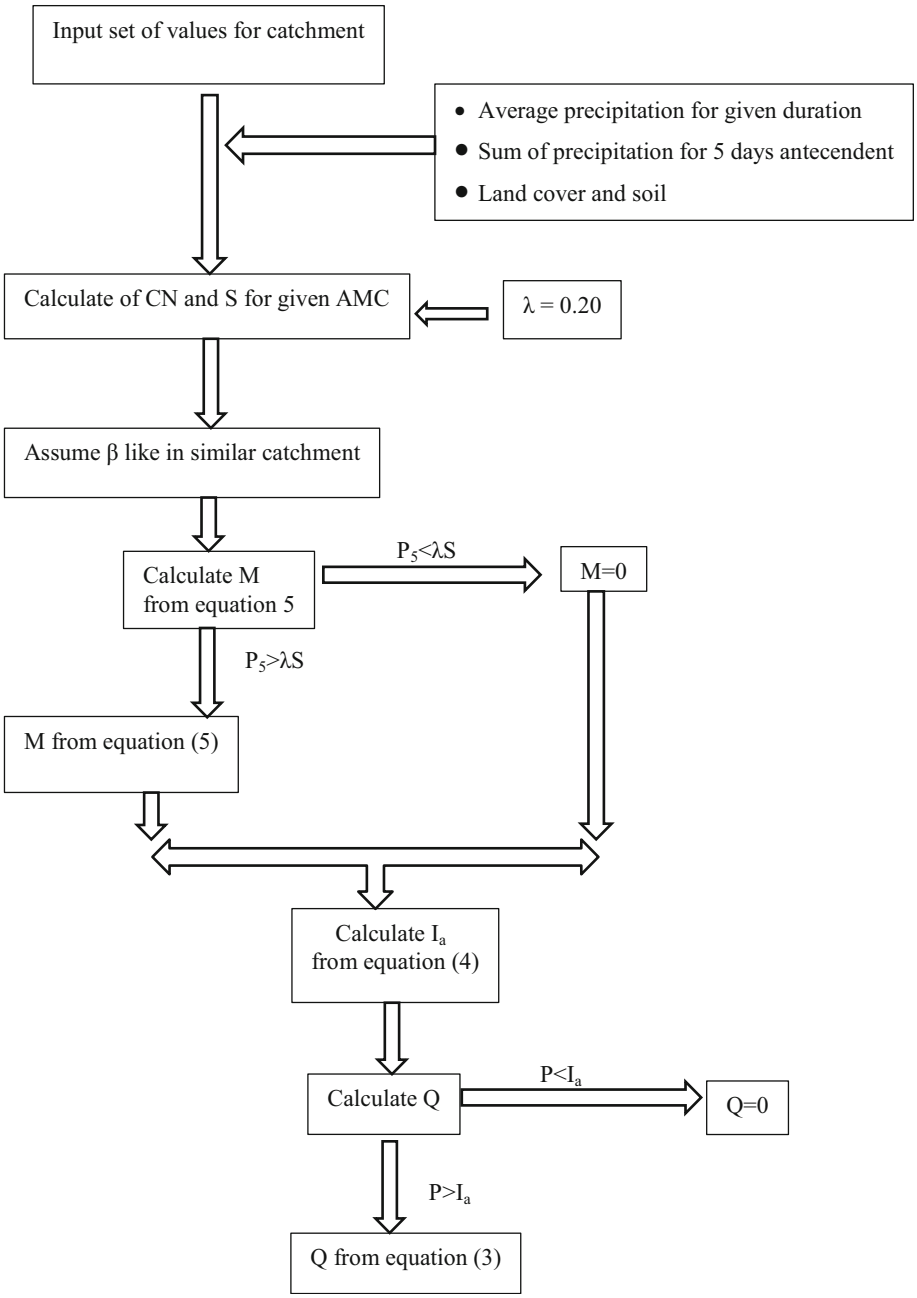


Fig. 3 The flowchart for calculation of direct outflow according to the modified SME method

was derived also for input parameters varying in the range from −90 to +90% around the parameters of the rainfall event that caused the greatest runoff. In Fig. 4, the plots of flexibility of *M* and *Q* to various input parameters were shown.

Both *M* and *Q* were insensitive to *S* (Fig. 4a) even though higher values of e(M) than e(*Q*) were observed. e(*M*) was positive, so an increase in *S* caused an increase in *M*. This was due to the fact that an increase in potential amount of water that can be accumulated in the soil profile (expressed by *S*) means that more of the net rainfall occurring prior to the flood can be accumulated in the soil profile creating the moisture M. M on the day of rainfall onset is assumed to be the amount of water infiltrated due to the antecedent 5-day rainfall (P5), prior to which the soil was completely dry. On the other hand, M indicates the amount of moisture added to a dry soil profile by rain P5 (Sahu et al. 2010). This amount of soil moisture depends on the substrate retention capacity determined by soil permeability and land cover.

Thus, the use of the maximum potential retention *S* as a soil moisture parameter is appropriate. The use of the *S* parameter from the USDA guidelines allows for the application of the modified SME method in uncontrolled

Table 5 Flexibility coefficients for SME method (model parameters adopted for the rainfall event that cause the greatest runoff)

Variable	Flexibility	Result
$S \rightarrow M$	0.84	Inflexible
$\boldsymbol{\beta \rightarrow M}$	**1.00**	**Flexible**
$P_5 \rightarrow M$	0.16	Inflexible
$\lambda \rightarrow M$	−0.006	Inflexible
$M \rightarrow Q$	0.02	Inflexible
$\boldsymbol{P \rightarrow Q}$	**1.08**	**Flexible**
$S \rightarrow Q$	−0.10	Inflexible
$P_5 \rightarrow Q$	0.003	Inflexible
$\lambda \rightarrow Q$	−0.02	Inflexible
$\beta \rightarrow$ Q	0.02	Inflexible

catchments. In the original SME method, in turn, the S parameter, that is functionally independent of the tabulated CN, has to be calibrated. This limits the use of the original SME method in uncontrolled catchments.

The flexibility e(Q) was negative; thus, the increase of S caused a decrease of Q. Low S values are typical for the substrate of low retention capacity (poorly permeable soil, highly sealed catchment). If these conditions are met, the potential for surface runoff formation is high. Increase of S results in retaining a substantial amount of precipitation in the soil profile due to infiltration and this limits the surface runoff. It is worth mentioning that the low sensitivity of the model to S changes is associated with low S values, i.e., in catchments with low retention capacity. Changes in the runoff amount Q and antecedent moisture

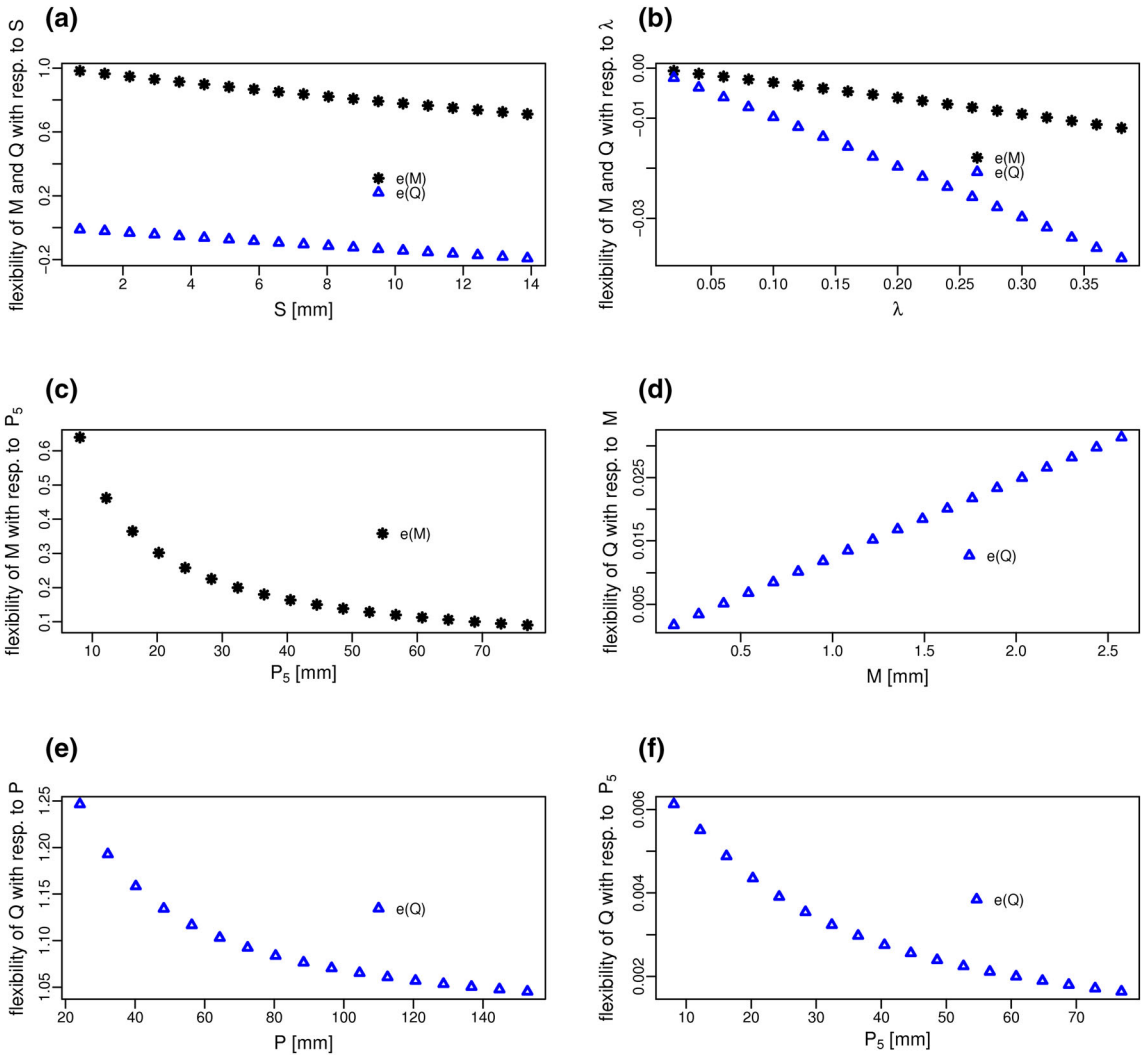


Fig. 4 Plots of flexibility of M and Q with respect to various input variables in the Stobnica catchment. In the panels the following plots are depicted: **a** flexibility of M and Q to S, **b** flexibility of M and Q to λ, **c** of M to P5, **d** of Q to M, **e** of Q to P, **f** of Q to P5

M yielded by the SME method were not sensitive to the changes in λ (Fig. 4b).

The flexibility e(m) was positive so an increase in λ resulted in rising M. Initial abstraction of precipitation is primarily due to infiltration and then to filling land depressions, interception, and evapotranspiration. Therefore, an increase in λ would increase the potential amount of precipitation that would improve the soil moisture M via infiltration. The e(Q) values were negative for runoff and this suggested that the runoff layer became thinner together with rising λ. The precipitation lost in the above-mentioned processes would not affect the formation of surface runoff, which is why the described relationships were observed. An increase in P_5 was accompanied by a rise in M that was reflected in positive values of e(M) (Fig. 4c). However, due to precipitation loss for evapotranspiration or interception, not the entire P_5 amount was converted into soil moisture. If the initial abstraction was huge (as determined by large S values) and P_5 precipitation was relatively low, the soil would not be supplied with water and its moisture and M would not increase. Thereby, M was insensitive to P_5. Similar pattern was observed for Q (Fig. 4f). Sensitivity analysis demonstrated that final result of the SME method computations, i.e., the change in direct runoff Q, was not strongly dependent on the changes in M (Fig. 4d). Increased soil moisture content would primarily reduce initial abstraction I_a (formula 4) and accelerate runoff formation. An increase in P_5 together with M growth may be explained by the fact that high soil moisture reduces initial precipitation loss, which means that most of the rainfall is converted into runoff. Sensitivity analysis indicated significant changes in M caused by variations in β values (Table 5). The β parameter in the formula (5) means that M due to P_5 at the beginning of a rain event was equal to a fraction β of the amount of moisture/water infiltrated during antecedent 5-day period P_5. The assumption is based on the fact that only a fraction, in general, of moisture/water added to the soil profile will contribute to M because of evapotranspiration, drainage, and other losses in antecedent 5-day period (Sahu et al. 2010). An increase in β caused an increase in Q but no flexibility of the model was demonstrated. Rising β was conducive to an increase in M and this consequently reduced the initial abstraction I_a. As mentioned before, the reduction of initial abstraction intensified the formation of direct runoff. In general, the SME method was very sensitive to changes in precipitation P causing a flood (Fig. 4e). This was because precipitation remains the main cause of the surface runoff formation.

Considering limited effects of the initial values, except for P, on the final computations of the runoff, the modified SME method seems appropriate for runoff estimation in uncontrolled catchments. To do that, λ must be assumed as 0.20 and S must be calculated from the CN according to USDA. The problem here is to calibrate β. It should be estimated from observed rainfall–runoff events that represent typical conditions of runoff formation in a specific region. Considering that SME method is not flexible toward changes in β and clear effects of catchment characteristics on its values may be observed (Table 3), the modified SME method seems to be suitable for the estimation of direct runoff in the catchments similar to those analyzed in this paper. The described relationships require of course further verification in a greater number of catchments with more diverse characteristics and different locations.

The flexibility of Q to β was also derived. It increased linearly with β and ranged from 0.02 to 0.03 for β varying from −90 to 90% of 0.22. Thus, the sensitivity of SME method to changes in β was low.

Advantages of the modified SME method

The following advantages of the modified SME method were shown in the three selected catchments:

1. It has a high quality performance that is higher than the quality of the original SME and of the SCS-CN models,
2. It is easier applicable than the original SME method,
3. It has a low sensitivity to input parameters, apart from rainfall.

The essential limitation of the modified SME method is the requirement to specify the parameters β and λ for a given catchment.

Conclusions

The study outcomes allowed us to draw the conclusions that

1. The modification of the SME method where the SCS-CN-based potential retention parameter S was included yielded direct runoff values which were very similar to the observed values.
2. Changes in runoff amount yielded by the modified SME method were highly sensitive to changes in rainfall. The model was insensitive to changes in any other parameters.
3. Errors in estimation of direct runoff were much smaller in the modified SME method than in the original SCS-CN method.
4. The empirical CNs were visibly higher than the tabulated CNs, derived from the SCS tables. This resulted in a considerable underestimation of the direct runoff when the SCS-CN method was used.

Acknowledgments The authors would like to thank the Dean of the Faculty of Environmental Engineering and Land Surveying, University of Agriculture in Krakow, for financial support. We also wish to thank the Institute of Meteorology and Water Management National Research Institute for daily flow data.

Compliance with ethical standards

Conflict of interest The authors do not declare any conflict of interest.

Appendix

$$\frac{\partial Q}{\partial P} = \frac{S^2\lambda^2 - \left(S^2 + 2PS\right)\lambda + (2P + M)S + P^2}{(P - \lambda S + S)^2}$$

$$\frac{\partial Q}{\partial S} = -\frac{S^2\lambda^3 - \left(S^2 + 2PS\right)\lambda^2 + P^2\lambda + P^2 + MP}{(P - \lambda S + S)^2}$$

$$\frac{\partial Q}{\partial \lambda} = -\frac{S^3\lambda^2 - \left(2S^3 + 2PS^2\right)\lambda + (2P + M)S^2 + P^2S}{(P - \lambda S + S)^2}$$

$$\frac{\partial Q}{\partial M} = \frac{P - \lambda S}{P - \lambda S + S}$$

$$\frac{\partial M}{\partial \beta} = \frac{P_5 S - \lambda S^2}{P_5 - \lambda S + S}, \frac{\partial M}{\partial P_5} = \frac{S^2\beta}{(P_5 - \lambda S + S)^2},$$

$$\frac{\partial M}{\partial S} = \frac{S^2\beta\lambda^2 - \left(S^2 + 2PS\right)\beta\lambda + P_5^2\beta}{(P_5 - \lambda S + S)^2}, \frac{\partial M}{\partial \lambda} = \frac{-S^3\beta}{(P_5 - \lambda S + S)^2}$$

References

ASCE (2009) Curve number hydrology: state of the practice. In: Hawkins RH, Ward TJ, Woodward DE, van Mullem JA (eds) American Society of Civil Engineers, Reston, USA

Banasik K, Woodward DE (2010) Empirical determination of runoff Curve Number for a small agriculture catchment in Poland. In: Proceedings of the 2nd joint federal interagency conference, vol 27. Las Vegas, NV, USA

Banasik K, Rutkowska A, Kohnová S (2014a) Retention and curve number variability in a small agricultural catchment: the probabilistic approach. Water 6:1118–1133

Banasik K, Krajewski A, Sikorska A, Hejduk L (2014b) Curve Number estimation for a small urban catchment from recorded rainfall-runoff events. Arch Environ Prot 40(3):75–86

Bedient PB, Huber WC, Vieux BE (2013) Hydrology and floodplain analysis. Pearson, London, p 813

Calver A, Stewart E, Goodsell G (2009) Comparative analysis of statistical and catchment modelling approaches to river flood frequency estimation. J Flood Risk Manage 2:24–31

Ciepielowski A. Study on relationships between parameters of the flood waves in selected cross sections. Scientific thesis and dissertations. SGGW, Warsaw, 1987

Darvishan AK, Banasik K, Sadeghi SH, Gholami L, Hejduk L (2015) Effects of rain intensity and initial soil moisture on hydrological responses in laboratory conditions. Int Agroph. 29:165–173

De Paola F, Ranucci A, Feo A (2013) Antecedent moisture condition (SCS) frequency assessment: a case study in southern Italy. Irrig Drain. 62:61–71

Deshmukh DS, Chaube UC, Hailu AE, Gudeta DA, Kassa MT (2013) Estimation and comparison of curve numbers based on dynamic land use land cover change, observed rainfall-runoff data and land slope. J Hydrol 492:89–101

Di Baldassarre G, Schumann G, Bates PD, Freer JE, Beven KJ (2010) Flood-plain mapping: a critical discussion of deterministic and probabilistic approaches. Hydrol Sci J 55(3):364–376

Dingman SL (2008) Physical hydrology, 2nd edn. Waveland Press, Long Grove, p 395

Domeneghetti A, Vorogushyn S, Castellarin A, Merz B, Brath A (2013) Probabilistic flood hazard mapping: effects of uncertain boundary conditions. Hydrol Earth Syst Sci 17:3127–3140

Ebrahimian M, Nuruddin AAB, Soom MA, Neng LJ (2012) Runoff estimation in steep slope catchment with standard nd slope-adjustment curve number method. Pol J Environ Stud. 21:1191–1202

Grimaldi S, Petroselli A, Romano N (2013a) Green-ampt curve number mixed procedure as an empirical tool for rainfall-runoff modelling in small and ungauged basins. Hydrol Process 27:1253–1264

Grimaldi S, Petroselli A, Romano N (2013b) Curve-number/green–ampt mixed procedure for streamflow predictions in ungauged basins: parameter sensitivity analysis. Hydrol Process 27:1265–1275

Hawkins RH, Ward TJ, Woodward DE, van Mullem JA (2009) Curve number hydrology: state of the practice. American Society of Civil Engineers, Reston

Hingray B, Picouet C, Musy A (2014) Hydrology: a science for engineers. CRC Press, Boca Raton

Hoes O, Nelen F (2005) Continuous simulation or event-based modelling to estimate flood probabilities? WIT Trans Ecol Environ 80:3–10

Hong Y, Adler RF (2007) Estimation of global SCS curve numbers using satellite remote sensing and geospatial data. Int J Remote Sens 29(2):471–477

Kowalik T, Wałęga A (2015) Estimation of CN parameter for small agricultural watersheds using asymptotic functions. Water 7(3):939–955

Krzanowski S, Miler AT, Walega A (2013), The effect of moisture conditions on estimation of the CN parameter value in the mountain catchment. Infrastructure and ecology of rural areas. 3/IV, 105–117 (In Polish)

Maidment DW, Hoogerwerf TN (2002) Parameter sensitivity in hydrologic modeling, technical report. The University of Texas at Austin, Austin

McCuen R (2003) Modeling hydrologic change: statistical methods. Lewis Publishers, Boca Raton

Mishra SK, Singh VP (2002) SCS-CN-based hydrologic simulation package. In: Singh VP, Frevert DK (eds) Mathematical models of small watershed hydrology and applications, water resources publs. LLC, Highlands Ranch, pp 391–464

Mishra SK, Singh VP (2003) Soil conservation service curve number (SCS-CN) methodology. Kluwer Academic Publ, Dordrec

Mishra SK, Jain MK, Bhunya PK, Singh VP (2005) Field applicability of the SCS-CN-based Mishra-Singh general model and its variants. Water Resour Manag. 19(1):37–62

Moriasi DN, Arnold JG, van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Am Soc Agric Biol Eng. 50:885–900

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models, part-I: a discussion of principles. J Hydrol 10:282–290

Petroselli A, Grimaldi S, Romano N (2013) Curve-number/green-ampt mixed procedure for net rainfall estimation: a case study of the mignone watershed, IT. Proced Environ Sci 19:113–121

Policht-Latawiec A, Kanownik W, Jurek A (2016) The effect of cooling water discharge from the power station on the quality of the Skawinka River water. Carpath J Earth Environ Sci 11(2):427–435

Ponce VM, Hawkins RH (1996) Runoff curve number: has it reached maturity? J Hydrol Eng 1(1):11–19

Rallison RE, Miller N (1981). Past, present, and future SCS runoff procedurę. In: Rainfall-runoff relationship. Proc of the international symphosium on rainfall-runoff modelling. Missisipi, Missisipi State University, p 353–364

Rutkowska A, Kohnová S, Banasik K, Szolgay J, Karabová B (2015) Probabilistic properties of a curve number: a case study for small polish and slovak carpathian basins. J Mt Sci 12(3):533–548

Sahu RK, Mishra SK, Eldho TI, Jain MK (2007) An advanced soil moisture accounting procedure for SCS curve number method. Hydrol Process 21(21):2872–2881

Sahu RK, Mishra SK, Eldho TI (2010) An improved AMC-coupled runoff curve number model. Hydrol Process 24(20):2834–2839

Sahu RK, Mishra SK, Eldho TI (2012) Performance evaluation of modified versions of SCS curve number method for two watersheds of Maharashtra, India. ISH J Hydraul Eng 18(1):27–36

Serinaldi F, Grimaldi S (2011) Synthetic design hydrographs based on distribution functions with finite support. J Hydrol Eng 16(5):434–446

Singh PK, Mishra SK, Berndtsson R, Jain MK, Pandey RP (2015) Development of a modified SMA based MSCS-CN model for runoff estimation. Water Resour Manage. doi:10.1007/s11269-015-1048-1

Soulis KX, Valiantzas JD (2012) SCS-CN parameter determination using rainfall-runoff data in heterogeneous watersheds – the two-CN system approach. Hydrol Earth Syst Sci 16:1001–1101

USDA Natural Resources Conservation Service (2004), Hydrology. In: National engineering handbook; USDA soil conservation service: Washington, Chapter 10

Wałęga A, Rutkowska A (2015) Usefulness of the modified NRCS-CN Method for the assessment of direct runoff in a mountain catchment. Acta Geophys 63(5):1423–1446

Wałęga A, Drożdżal E, Piórecki M, Radoń R (2012) Some problems of hydrology modeling of outflow from ungauged catchments with aspects of flood maps design. Acta Scientiarum Polonorum, Formatio Circumiectus 11(3):57–68 (**in Polish**)

Wałęga A, Rutkowska A, Policht-Latawiec A (2014) Sensitivity of Beta and Weibull synthetic unit hydrographs to input parameter changes. Pol J Environ Stud 23(1):221–229

Wałęga A, Michalec B, Cupak A, Grzebinoga M (2015) Comparison of SCS-CN Determination methodologies in a heterogeneous catchment. J Mt Sci 12(5):1084–1094

Acta Geophys. (2017) 65:377–394
DOI 10.1007/s11600-017-0036-7

RESEARCH ARTICLE

Changes in seasonal streamflow extremes experienced in rivers of Northwestern South America (Colombia)

J. O. Pierini[1] · J. C. Restrepo[2] · J. Aguirre[3] · A. M. Bustamante[3] · G. J. Velásquez[3]

Received: 21 March 2017 / Accepted: 28 March 2017 / Published online: 1 April 2017

Abstract A measure of the variability in seasonal extreme streamflow was estimated for the Colombian Caribbean coast, using monthly time series of freshwater discharge from ten watersheds. The aim was to detect modifications in the streamflow monthly distribution, seasonal trends, variance and extreme monthly values. A 20-year length time moving window, with 1-year successive shiftments, was applied to the monthly series to analyze the seasonal variability of streamflow. The seasonal-windowed data were statistically fitted through the Gamma distribution function. Scale and shape parameters were computed using the Maximum Likelihood Estimation (MLE) and the bootstrap method for 1000 resample. A trend analysis was performed for each windowed-serie, allowing to detect the window of maximum absolute values for trends. Significant temporal shifts in seasonal streamflow distribution and quantiles (QT), were obtained for different frequencies. Wet and dry extremes periods increased significantly in the last decades. Such increase did not occur simultaneously through the region. Some locations exhibited continuous increases only at minimum QT.

Keywords Discharge · Time series analysis · Streamflows · Extremes analysis · Trends

✉ J. C. Restrepo
restrepocj@uninorte.edu.co

J. O. Pierini
jpierini@criba.edu.ar

J. Aguirre
julian.aguirre@upb.edu.co

[1] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (IADO, CONICET, UNS), Florida 7500, Complejo CCTBB. Edificio E1, B8000FWB, Bahía Blanca, Argentina

[2] Grupo de Investigación en Geociencias, Departamento de Física, Universidad del Norte, km 5 vía Puerto Colombia, Barranquilla, Colombia

[3] Grupo de Investigaciones Ambientales GIA. Escuela de Ingenierías, Universidad Pontificia Bolivariana, Circular 1 No. 70-01 Bloque 11 Oficina 225, Medellín, Colombia

Introduction

Streamflow variability is a key element within hydrologic cycle analysis, especially when dealing with evaluating combined effects of global change and anthropogenic intervention (Vörösmarty and Sahagian 2000; Walling and Fang 2003; Hungtinton 2006; Milliman et al. 2008; Stosic et al. 2016). Several studies have analyzed streamflow changes for detecting significant trends, identifying major oscillations periods, and determining relationships between hydrological responses and climate forcings (e.g., Probst and Tardy 1987; Genta et al. 1998; Robertson and Mechoso 1998; Pekarova et al. 2003; Labat et al. 2004; Pasquini and Depetris 2007; Labat 2008, 2010; Milliman et al. 2008; Dai et al. 2009; Telesca et al. 2012, 2013a, b; Pierini et al. 2012, 2015; Shaban et al. 2014). Some of these studies have shown contrasting results regarding streamflow trends and variability. For example, Milliman et al. (2008) indicated that cumulative discharge of 137 rivers, representative for the entire regions of the world, remained statistically unchanged between 1951 and 2000, offering little support to a global intensification of the hydrological cycle. Pekarova et al. (2003) also found no evidence of significant trends (neither increasing nor decreasing) in the annual streamflow of the 24 major rivers of the world. However, both noted significant changes in individual rivers at regional levels and relatively accelerated

transitions between wet and dry periods (Pekarova et al. 2003; Milliman et al. 2008). On the contrary, dramatic changes in the hydrological patterns and long-term trends, as well as a crescent anthropogenic intervention, were reported for several major rivers of the world, especially during the last two decades (e.g., Probst and Tardy 1987; Labat et al. 2004; Pinter et al. 2006; Varis et al. 2012; Walling and Fang 2003; Stosic et al. 2016). A comprehensive analysis of trends in the runoff from major rivers worldwide, from 1920 to 1995, showed an increase in the world continental runoff during the twentieth century (Labat et al. 2004). According to Hungtinton (2006), these latter results, in conjunction with evidence of increasing runoff from smaller rivers in the northern hemisphere, provide possible evidence for a warming-associated intensification of the water cycle. There is a general agreement regarding those differences in regional responses to climatic oscillation and anthropogenic intervention lead to significant differences in streamflow long-term trends (Alcamo et al. 1997; Vörösmarty and Sahagian 2000; Pekarova et al. 2003; Milliman et al. 2008; Dai et al. 2009).

A central question in fluvial hydrology is to determine whether long-term trends might mask or hide short-term changes, especially in rivers where seasonal changes of streamflow are higher than interannual changes. In the rivers draining the Caribbean alluvial plain of Colombia (northern South America) (Fig. 1) the annual cycle, related to Inter Tropical Convergence Zone—ITCZ—shiftment, appears as the main oscillatory component of hydrologic variability, wherein the ENSO-related band (3–7 year) and the quasi-decadal band (associated with sea surface temperatures anomalies in the Tropical North Atlantic) represent the second-order oscillatory components of this variability (Restrepo et al. 2012, 2014). Despite most of these rivers do not exhibit any long-term trends in their streamflows (e.g., Mesa et al. 1997; Poveda 2004; Restrepo et al. 2012, 2014), some evidence suggests that hydrological changes occurred at intra-annual scale. A Mann–Kendall Test applied to monthly data highlighted significant upward trends during wet seasons in most of these rivers (Restrepo et al. 2014). These trends could lead to substantial differences in quantiles within a hydrologic year, which poses new challenges for effective water resource management across the Caribbean littoral of Colombia. For example, the Magdalena River experienced extremely low streamflow values at the beginning of 2010, during the dry season, whereas the wet season was one of the most intense and extensive of the hydrological record (1941–2010) (Hoyos et al. 2013). Analysis of historical discharges in these rivers showed a strong connection between long-term trends and hydrologic periodicities (Restrepo et al. 2014). A detailed study of intra-annual variability of streamflow, comprising the detection of

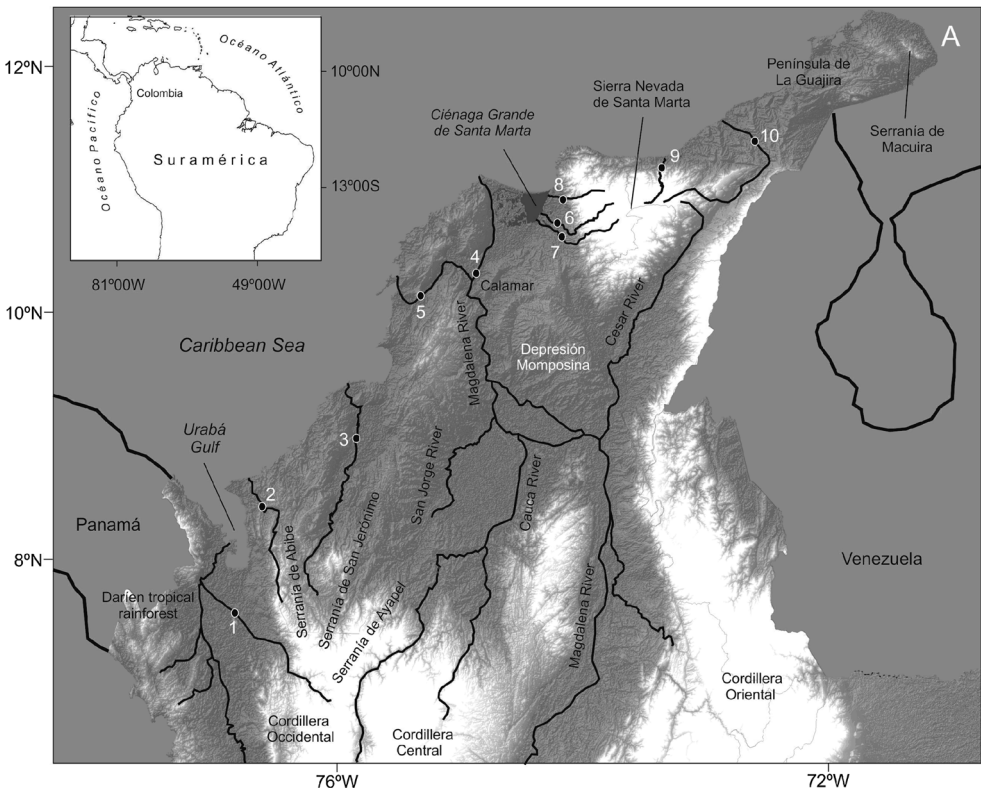


Fig. 1 Caribbean plain of Colombia in northwest South America. Major topographic features, selected rivers and gauge stations: *1* Sucío River, *2* Mulatos River, *3* Sinú River, *4* Magdalena River (Calamar), *5* Canal del Dique, *6* Aracataca River, *7* Fundación River, *8* Frío River, *9* Palomino River, and *10* Ranchería River

gradual or abrupt changes in season length and minimum and maximum streamflow values (e.g., Xu 2000; Chen et al. 2010), will improve our capabilities for monitoring and predicting the consequences of changing hydrologic regimes.

Many methods exist for detecting potential changes in the historical hydrological records (e.g., linear trend analysis, Mann–Kendall test, Wavelet analyses, scanning t test, scanning F test). Some of these tests were previously used in Colombian rivers, mainly at annual scale (Mesa et al. 1997; Poveda 2004; Restrepo et al. 2012, 2014). The streamflow extremes are typically defined based on the frequency of occurrence and/or their return periods. Thus, several fundamental probability frequency distributions can be used for estimating streamflow values; amongst which the most commonly used are the Generalized Extreme Value (GEV), Gumbel, Gamma and Weibull distributions (e.g., Wilks and Eggleston 1992; Vogel and Wilson 1996; Martins and Stedinger 2000; Katz et al. 2002; McMahon et al. 2007). In this study, we propose to establish a measure of the variability experienced in the extreme streamflows at seasonal scale, from the middle of the twentieth century to present time. Besides, we seek to detect whether there are modifications in their statistical characteristics, as well as changes in the variance of seasonal streamflow of the Caribbean Colombian rivers. We analyze in detail quantiles defined through a theoretical distribution function over (below) the 90th (10th) percentile, and variability of extreme streamflows amongst different stations. In fact, a change in a climate variable will also result in a change in the shape of its distribution. The main aim is focusing on the changes produced in extreme dry and wet values through time, to discern the impact of climate change in streamflow variability of the rivers draining the Caribbean littoral of Colombia. So far, no studies have dealt with changes in streamflow series as long as those hereby shown, available in Colombian, pursuing analysis on seasonal extremes over (below) 80th, 90th, 95th (20th, 10th, 5th) percentiles. In recent years, such knowledge has gained importance due to an increase in the number, duration and intensity of hydrological events such as floods and droughts (Hoyos et al. 2013). Thus, this study aims to establish the characteristics of temporal changes in the extreme streamflows of the conterminous Caribbean Colombian rivers, during the latter half of the twentieth century. An additional aim is to determine whether the changes in extreme streamflows could be characterized as a gradual change or as an abrupt change.

Study area

The Fig. 2 presents streamflow histograms for the rivers draining the Caribbean plain of Colombia, based on the entire record of monthly data (Table 1). Almost all locations are characterized by a bimodal distribution with peaks around May and October; it has been estimated that these seasonal changes are higher than the interannual changes of streamflow (Restrepo et al. 2014). According to Restrepo et al. (2014) the mean monthly streamflow of these rivers ranges between 4.63 and 6497 $m^3\ s^{-1}$, pouring collectively $\sim 330.5\ km^3\ year^{-1}$ of freshwater into the Caribbean Sea. The Magdalena River provides the largest supply of freshwater into the Caribbean Sea, with a mean discharge of 205.1 $km^3\ year^{-1}$.

Spectral analyses performed by Restrepo et al. (2012, 2014) revealed that the annual (related to Intertropical Convergence Zone—ITCZ—smigration) and quasi-decadal (associated with Sea Surface Temperatures anomalies) bands appear as the main oscillatory components of the hydrologic variability of the rivers draining the Caribbean alluvial plain of Colombia, whereas the ENSO-related band represents a second-order oscillatory component of this variability. Thus, the meridional oscillation of the ITCZ controls the annual hydrological cycle, defining two rainy seasons in the Caribbean plain. The first season extends from April to May, when the ITCZ is migrating from south to north; whilst the second and stronger season goes from September to November, when the ITCZ shifts southward. However, some local patterns can be distinguished as a consequence of orographic effects over air masses dynamics (Bernal et al. 2006; Poveda 2004; Poveda et al. 2006). Humid air masses from the Pacific Ocean are advected by the Chocó Jet. These masses rise rapidly along the Cordillera Occidental, promoting their deep convection, which in turn enhances the mesoscale convective systems causing intense rainfall rates (Poveda and Mesa 2004). The Caribbean or San Andrés jet current causes a humidity divergence in the northwest of South America that enhances the uplifting of air masses along the Sierra Nevada de Santa Marta slopes and causes strong superficial winds and dryness in the Peninsula de la Guajira (Bernal et al. 2006). The stronger rainy season coincides with maximum intensity of Chocó Jet, whilst the dry season corresponds to the maximum strength of the Caribbean Jet (Poveda et al. 2006). Hence, the western of the Caribbean plain, close to the Darien tropical rainforest, and the Sierra Nevada de Santa Marta exhibit the maximum rainfall rates and the lower mean annual temperatures, with values of $>2000\ mm\ year^{-1}$ and $<20\ °C$, respectively. On the contrary, the lowlands are hotter and drier with mean annual temperatures $>27\ °C$ and rainfall rates below to $1000\ mm\ year^{-1}$ (Mesa et al. 1997). At interannual scales, the major anomalies in hydrological patterns have been associated to both phases of El Niño/Southern Oscillation (ENSO) and to climatic and oceanographic processes of low-periodicity (i.e., quasi-decadal) such as Pacific Decadal Oscillations (PDO) and/or Tropical North Atlantic (TNA) (Restrepo et al. 2014). The ENSO warm phase (El

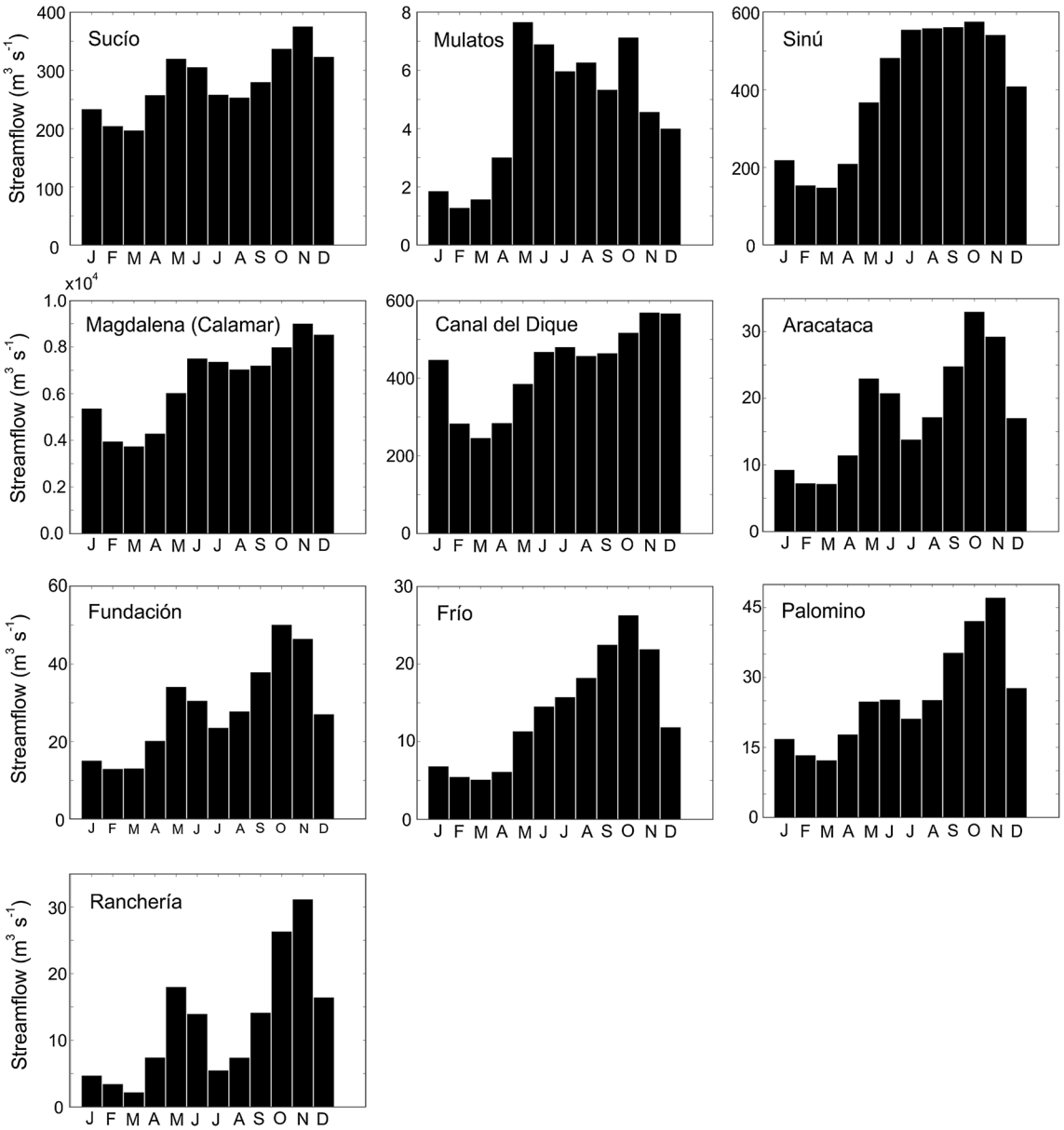


Fig. 2 Histograms of monthly streamflow for each gauge stations

Niño) promotes an increase in mean air temperature; a decrease in soil moisture and vegetation index, and thus the rainfall rates diminishes. The opposite anomalies during ENSO cold phase (La Niña) generate abundant and intense rainfalls (Poveda et al. 2001; Poveda 2004). Furthermore, the climate/oceanographic oscillations of low frequency act as a source of variability that enhance or diminish the ENSO effects on these rivers. For example, the concurrence of the strengthening of the quasi-decadal oscillation, between 1990 and 2010, and the major oscillations of the ENSO-related band, during 1998–2002 and 2009–2010, led to periods of intense hydrologic activity, wherein anomalously high streamflows occurred (Hoyos et al. 2013; Restrepo et al. 2014).

Data and methods

Data

The IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia) provided monthly streamflow data in ten drainage basins located along the Caribbean coast of Colombia. These drainage basins comprise different climatological and topographical settings (Fig. 1; Table 1). Time series selection was based both on its length and reliability. The verification of missing data was carried out before performing the statistical analysis. Time series employed in this study correspond to the locations listed in Table 1. The streamflow

Table 1 Rivers and gauging stations

River		Gauging station	Location				Monthly record
			Elevation (m.a.s.l.)	Drainage basin (10^3 km^2)	Longitude	Latitude	
1	Sucío	Mutata	132	4.52	76°26W	7°13N	1976–2010
2	Mulatos	Pueblo Bello	84	0.01	76°31W	8°12N	1977–2010
3	Sinú	Cotoca Abajo	5	14.73	75°51W	9°13N	1970–2010
4	Magdalena	Calamar	8	257.43	74°55W	10°15N	1941–2010
5	Canal del Dique	Santa Helena	3		75°24W	10°04N	1979–2010
6	Aracataca	Puente Ferrocarril	37	0.93	74°11W	10°35N	1965–2010
7	Fundación	Fundación	55	1.87	74°11W	10°31N	1958–2010
8	Frío	Rio Frío	30	0.32	74°09W	10°34N	1965–2009
9	Palomino	Puente Carretera	30	0.68	73°34W	11°14N	1973–2010
10	Ranchería	Hacienda Guamito	76	4.23	72°37W	11°10N	1976–2007

time series measured at the mouth of the watershed might be considered as a valuable integrated signal between gain and loss of the continental water cycle (i.e., precipitation, evapotranspiration, runoff). Thus, gauging stations analyzed in this study are situated close to the downstream part of the basin. Besides, they are located far enough apart to avoid spatial dependence problems. According to the descriptions performed by Restrepo et al. (2014) each of the selected sites can be associated to a hydrologic region.

Methods

Differentiation between natural variability and trends constitutes a major concern amongst hydrologist. Hydrological time series are usually analyzed by concentrating on differences in 30-year normal's along the whole period of record. This timeframe is also suitable for describing hydrologic time series as non-stationary signals with local periods of stationary patterns. Consequently, the data set length of our study, mostly 30 years (Table 1), suffices the minimum required length for examining evidence of climatic change in hydrologic time series. The local significance of long-term trends of streamflow was analyzed by Restrepo et al. (2014) through the non-parametric Mann–Kendall (MK) test. The MK test is a rank-based procedure, especially suitable for non-normally distributed data, which contains outliers and non-linear trends (Salas 1993). It has been demonstrated that the presence of serial correlation decreases the power of the MK test and leads to an erroneous rejection of the null hypothesis (Type II error) (Helsel and Hirsch 1992; Kulkarni and von Storch 1995; Yue et al. 2002; Yue and Wang 2002; Yue and Pilon 2003). One of the most common corrections applied in previous studies has been to remove the serial correlation in the data by pre-whitening; applying the MK test to the series data ($X'_i = X_i - r_{-1}X_{i-1}$) and then extract the lag-1 serial correlation coefficient (r_{-1}) of the detrended series (Yue et al. 2002; Yue et al. 2003; Yue and Pilon 2003; Restrepo et al. 2014). The pre-whitening was applied only to time series with lag-1 serial positive correlation to test the effect of the pre-whitening on the results. Therefore, we analyzed both, original data as well as pre-whitened data (X'_i), because serial correlation coefficients were generally low for the seasonal time series. The differences between the two approaches were not large.

The original dataset was divided on Dry1 (i.e., the sum of streamflows from December to March—first Dry climatic season), Wet1 (i.e., the sum of streamflows from April to May), Dry2 (i.e., the sum of streamflows from June to August), and Wet2 (i.e., the sum of streamflows sum from September to November). The analysis of streamflow focused on changes experienced in the extremes values of the dry and wet seasons along the historic record. Streamflow extremes values are typically defined based on the frequency of occurrence by percentile (e.g., upper 5, 1, or 0.1%) or by the return period. Several fundamental Probability Density Functions (PDF) can be adopted for estimating streamflow amounts. Amongst which the most commonly used are the Generalized Extreme Value (GEV), Gumbel, Gamma and Weibull distribution. The Kolmogorov–Smirnov (K–S) test is used to establish if a sample comes from a hypothesized continuous PDF. It is based on the largest vertical difference between a theoretical and empirical PDF. After testing several fit tests for different theoretical distributions (i.e., Gamma, Weibull, GEV, Gumbel, Lognormal; not shown for brevity), the Gamma distribution provides the best statistical fit to the seasonal patterns of streamflow in the Caribbean Colombian plain. Thus, according to the available monthly streamflow data, the most suitable procedure is based on fitting data to Gamma distributions. The gamma distribution is an arbitrary, but convenient choice, to represent variations in streamflow (Wilks and Eggleston 1992; Wilks 1995). Within the continuous theoretical distributions of probability this function is defined by the PDF,

$$f(x) = \frac{(x/\beta)^{\alpha-1}\exp(-x/\beta)}{\beta\,\Gamma(\alpha)},\ x,\ \alpha,\ \beta > 0 \quad (1)$$

The quantity $\Gamma(\alpha)$ is a value of the mathematical Gamma function. The two distribution parameters are α and β; shape and scale parameter, respectively. These parameters control the distribution of the Gamma function. The variance of the Gamma distribution is equal to $\alpha\beta^2$ (Wilks 1995).

Non-stationarity properties are present in time series of streamflow. Thus, a return period estimation for a given event should exhibit significant changes when different subseries are extracted from the whole hydrological record, demonstrating significant changes over the streamflow monthly events (magnitude and frequency). Furthermore, it is known that return periods are commonly used for fitting Gamma probability distributions (Yu 2003; Davison and Hinkley 1997; Hall et al. 2004). Thus, return periods (RP) of 5, 10 and 20 years were selected within this study to fit the Gamma distribution to each seasonal pattern of streamflow. The extreme dry or wet quantile (QT), such as the streamflow value that reaches or exceeds the probability of 0.20, 0.10, 0.05 or 0.80, 0.90, 0.95, respectively, was defined for each RP. The evaluation of the time variation of the oscillation processes employed a 20-year time-windows, with 1-year shifting through the series. Each time-window selects a subseries of seasonal streamflow data (i.e., 20 years for Dry1/Dry2/Wet1/Wet2, is thus used for each extracted series). For example, in the Magdalena the first data subseries of Dry1 season is (1941–1960), the second is (1942–1961), and so on until the last subseries (1991–2010). We employed a length of 20 years for the window after several evaluations on the concurrence of the major hydrological oscillation processes and the occurrence of the maximum power of the interannual and quasi-decadal bands, which defined period of intense hydrological activity (Restrepo et al. 2014, 2015). In addition, Restrepo et al. (2014) highlighted a change in the variability patterns of the fluvial systems of the Colombian Caribbean between 2000 and 2010, characterized by a shift toward the domain of quasi-decadal processes. Thus, it is considered that these sub-periods guarantee the probability of occurrence of significant oscillation processes during this time framework (i.e., 20 years).

When adjusting a theoretical distribution to a finite data sample, the estimation of the parameters presents uncertainties. Therefore, such procedure has an effect on the calculations of the quantiles and return periods. Thus, to minimize such effects, the maximum likelihood estimation (MLE) and a parametric bootstrap approach with 1000 iterations were used to estimate the parameters of the Gamma distribution. This method is based on random samples which are drawn from the fitted Gamma distribution. The subsequent analysis of quantiles employed the mean parameters from the bootstrapping to estimate the changes in the occurrence of wet and dry years. The confidence intervals of quantiles for different periods are compared with each other in order to decide whether a change in the mean QT estimate is statistically significant at a given error.

Results

Times series analysis

The general procedure employed for all the time series is presented in detail for the Magdalena River. Table 2 presents the Gamma distribution parameters (α and β) for the corresponding Wet and Dry streamflow seasons (Dry1, Wet1, Dry2, Wet2) in the same six windows, as mentioned in Fig. 3. Both parameters were estimated using the MLE and 1000 bootstrap methods. The 95% confidence interval estimated for such procedures is shown (Table 2). The skewness and kurtosis were added, as well as the variance estimated through bootstrapping. Wet and Dry streamflow seasons showed an increase in its variability since 1941 until recently. Figure 3 shows the time series and mean value of streamflow for the Dry1 (Fig. 3a), Wet1 (Fig. 3b), Dry2 (Fig. 3c) and Wet2 (Fig. 3d) seasons (upper panel) and the corresponding Gamma distributions (bottom panel), only for six different 20-year window periods (i.e., 1941–1960, 1951–1970, 1961–1980, 1971–1990, 1981–2000, 1991–2010). The rest of window periods are not shown for clarity of the figure. The streamflow of the Magdalena River exhibited a trend in the mean value, as well as in the extreme values during wet and dry years. The bottom panel illustrates changes in the location and shape of Gamma distributions. A shift of the Gamma distribution toward a lower mean (i.e., positive skewness) reflects a decrease in the streamflow when comparing different periods. On the contrary, the years of high streamflows observed after the second 20-year period and on, led to a negative skewness. Moreover, a positive trend in the streamflow was observed during the shift of the Gamma distribution toward higher means, from the first to the last of the six time-window periods shown in the Fig. 3. For example, the Wet1 and Wet2 seasons experienced an increase in its variability during the last decades, particularly between 1991 and 2010. Such change was reflected in a deformation (flattening and/or widening) of the Gamma distribution curve with respect to the other periods.

The Fig. 4 presents the linear trend values and its respective variance for each streamflow season and time-window evaluated in the Magdalena. The Dry1 (Fig. 4a) showed a period with positive significant trend values

Table 2 Mean value of shape (α) and scale (β), Skewness, Kurtosis and Variance parameters of Gamma distribution estimated for different 20-year period of Dry1/Wet1/Dry2/Wet2 streamflow time series for Magdalena using method of maximum likelihood, MLE, and 1000 bootstrap samples with 95% confident intervals

Period	(α) MLE	(β) MLE	(α) Bootstrap	(β) Bootstrap	Skewness	Kurtosis	Variance bootstrap ($\times 10^7$)
Dry1							
1941–1960	11.1	1758.2	12.7 ± 4.0	1671.8 ± 461.9	0.58 ± 0.08	0.52 ± 0.15	3.55 ± 0.09
1951–1970	20.8	972.8	24.4 ± 8.7	926.9 ± 305.9	0.42 ± 0.07	0.27 ± 0.09	2.10 ± 0.08
1961–1980	19.5	1093.1	24.0 ± 13.1	1060.0 ± 400.2	0.44 ± 0.08	0.30 ± 0.10	2.70 ± 0.21
1971–1990	17.6	1279.6	21.0 ± 8.9	1211.5 ± 393.6	0.46 ± 0.08	0.32 ± 0.10	3.08 ± 0.14
1981–2000	17.0	1302.1	20.7 ± 7.6	1203.4 ± 417.2	0.46 ± 0.08	0.33 ± 0.11	3.00 ± 0.13
1991–2010	15.2	1531.2	17.3 ± 5.2	1456.8 ± 411.2	0.49 ± 0.07	0.37 ± 0.10	3.67 ± 0.09
Wet1							
1941–1960	15.4	640.0	17.6 ± 4.9	604.0 ± 165.0	0.49 ± 0.06	0.36 ± 0.09	0.64 ± 0.01
1951–1970	31.9	288.3	36.0 ± 11.5	277.9 ± 76.0	0.34 ± 0.05	0.18 ± 0.05	0.28 ± 0.01
1961–1980	14.4	682.5	17.3 ± 6.5	647.1 ± 238.3	0.50 ± 0.09	0.39 ± 0.13	0.72 ± 0.04
1971–1990	14.4	716.6	16.9 ± 6.1	682.4 ± 227.3	0.51 ± 0.08	0.40 ± 0.13	0.79 ± 0.03
1981–2000	17.4	601.6	20.3 ± 7.0	573.7 ± 186.8	0.46 ± 0.07	0.33 ± 0.10	0.67 ± 0.02
1991–2010	15.3	730.9	17.1 ± 5.2	706.8 ± 192.8	0.50 ± 0.07	0.38 ± 0.10	0.85 ± 0.02
Dry2							
1941–1960	19.7	1046.2	22.0 ± 6.6	1008.1 ± 266.9	0.44 ± 0.06	0.29 ± 0.08	2.24 ± 0.05
1951–1970	25.3	854.7	28.1 ± 7.4	812.6 ± 178.0	0.39 ± 0.04	0.23 ± 0.05	1.86 ± 0.02
1961–1980	28.7	759.9	33.3 ± 11.2	714.9 ± 198.4	0.36 ± 0.05	0.20 ± 0.06	1.70 ± 0.04
1971–1990	24.5	891.4	29.7 ± 14.1	848.9 ± 297.5	0.39 ± 0.07	0.23 ± 0.08	2.14 ± 0.12
1981–2000	16.1	1379.4	18.9 ± 6.4	1289.0 ± 386.3	0.48 ± 0.07	0.35 ± 0.11	3.14 ± 0.10
1991–2010	12.6	1835.4	14.3 ± 4.4	1745.4 ± 448.3	0.54 ± 0.07	0.45 ± 0.12	4.36 ± 0.09
Wet2							
1941–1960	29.2	730.3	32.9 ± 10.2	699.6 ± 175.3	0.36 ± 0.05	0.20 ± 0.05	1.61 ± 0.03
1951–1970	35.0	644.4	40.8 ± 13.5	600.3 ± 162.3	0.32 ± 0.05	0.16 ± 0.04	1.47 ± 0.04
1961–1980	27.5	899.0	31.2 ± 9.3	859.5 ± 236.4	0.37 ± 0.05	0.21 ± 0.06	2.30 ± 0.05
1971–1990	24.0	1078.2	26.7 ± 6.3	1019.3 ± 235.5	0.39 ± 0.04	0.24 ± 0.05	2.77 ± 0.03
1981–2000	18.3	1390.6	21.5 ± 8.5	1346.0 ± 466.7	0.45 ± 0.08	0.32 ± 0.11	3.90 ± 0.19
1991–2010	11.3	2240.9	13.0 ± 4.5	2154.2 ± 686.6	0.58 ± 0.09	0.51 ± 0.15	6.03 ± 0.21

during the central years surveyed—the time-window between 1958 and 1977 (Table 3), corresponding to the extreme interval in Fig. 4a and an increasing trend since the 1986–2005 time-window, which exhibited a fairly consistent pattern. The maximum positive trend in streamflow (Fig. 4a) was more noticeable to this season during the 1957–1976 time-window, whilst a dramatic reversion in the sign of the trend was observed afterwards (Table 3). This remarkable feature was present also in other time series from Table 4.

Variances were analyzed to distinguish whether there is a climate change associated to the streamflow variability. For the Magdalena, the present Dry1 streamflow variances exhibited such high values as those at the beginning of the period surveyed [3.4×10^7 $(m^3\ s^{-1})^2$]; whilst in central years, the fluctuations were around 2.2×10^7 $(m^3\ s^{-1})^2$. The variance at the beginning and at the end of the surveyed period exhibited comparable high values (Fig. 4a). The streamflow pattern of the Wet1 season indicates that observed variances have been lower than the other seasons, exhibiting strong variability in the last decades (Fig. 4b). During the Wet2 season (Fig. 4d) the streamflow pattern indicates that actual variances have overcome the previous values observed through the entire record [5.8×10^7 $(m^3\ s^{-1})^2$]. Thus, there has been a shift in the streamflow distributions, as presented in Fig. 3, as well as irregular changes in the variance throughout time of the Wet1 and Dry2 streamflow seasons; whilst the streamflows of the Dry1 and Wet2 seasons experienced a greater variability. Along the whole record, the final segments of Wet1 and Dry2 seasons exhibited the highest variances [$\sim 4.1 \times 10^7$ $(m^3\ s^{-1})^2$], whereas the variance was close to 2.1×10^7 $(m^3\ s^{-1})^2$ in the central years (1970–1995) (Fig. 4). All seasons patterns indicate that their variability has increased significantly during the last decades; other rivers, such as Frío, Sinú, C. Dique, among others, exhibited similar

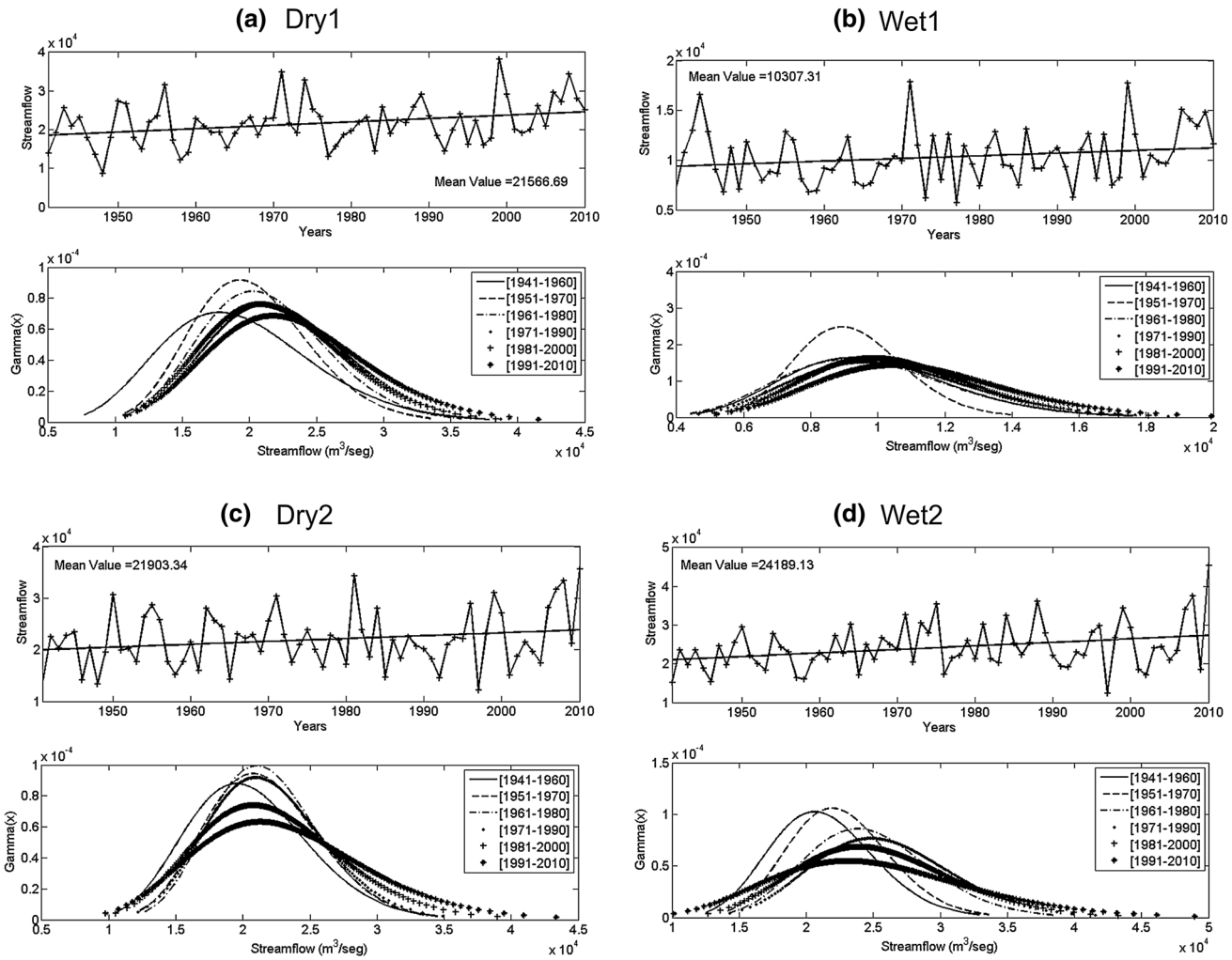


Fig. 3 Magdalena river seasonal streamflow values (*top*) and Gamma distribution for six different 20-year period (*bottom*) during 1941–2010 period in **a** Dry1, **b** Wet1, **c** Dry2, **d** Wet2 seasons

patterns (Table 4). From a retrospective view, the variance has increased four times from its original values (Fig. 4d).

Quantile analysis

Figure 5 represents the temporal evolution of QTs (in m^3 s^{-1}). The extremes from Dry1/Wet1/Dry2/Wet2 (upper/lower frame) are plotted for three return periods (RPs): 20, 10 and 5 years. Figure 5a, c correspond to Dry1 and Dry2, whereas Fig. 5b, d, represent the Wet1 and Wet2. If we consider the temporal variation for the entire 20-year time-window periods, Dry1 and Wet1 extremes (Fig. 5a, b) showed an increment throughout the century for the same RP (upper frame). This positive trend was enhanced during almost all the twentieth century in the Dry2 and Wet2 seasons (Fig. 5c, d). Besides, recent QTs for a return period of 20 years (percentile 0.8 in Fig. 5c, d), for instance, are similar to those observed at the beginning of the record. Additionally, some inter-decadal fluctuations can be observed. During the first part of the record, QTs were dominated by an oscillatory trend reaching a maximum peak around 1980–1990. Then, they exhibited an oscillatory pattern with decreasing trend. The same analysis applied to Dry1 and Wet1 seasons showed increasing streamflows in correspondence with the positive trends experienced from 1980.

Confidence Interval (CI) was computed for each one of the seasonal QTs obtained through the 20-year time-windows. The last time-window corresponds to the period 1991–2010. Its Confidence Interval was set as a Control Interval, in order to provide a statistical value to compare and measure the observed changes. Hence, we established the year in which the difference between the CI and the Control Interval exceed the zero value or is less than zero, depending if a negative or positive trend occurred. Thus, when there is no overlapping between both segments, the observed QT becomes significant for identifying changes in streamflow. Then, significant climate change periods were

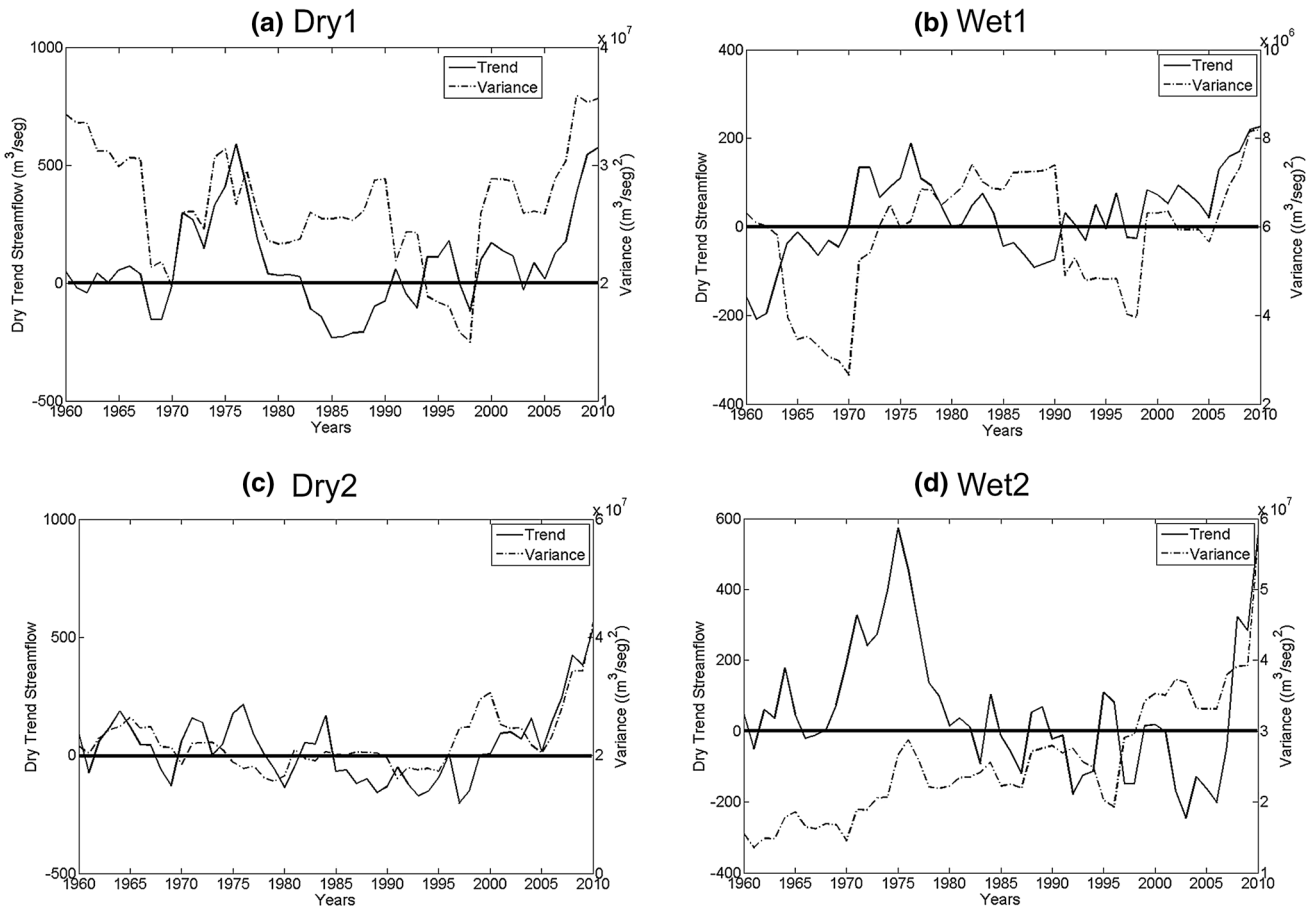


Fig. 4 Variance (*dashed line*) and linear trend (*full line*) for all 20-year window of Magdalena river seasonal streamflow in **a** Dry1, **b** Wet1, **c** Dry2, **d** Wet2 seasons (Note that the values are assigned to the end of the interval analyzed, e.g., 1941–1960 is assigned to 1960 and so on)

defined as those successive years in which there was no overlapping between the observed CI and Control Interval.

Figures 6 and 7 display the estimated QTs and CI for the Dry1/Dry2 (left) and Wet1/Wet2 (right) seasons in the Magdalena and Mulatos rivers, respectively. For these examples, the extremes at 10 and 90% probability levels were selected. For these examples, the extremes for the 10 and 90% probability levels were selected. Horizontal band corresponds to the Control Interval and the bars to the CI, for each QT. They were estimated using the 1000 bootstrap procedure. In general, a QTs increase were observed along the surveyed period with a negative difference between CI and Control Interval (extremes or mean CI minus extremes or mean Control Interval, respectively). Besides, a difference (positive or negative) is statistically significant (at a given error) when the 100% of the CI of the associated QT do not overlap with the Control Interval. It is then considered that the corresponding QT differs significantly from the Control Interval; and thus, the last of the 20-year time-windows is identified as the starting year of climate change.

Further, the QT for Dry1 (10% probability level), as well as the CI, in the Magdalena (Fig. 6a) showed a significant difference, with an oscillating pattern from the beginning of the record until the 1948–1967 time-window, when a major leap in the streamflow occurred. From this point until the mid-1980s the QT trend was positive and remained almost constant through the end of the record. When it reached approximately 16,000 $m^3\ s^{-1}$. Similarly, for Wet1 at the 10% probability level (Fig. 6b), it can be appreciated an oscillatory pattern from the beginning to the 1958–1977 time-window, when a significant difference emerged up. Subsequently, an increasing in the variability and a positive trend of QT were experienced until the end of the surveyed period. The values in the trend reached approximately 7900 $m^3\ s^{-1}$ at this final period. The QTs values from the Dry2 and Wet2 seasons are shown in Fig. 6c, d. Values estimated at the 10% probability level showed a remarkable wave form, with a period of approximately 24 years. According to the aforementioned definition, there are no years in which significant changes in QTs can be detected for Dry2 (Fig. 6c). Instead, for Wet2 (Fig. 6d) there were some significant changes for QT between the 1966–1985 and 1973–1992 time-windows,

Table 3 Seasonal Kendall trend significance test estimated for different 20-year period of Dry1/Wet1/Dry2/Wet2 streamflow time series for Magdalena and Mulatos

Dry1	Significant period with p: 0.05				
Magdalena	[1956–1975] ↑	[1957–1976] ↑	[1958–1977] ↑	[1990–2009] ↑	[1991–2010] ↑
Mulatos	[1983–2002] ↑	[1984–2003] ↑	[1985–2004] ↑		
Wet1	Significant period with p: 0.05				
Magdalena	[1957–1976] ↑	[1990–2009] ↑	[1991–2010] ↑		
Mulatos	None				
Dry2	Significant period with p: 0.05				
Magdalena	[1991–2010] ↑				
Mulatos	None				
Wet2	Significant period with p: 0.05				
Magdalena	[1956–1975] ↑	[1957–1976] ↑	[1958–1977] ↑		
Mulatos	None				

↑, positive trend; ↓, negative trend, −, no trend

which lead to an increase in the streamflow variability over the end of the surveyed period.

In the QTs analysis at the 90% probability level, the Magdalena showed the lowest extreme values for the Dry1 (Fig. 6e) in the 1946–1965 and 1979–1998 time-windows. The Wet1 experienced a similar pattern of significant changes during the same time-windows (Fig. 6f); although, a sudden increase in both streamflow and variability were experienced until the end of the surveyed period. The opposite occurred for the Dry2 values (Fig. 6g). There is evidence of significant changes in the QTs from the beginning to 1951–1970, 1955–1974 to 1961–1980, 1972–1991 to 1977–1996 and 1982–2001 to 1987–2006 time-windows; during the rest of period there was a continuous increase. Figure 6h shows the QT for Wet2 at the 90% probability level. It is observed a significant difference with an oscillatory pattern from the beginning of the record until the 1955–1974 time-window, where a little jump in the streamflow was generated. From this point until the mid-1980s the QTs trend is practically stable. Then, it remains oscillating until the end of the record, where it reached values of approximately 35,000 $m^3 s^{-1}$.

Figure 7a shows the QTs and the CI, at the 10% probability level, for Dry1 in the Mulatos river. The analysis highlighted a significant difference from the beginning of the record until the 1981–2000 time-window; afterwards, it exhibited oscillatory patterns, reaching its highest variability and values of approximately 3.1 $m^3 s^{-1}$. Similarly, for the Wet1 at the 10% probability level (Fig. 7b), it can be appreciated an oscillatory pattern with a positive trend since the beginning of the record. Such patterns generated a significant difference after the 1981–2000 time-window, reaching values of approximately 5 $m^3 s^{-1}$. The streamflow QTs values, estimated at the 10% probability level, for Dry2 and Wet2 are shown in Fig. 7c, d. According to the aforementioned definition, there were no years in which significant changes in QTs can be detected. Instead, for the Dry2 and Wet2 (Fig. 7c, d) there were some trend changes since the 1981–2000 time-window; particularly, the decrease in the streamflow values over the end of the surveyed period.

In the QTs analysis at the 90% probability level, Mulatos showed the extremes values for the Dry1 season (Fig. 7e) in the 1946–1965 and 1979–1998 time-windows, exhibiting also a similar variability in the intervening years. Such pattern was similar in the Wet1 (Fig. 6f) during the same time-windows. Afterwards, the streamflow and variability experienced a sudden increase at the end of the surveyed period. On the contrary, the Dry2 values did not exhibited significant changes in the QTs from the beginning of the record to the 1979–1998 time-window. The rest of the period underwent a continuous increase (Fig. 6g). Figure 6h shows the QT for Wet2 at the 90% probability level. There were no years in which significant changes in QTs can be detected. The oscillatory pattern was prominent. Such pattern remained stable until the end of the record, when values of approximately 29 $m^3 s^{-1}$ were reached.

Applying in detail the procedure described with the rest of time series corresponding to locations in Table 1, years with statistically significant changes in streamflow QTs were obtained, for both Dry and Wet seasons. As summary, the Dry1/Wet1/Dry2/Wet2 seasons with changes at the 10 and 90% probability level for each site

Table 4 Seasonal Kendall trend significance estimated for different 20-year period of Dry1/Wet1/Dry2/Wet2 streamflow time series

Dry1	Significant period with p: 0.05					
Frio	None					
Palomino	None					
Rancheria	None					
Sucio	None					
Sinú	[1985–2004] ↑	[1986–2005] ↑	[1990–2009] ↑	[1991–2010] ↑		
C. Dique	[1990–2009] ↑	[1991–2010] ↑				
Aracataca	[1972–1991] ↓	[1973–1992] ↑	[1974–1993] ↑	[1975–1994] ↑	[1976–1995] ↑	[1977–1996] ↑
Aracataca	[1978–1997] ↑	[1979–1998] ↑	[1980–1999] ↑	[1981–2000] ↑	[1982–2001] ↑	[1983–2002] ↑
Aracataca	[1984–2003] ↑	[1985–2004] ↑				
Fundación	[1966–1985] ↓	[1967–1986] ↓	[1968–1987] ↓	[1969–1988] ↓		
Wet1	Significant period with p: 0.05					
Frio	[1973–1992] ↑	[1974–1993] ↑	[1975–1994] ↑	[1976–1995] ↑	[1982–2001] ↑	[1983–2002] ↑
Frio	[1984–2003] ↑	[1985–2004]↑				
Palomino	[1973–1992] ↑	[1974–993] ↑				
Rancheria	[1988–2007] ↑					
Sucio	None					
Sinú	[1989–2008] ↑	[1990–2009] ↑	[1991–2010] ↑			
C. Dique	[1980–1999] ↑	[1981–2000] ↑	[1990–2009] ↑			
Aracataca	None					
Fundación	None					
Dry2	Significant period with p: 0.05					
Frio	[1971–1990] ↑	[1972–1991] ↑	[1973–1992] ↑	[1974–1993] ↑	[1975–1994] ↑	[1976–1995] ↑
Frio	[1977–1996] ↑	[1980–1999] ↑				
Palomino	[1983–2002] ↓					
Rancheria	None					
Sucio	[1991–2010] ↑					
Sinú	[1989–2008] ↑	[1990–2009] ↑	[1991–2010] ↑			
C. Dique	[1988–2007] ↑	[1989–2008] ↑	[1990–2009] ↑	[1991–2010] ↑		
Aracataca	None					
Fundación	[1971–1990] ↑	[1973–1992] ↑				
Wet2	Significant period with p: 0.05					
Frio	None					
Palomino	[1983–2002] ↓	[1984–2003] ↓	[1986–2005] ↓			
Rancheria	None					
Sucio	[1989–2008] ↑	[1991–2010] ↑				
Sinú	None					
C. Dique	[1991–2010] ↑					
Aracataca	None					
Fundación	[1971–1990] ↑	[1972–1991] ↑	[1984–2003] ↓	[1985–2004] ↓	[1986–2005] ↓	[1987–2006] ↓

↑, positive trend; ↓, negative trend, –, no trend

included in Table 1 were evaluated. However, all results are not shown due to brevity. Only the periods of consecutive 20-years time-windows with significant variability in seasonal streamflow of the Magdalena and Mulatos are shown (Figs. 6 and 7). Figures 8 and 9 show the statistical significance (black) for the QTs at the 10 and 90% probability level, respectively, for each station through the 20-year time-windows.

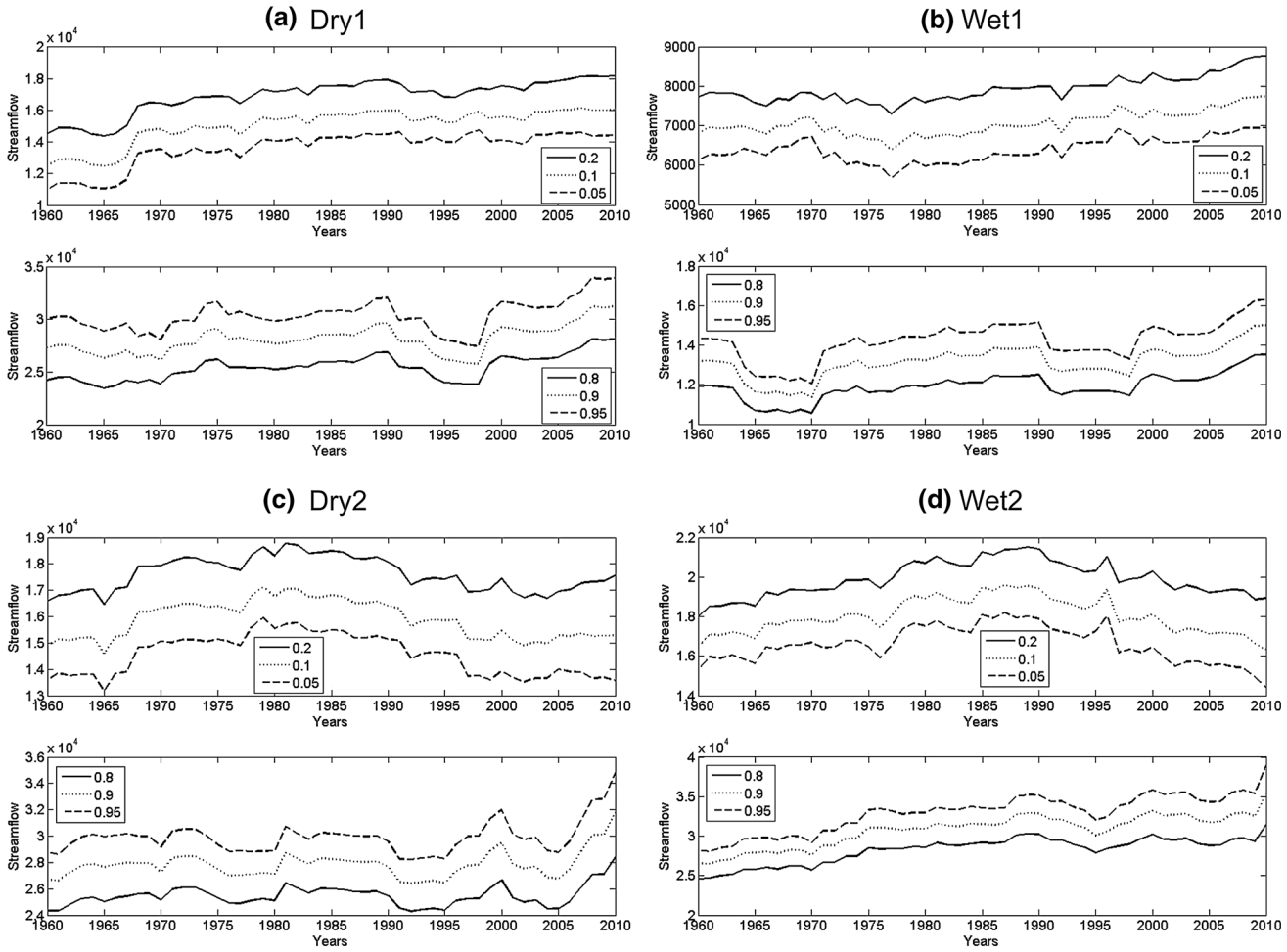


Fig. 5 Magdalena river Quantiles ($m^3\ s^{-1}$) for extreme Dry/Wet seasonal streamflow (*top/bottom panel*) with return periods of 20 (*full*), 10 (*point*) and 5 (*dashed*) years. **a** Dry1, **b** Wet1, **c** Dry2, **d** Wet2 (Note that the values are assigned to the end of the interval analyzed, e.g., 1941–1960 is assigned to 1960 and so on)

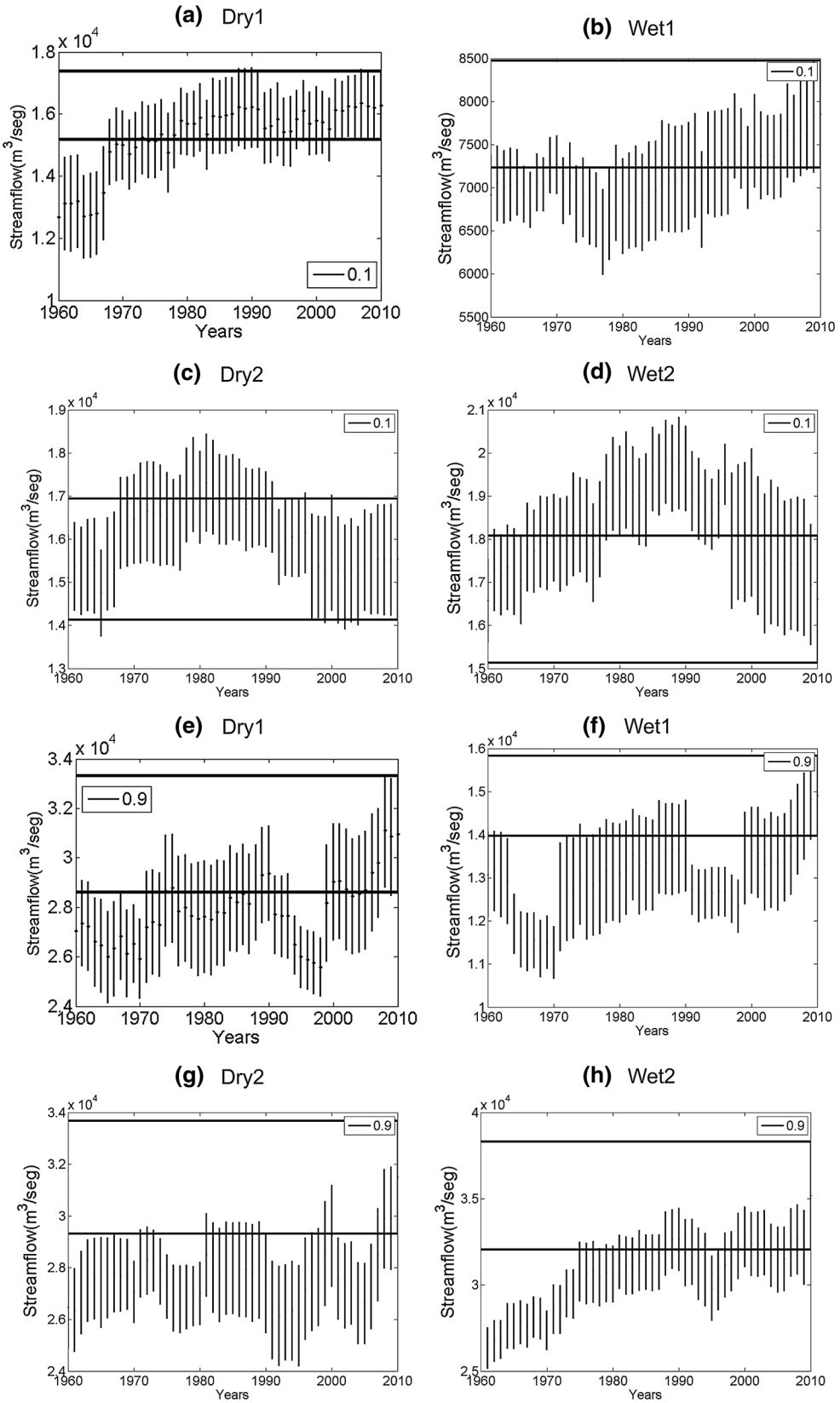
Discussion

Concerning the seasonal streamflow, the region under study presents well differentiated wet and dry seasons (Fig. 1). However, extreme dry QTs have increased through the record, except for the Mulatos at the 10% probability level (Fig. 7a). On the contrary, no significant QT variability have been found for the wet extreme in the locations, except for the Magdalena (Calamar) at the 90% probability level (Fig. 6e–h). Thus, the streamflow values of Dry1 and Dry2 seasons have increased through the entire record, indicating that dry extremes in the current distribution of streamflows are not as low as used to be before.

QT for extreme dry conditions increased over practically all the study region, except in the Mulatos River (i.e., Dry1 and Dry2), where no significant differences in QTs of the dry extreme streamflow were detected (Fig. 7). However, Wet1 and Wet2 extremes (at the 90% probability level) showed positive differences across the whole study area, indicating that current values are significantly higher than those experienced previously. For a fixed return period, the extreme Wet1 and Wet2 streamflows have been rising over the surveyed rivers, although with different time length and different change points. For example, the Caribbean Colombian region, represented by the Magdalena river, experienced a positive change several decades later.

Currently, the evidence indicates that the occurrence of hydrologic disasters in the Caribbean plain of Colombia is favored as a consequence of anthropogenic activities, the

Fig. 6 Magdalena river Quantiles and bootstrapping confidence intervals for each consecutive 20-year window of seasonal streamflow (Dry1/Wet1/Dry2/Wet2) at probability of 10% (**a–d**) and 90% (**e–h**). Horizontal full line band shows confidence interval (Control) of last 20-year period. See text for details (note that the values are assigned to the end of the interval analyzed, e.g., 1941–1960 is assigned to 1960 and so on) ▶


(a) Dry1
x 10^4
Streamflow(m^3/seg)
Years
0.1
(b) Wet1
(c) Dry2
(d) Wet2
(e) Dry1
0.9
(f) Wet1
(g) Dry2
(h) Wet2

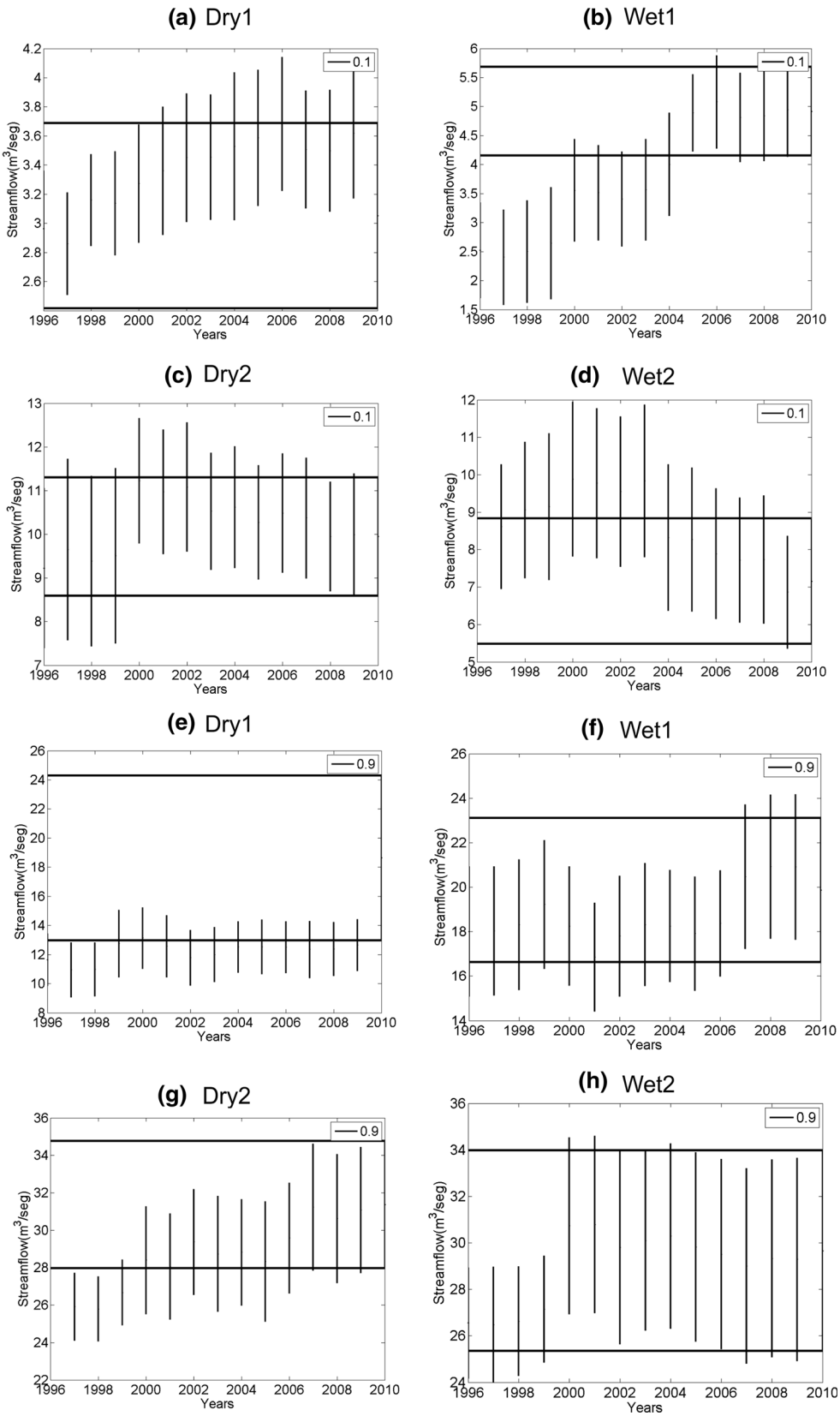


Fig. 7 Mulatos river Quantiles and bootstrapping confidence intervals for each consecutive 20-year window of seasonal streamflow (Dry1/Wet1/Dry2/Wet2) at probability of 10% (**a**–**d**) and 90% (**e**–**h**). Horizontal full line band shows confidence interval (Control) of last 20-year period. See text for details (note that the values are assigned to the end of the interval analyzed, e.g., 1977–1996 is assigned to 1996 and so on)

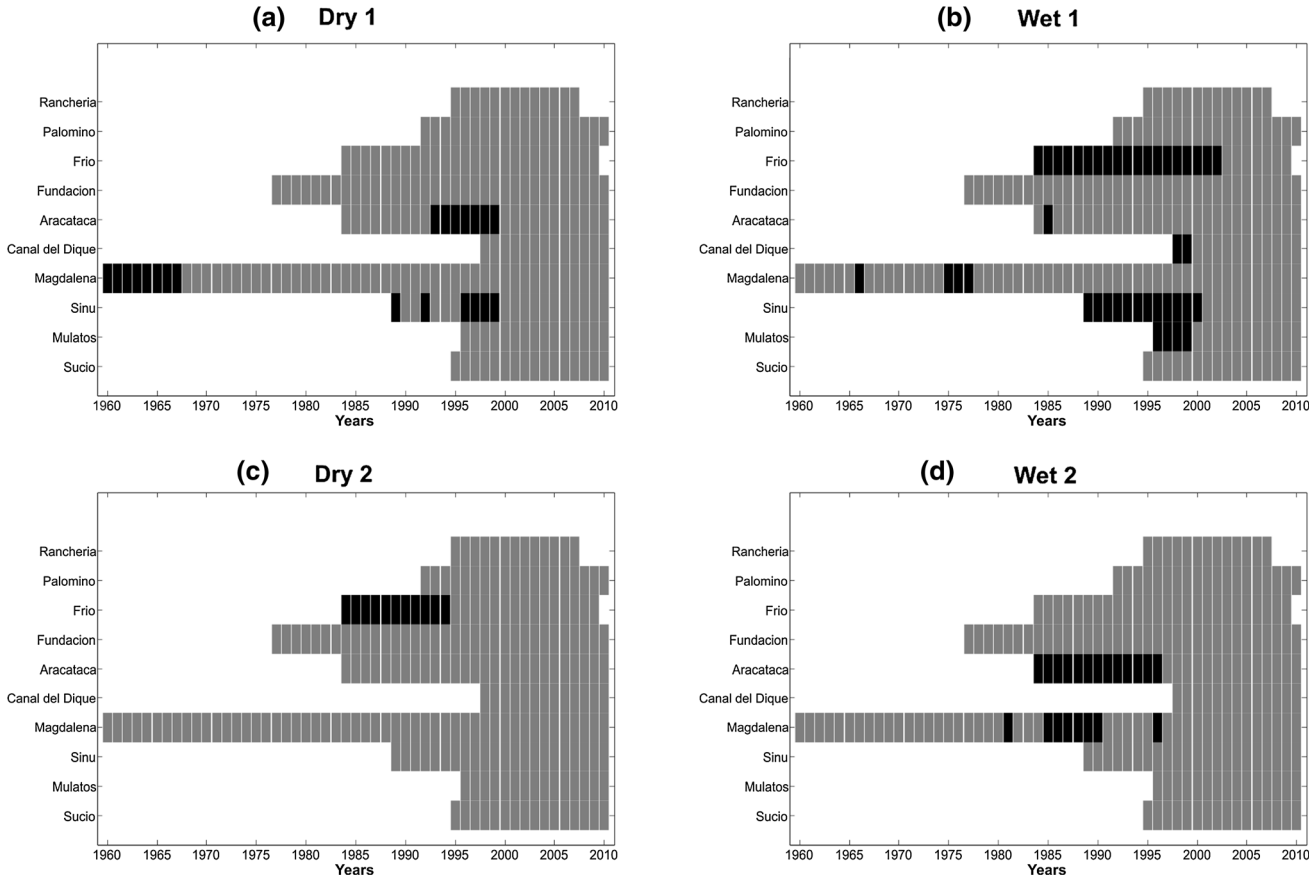


Fig. 8 Significant difference between Control and bootstrapping confidence intervals for each consecutive 20-year window of seasonal streamflow (Dry1/Wet1/Dry2/Wet2) at probability of 10% (**a**–**d**) (note that the values are assigned to the end of the interval analyzed, e.g., 1977–1996 is assigned to 1996 and so on)

increase of population and infrastructure development as well as others factors that modify land use conditions and the water cycle (Hoyos et al. 2013; Restrepo et al. 2015). In such context, it is useful to examine the observed shift in the theoretical distribution of streamflow with respect to both, time and space variability. In the evaluation of the Gamma parameters for each station, it becomes apparent that locations receiving a minimal amount of streamflow are described by either large shape or scale parameters. Values of shape and scale parameters in the Magdalena River (Table 2) exhibited substantial change during the 1951–1970 period, as well as a significant increasing trend since the 1955–1974 period. Distributions with a low shape parameter are positively skewed; as the shape value increases, the distribution curve becomes more symmetrical. These results indicate that significant changes in the frequency and intensity of extreme events might result from a relatively small shift in the average of streamflow distribution (e.g., Karl et al. 2008). Whenever a trend of extreme events is identified and a kind of persistence is determined, more evidence is added to the probable human effect on hydrological changes. However, a reverse in the trend sign and return to initial conditions would reflect the natural fluctuations of climate. Such is the case for the Magdalena and Mulatos rivers highlighted in the present study.

Although the QTs of the Wet1 and Wet2 showed a sustained increase during the last decades (Fig. 6), these values were similar to those observed at the beginning of the records. Both locations, Magdalena and Mulatos, experienced a shift in their mean streamflow values (see Figs. 6, 7) during the surveyed period. Opposite, the variance for Mulatos did not show a noticeable increase, just a slight fluctuation or even a decrease respect to its initial values; whilst at the Magdalena, a positive trend accompanies the larger variances. Thus, it is possible to distinguish certain period of years in which the increasing trends are followed by increasing variances, reflecting the increase in the intensity and variability of streamflows. This particular pattern is also present in the other locations along the surveyed area, such as in the Sucio, Sinú, Canal del Dique Magdalena and Mulatos rivers. Analyzing the properties of the distribution parameters whilst assessing long-term trends constitutes a primary approach aimed to

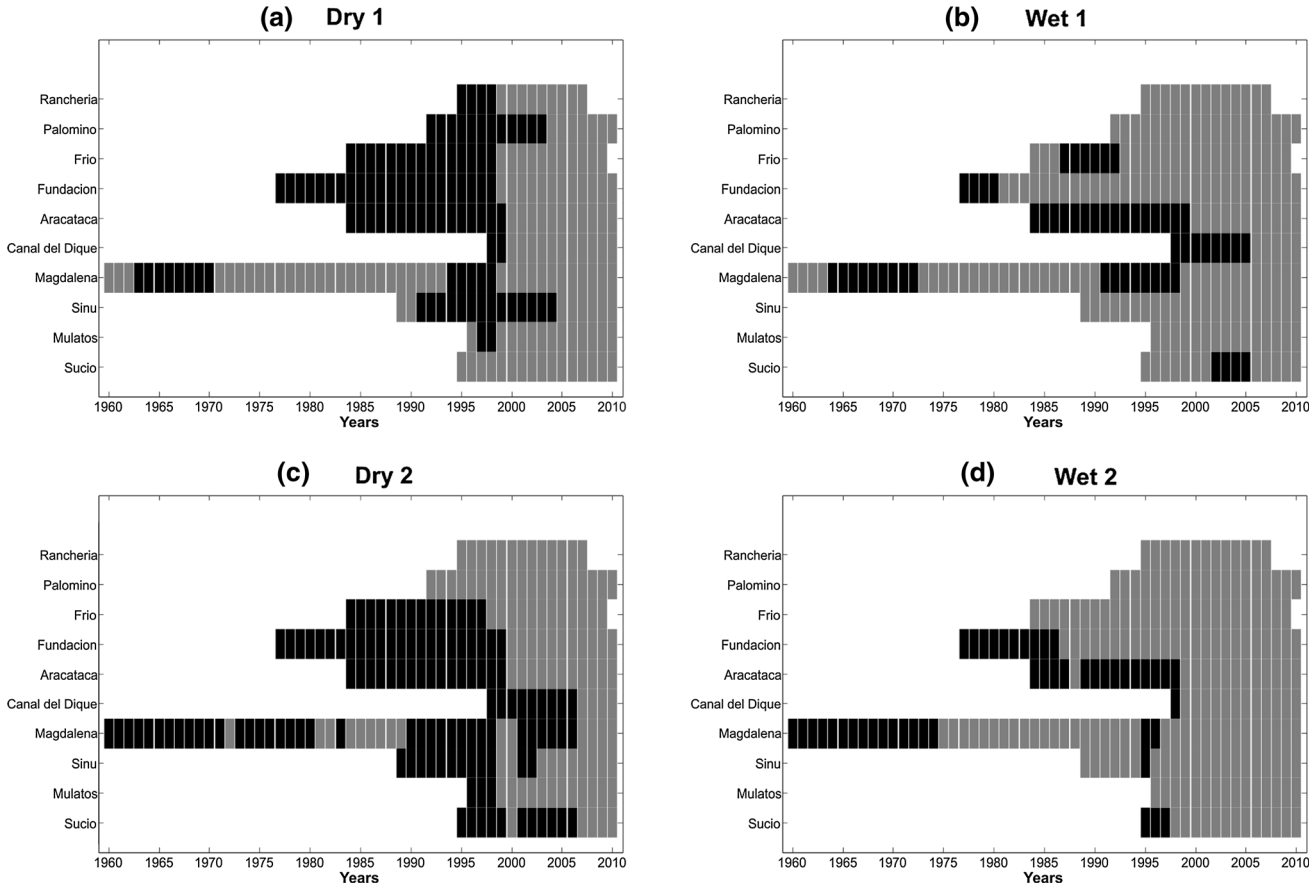


Fig. 9 Significant difference between Control and bootstrapping confidence intervals for each consecutive 20-year window of seasonal streamflow (Dry1/Wet1/Dry2/Wet2) at probability of 90% (**a**–**d**) (note that the values are assigned to the end of the interval analyzed, e.g., 1977–1996 is assigned to 1996 and so on)

classify climatic change, with respect to the location of the station.

The significant difference of QTs between the control and the last of the 20-year time-windows was identified as the starting year of climate change. Extreme dry QTs have changed significantly through the records, except in the Aracataca, Frío and Sinú rivers, between 1960 and1985 (see Fig. 8a, c at the 0.1 probability level). On the contrary, no significant QTs variability have been found for the wet extreme, except for the Frio, Aracataca, Magdalena, Sinú and Mulatos rivers at the 0.1 probability level (Fig. 8b, d). Similarly, extreme QTs at the 90% probability has changed within surveyed area, showing a significant change in the dry season since 1960–1980 in most of the Caribbean plain. Such changes have been triggered by the increase in climate variability generated by the strengthening of the ENSO during those decades (Poveda et al. 2001). However, wet extremes (at the 0.9 probability level) showed no significant differences within the surveyed area. The actual values are significantly lower than those observed previously. For a fixed return period, the extreme wet have changed as a whole for the study regions although with different time length and changing points.

Conclusion

Changes in quantiles of seasonal streamflow extremes were detected in ten different rivers through the Caribbean Colombian plain. Using a 20-year time-window to analyze series larger than fifty years of streamflow, it has been shown that extremes values from Dry1/Wet1/Dry2/Wet2 have varied (increased) significantly in the last decades. However, such changes have not occurred simultaneously in the whole region. The general diminishing in trend values observed at Mulatos, Sinú and Sucio (Tables 3, 4) provides an explanation of why this area exhibit a very particular behavior, which might be related to their highest susceptibility to the variability of sea surface temperature of the Caribbean The proximity of this watersheds to the warm pool forming off the coast of Urabá-Cordoba might induce anomalies in rainfall patterns due to the advection of humidity from the ocean and their relatively low

buffering capacity to filter out meteorological inputs (i.e., small/medium drainage basins).

In addition, the analysis highlighted significant differences between the Wet and Dry seasonal extremes. While increases in streamflow were experienced in both quantiles, some rivers showed continuous increases just at minimum QTs. Others rivers presented a pluri-annual variability of different length, whereas certain regions suffered a noticeable change in trend sign.

One of the mayor challenges encountered when quantifying extreme climatic events and their changes is the lack of information on long records, which besides are not always presented adequately for the analysis. Even when this information is available, it must go through processes of digitally processing, quality controls and homogeneity testing, and data interpolation. The lack of any of these factors affects the analysis. The monthly series employed in this work comes from the IDEAM; thus, comprising "long" reliable and official series subjected to well-known systematic control processes.

With the experience acquired performing this research, and considering that significant changes in quantiles at the seasonal scale are driven by changes in the amount of flow associated with daily rainfall or severe storms, further analysis of extremes based in daily and hourly data on streamflow and rainfall are needed in the future.

Acknowledgements This research was funded by the Universidad del Norte (Dirección de Investigación, Desarrollo e Innovación—DIDI Grant) and the Universidad Pontificia Bolivariana—Sede Medellín (Centro Integral para el Desarrollo de la Investigación—CIDI Grant).

References

Alcamo J, Doll P, Kaspar F, Siebert S (1997) Global change and global scenarios of water use and availability: an application of water GAP1.0. [online]. Kassel, University of Kassel. Available from: http://www.usf.uni-kassel.de/usf/archiv/dokumente/projekte/watergap.teil1.pdf [Accessed 27 July 2013]

Bernal G, Poveda G, Roldan P, Andrade C (2006) Patrones de variabilidad de las temperaturas superficiales del mar en la costa Caribe colombiana. Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales 30(115):195–208

Chen Y, Zhang Q, Xu Ch, Lu X, Zhang S (2010) Multiscale streamflow variations of the Pearl River basin and possible implications for the water resource management within the Pearl River Delta, China. Quat Int 226(1–2):44–53

Dai A, Qian T, Trenberth K, Milliman J (2009) Changes in continental freshwater discharge from 1948 to 2004. J Clim 22(10):2773–2792

Davison AC, Hinkley DV (1997) Bootstrap methods and their application. Cambridge University Press, Cambridge, p 582

Genta JL, Pérez Iribarren G, Mechoso CR (1998) A recent increasing trend in the streamflow of rivers in Southeastern South America. J Clim 11(11):2858–2862

Hall MJ, van den Boogaard HFP, Fernando RC, Mynett AE (2004) The construction of confidence intervals for frequency analysis using resampling techniques. Hydrol Earth Syst Sci 8(2):235–246

Helsel DR, Hirsch RM (1992) Statistical methods in water resources, chap 3. USGS, Technical Report, p 524

Hoyos N, Escobar J, Restrepo JC, Arango AM, Ortiz J (2013) Impact of the 2010–2011 La Niña Phenomenon in Colombia, South America: The Human Toll of an Extreme Weather Event. Appl Geogr 39:16–25

Hungtinton T (2006) Evidence for intensification of the global water cycle: review and synthesis. J Hydrol 319(1–4):83–95

Karl TR, Gerald AM, Christopher DM, Susan JH, Anne MW, William LM (2008) Weather and climate extremes in a changing climate regions of focus: North America, Hawaii, Caribbean, and U.S. Pacific Islands. Technical Report. U.S. Climate Change Science Program and the Subcommittee on Global Change Research. pp 176

Katz R, Parlange M, Naveau P (2002) Statistics of extreme in hydrology. Adv Water Resour 25(8–12):1287–1304

Kulkarni A, von Storch H (1995) Monte Carlo experiments on the efect of serial correlation on the Mann-Kendall test of trend. Meteorol Z (NF) 4:82–85

Labat D (2008) Wavelet analysis of the annual discharge records of the world's largest rivers. Adv Water Resour 31(1):109–117

Labat D (2010) Cross wavelet analyses of annual continental freshwater discharge and selected climate indices. J Hydrol 385(1–4):269–278

Labat D, Godderis Y, Probst J, Guyot JL (2004) Evidence for global runoff increase related to climate warming. Adv Water Resour 27(6):631–642

Martins E, Stedinger J (2000) Generalized maximum-likelihood extreme-value quantile estimators for hydrologic data. Water Resour Res 36(3):737–744

McMahon T, Vogel R, Peel M, Pegram G (2007) Global streamflows: Part 1: characteristics of annual streamflows. J Hydrol 347(3–4):243–259

Mesa O, Poveda G, Carvajal L (1997) Introducción al clima de Colombia. Universidad Nacional de Colombia, Bogotá

Milliman JD, Farnsworth K, Jones P, Xu K, Smith LC (2008) Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951–2000. Global Planet Change 62(3–4):187–194

Pasquini A, Depetris P (2007) Discharge trends and flow dynamics of South American rivers draining the southern Atlantic seaboard: an overview. J Hydrol 333(2–4):385–399

Pekarova P, Miklanek P, Pekar J (2003) Spatial and temporal runoff oscillation analysis of the main rivers of the world during the 19th–20th centuries. J Hydrol 274(1–4):62–79

Pierini JO, Gómez EA, Telesca L (2012) Prediction of water flows in Colorado River, Argentina. Latin Am J Aquat Res 40:872–880

Pierini JO, Restrepo JC, Lovallo M, Telesca L (2015) Discriminating between different streamflow regimes by using the Fisher-Shannon method: an application to the Colombia rivers. Acta Geophys 63:533–546

Pinter N, Ickes B, Wlosinski J, van der Ploeg R (2006) Trend in flood stages: contrasting results from the Mississippi and Rhine River systems. J Hydrol 331(3–4):554–566

Poveda G (2004) La hidroclimatología de Colombia: una síntesis desde la escala inter-decadal hasta la escala diurna. Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales. 28(107):201–222

Poveda G, Mesa O (2004) On the existence of Lloro (the rainiest locality on Earth): enhanced ocean-atmosphere-land interaction by a low-level jet. Geophy Res Lett 27(11):1675–1678

Poveda G, Jaramillo A, Gil M, Quiceno N, Mantilla R (2001) Seasonality in ENSO-related precipitation, river discharges, soil moisture, and vegetation index in Colombia. Water Resour Res 37(8):2169–2178

Poveda G, Waylen P, Pulwarty R (2006) Annual and inter-annual variability of the present climate in Northern South America and southern Mesoamerica. Palaeogeogr Palaeoclimatol Palaeoecol 234(1):3–27

Probst JL, Tardy Y (1987) Long range streamflow and world continental runoff fluctuations since the beginning of this century. J Hydrol 94(3–4):289–311

Restrepo JC, Ortíz JC, Maza M, Otero L, Alvarado M, Aguirre J (2012) Estimating Fluvial Discharge in the Caribbean Seaboard of Colombia: Magnitude, Variability and Extreme Events. In: Lane P, McKee Smith J (eds) Proceedings of the international conference on coastal engineering 2012 (ICCE 2012), 1–6 July 2012. Santander, Spain

Restrepo JC, Ortíz JC, Pierini J, Schrottke K, Maza M, Otero L, Aguirre J (2014) Freshwater discharge into the Caribbean Sea from the rivers of Northwestern South America (Colombia): Magnitude, variability and recent changes. J Hydrol 509:266–281

Restrepo JC, Schrottke K, Traini C, Ortíz JC, Orejarena A, Otero L, Higgins A, Marriaga L (2015) Sediment transport regime and geomorphological change in a high discharge tropical delta (Magdalena River, Colombia): insights from a period of intense change and human intervention (1990–2010). J Coastal Res 32(3):575–589

Robertson AW, Mechoso CR (1998) Interannual and decadal cycles in river flows of Southeastern South America. J Clim 11(10):2947–2957

Salas JD (1993) Analysis and modeling of hydrologic time series (Chapter 19). In: Maidment DR (ed), Handbook of hydrology, McGraw Hill, pp. 72

Shaban A, Telesca L, Darwich T, Amacha N (2014) Analysis of long-term fluctuations in stream flow time series: an application to Litani River, Lebanon. Acta Geophys 62:164–179

Stosic T, Telesca L, de Souza Ferreira DV, Stosic B (2016) Investigating anthropically induced effects in streamflow dynamics by using permutation entropy and statistical complexity analysis: a case study. J Hydrol 540:1136–1145

Telesca L, Lovallo M, Lopez-Moreno I, Vicente-Serrano S (2012) Investigation of scaling properties in monthly streamflow and Standardized Streamflow Index (SSI) time series in the Ebro basin (Spain). Phys A 391:1662–1678

Telesca L, Vicente-Serrano SM, Lopez-Moreno JI (2013a) Power spectral characteristics of drought indices in the Ebro river basin at different temporal scales. Stoch Environ Res Risk Assess 27:1155–1170

Telesca L, Lovallo M, Shaban A, Darwich T, Amacha N (2013b) Singular spectrum analysis and Fisher-Shannon analysis of spring flow time series: an application to Anjar Spring. Lebanon, Physica A 392:3789–3797

Varis O, Kummu M, Silmivara A (2012) Ten major rivers in monsoon Asia–Pacific: an assessment of vulnerability. Appl Geogr 32(2):441–454

Vogel R, Wilson I (1996) Probability distribution of annual maximum, mean, and minimum streamflows in the United States. J Hydrol Eng 1(2):69–76

Vörösmarty CJ, Sahagian D (2000) Anthropogenic disturbance of the terrestrial water cycle. Bioscience 50(9):753–765

Walling D, Fang D (2003) Recent trends in the suspended sediment loads of the world's rivers. Global Planet Change 39(1–2):111–126

Wilks DS (1995) Statistical methods in the atmospheric sciences. International geophysics series 59. Academic Press, San Diego, p 464

Wilks DS, Eggleston KL (1992) Estimating monthly and seasonal precipitation distributions using the 30- and 90-day outlooks. J Clim 5:252–259

Xu C (2000) Modeling the effects of climate change on water resources in central Sweden. Water Resour Manage 14(3):177–189

Yu CH (2003) Resampling methods: concepts, applications, and justification. Pract Assess Res Eval 8(19):1–23

Yue S, Pilon P (2003) The interaction between deterministic trend and autoregressive process. Water Resour Res 39(4):1077. doi:10.1029/2001WR001210

Yue S, Wang CY (2002) Applicability of prewhitening to eliminate the influence of serial correlation on the Mann-Kendall test. Water Resour Res 38(6):1068. doi:10.1029/2001WR000861

Yue S, Pilon P, Phinney B, Cavadias G (2002) The influence of autocorrelation on the ability to detect trend in hydrological series. Hydrol Process 16:1807–1829

Yue S, Pilon P, Phinney B (2003) Canadian streamflow trend detection: impacts of serial and cross-correlation. Hydrol Sci J 48(1):51–63

Acta Geophys. (2017) 65:395–410
DOI 10.1007/s11600-017-0030-0

RESEARCH ARTICLE

Comparative analysis of three ionospheric broadcast models for global navigation satellite systems

Shuhui Li[1] · Lihua Li[1] · Shuqing Wang[1] · Junhuan Peng[1] · Yujian Xu[1] · Yaohui Zhao[1]

Received: 24 December 2016 / Accepted: 13 March 2017 / Published online: 31 March 2017

Abstract An ionospheric model and corresponding coefficients broadcasted via GNSS navigation message are generally used to estimate the time delay for single-frequency GNSS users. In this article, the capabilities of three ionospheric models, namely, Klobuchar model, NeQuick Galileo version (NeQuick G), and Neustrelitz TEC broadcast model (NTCM-BC), were assessed. The models were examined in two aspects: total electron content (TEC) prediction and ionospheric delay correction effects in single-point positioning. Results show that both NeQuick G and NTCM-BC models outperformed Klobuchar model for predicting global TEC values during all the test days. Compared with Slant TEC (STEC) along the receiver-to-satellite ray path derived from IGS global ionosphere map (GIMs), STEC from NeQuick G and NTCM-BC models tend to have less bias than those from Klobuchar model in most situations. The point positioning results were improved by applying ionospheric broadcast models especially at the mid- and low-latitude stations.

✉ Shuhui Li
li.shuhui@163.com

Lihua Li
lihuali@cugb.edu.cn

Shuqing Wang
w_sq_2002@163.com

Junhuan Peng
pengjunhuan@163.com

Yujian Xu
xu15210965369@163.com

Yaohui Zhao
857434359@qq.com

[1] School of Land Science and Technology, China University of Geosciences (Beijing), 29 Xueyuan Road, Haidian District, Beijing 100083, China

Keywords Global navigation satellite system (GNSS) · Total electron content (TEC) · Klobuchar model · NeQuick Galileo version (NeQuick G) · Neustrelitz TEC broadcast model (NTCM-BC)

Introduction

Global navigation satellite system (GNSS) measurements are strongly influenced by ionospheric effects due to signal propagation through the ionosphere, which can severely degrade position accuracy. Therefore, a model that limits or removes ionospheric error is necessary. For GNSS positioning applications, the dual-frequency combination method can be used to effectively eliminate the ionospheric effect (i.e., ionospheric-free combination), but it cannot be used in a single frequency receiver. For a single frequency user, the ionospheric correction coefficients from broadcast ephemeris and an empirical model are common methods in positioning applications. To achieve accurate positioning results, a suitable ionospheric correction model is necessary (Hochegger et al. 2000; Feltens 2011).

Several broadcast models can be used to consider the ionospheric correction term. At present, the Global Positioning System (GPS) applies an ionospheric correction algorithm known as Klobuchar model (Klobuchar 1987). The GPS broadcasts eight coefficients necessary for this model in the navigation message. The Klobuchar model is supposed to provide a 40–50% correction of the ionospheric time delay (Klobuchar 1996). Meanwhile, the NeQuick Galileo version (NeQuick G) is the official ionosphere correction model incorporated in the Galileo receivers (Galileo Information Center 2006), and the model coefficients are broadcasted as part of the Galileo navigation message. The NeQuick model was developed from

NeQuick 1 to NeQuick 2, and NeQuick G is the single-frequency ionospheric correction algorithm for navigation message. Using coefficients a_0, a_1, and a_2, the daily effective ionization levels (AZ) representing solar activity levels can be calculated and used in the NeQuick G model to achieve ionospheric slant total electron content (STEC) along the satellite-to-receiver ray path (European Commission 2015). Hoque and Jakowski (2015) and Hoque et al. (2015) proposed an alternative ionospheric correction algorithm called Neustrelitz TEC broadcast model (NTCM-BC), which is also applicable to global satellite navigation systems. The NTCM global and regional TEC model has been developed during the past two decades at the Institute of Communications and Navigation, German Aerospace Center (DLR) in Neustrelitz. The NTCM global empirical model was simplified for use as a broadcast ionosphere model for satellite navigation systems. To drive NTCM-BC, nine ionospheric correction coefficients need to be uploaded to the satellites for broadcasting in the navigation message.

Studies have been conducted on the performances of different ionospheric models in providing ionospheric delay correction (Yuan and Ou 2001; Orús et al. 2002; Yuan et al. 2008; Radicella et al. 2008; Farah 2009; Macalalad et al. 2013; Wang et al. 2016). Angrisano et al. (2013a) assessed their performance in single-point positioning of the NeQuick G and compared the performance with that of the Klobuchar model. The analysis was performed in measurements domain and the results showed that the NeQuick G model provided better ionospheric delay estimation in 98% of the tests. Bidaine et al. (2013) compared the improvement of position accuracy of the Klobuchar model with different NeQuick versions and found that vertical errors dropped by 56–64% as a result of the analyzed ionospheric corrections; however, horizontal errors decreased by 27% at most. Angrisano et al. (2013b) compared the single-point positioning performance of the NeQuick G model with that of the Klobuchar model used for GPS. They found that in general, the NeQuick G model exhibited better horizontal positioning performance than the Klobuchar model. Yu et al. (2015) assessed the performance of the NeQuick G model in predicting global TEC by ingesting the global ionosphere map (GIMs) data from previous day(s). The results showed the good performance of the GIMs-driven-NeQuick model. An average of 86% vertical TEC error less than 10 TECU was achieved, and the model performed better during years with low solar activity. Today the NeQuick G corrections are available in navigation message sent by Galileo satellites.

Despite the series of studies on the Klobuchar and NeQuick G models, slight attention has been given to NTCM-BC model. A study (Hoque and Jakowski 2015) showed that the TEC prediction performance of NTCM-BC was better than that of Klobuchar model and was comparable to that of NeQuick G model. In terms of complexity, the NTCM-BC was easier to handle than the NeQuick G, and in this respect comparable to Klobuchar model. Furthermore, although the performance specifications of NTCM-BC algorithms were expressed in terms of TEC in the study of Hoque and Jakowski (2015) and its impact on point positioning has been done to some extend in Hoque et al. (2015), its performance still remained the subject of investigation.

The overall aim of this paper is to assess the performance of the three ionosphere models, namely, Klobuchar, NeQuick G, and NTCM-BC models. First, we demonstrate the effectiveness of the models in terms of TEC, which is the primary parameter they are expected to provide. Subsequently, we focus on the positioning error reduced by the corrections from the models. The analysis is performed under different geomagnetic activity conditions. The output of this article can improve our understanding of these ionospheric models.

Ionospheric models

Klobuchar model

The Klobuchar model provides a daily vertical TEC profile comprising a cosine representation during the day and a constant value during the night (5×10^{-9}s). Both amplitude and period of the cosine term are represented by four broadcast coefficients ($\alpha_0 \sim \alpha_3$, $\beta_0 \sim \beta_3$) that define a third order polynomial of the geomagnetic latitude. The phase of the maximum is fixed at 14:00 local time. A mapping function based on a thin-shell approximation is applied to convert the vertical time delay at the pierce point to a slant delay along a given ray path (Klobuchar 1987). The slant time delay is written as:

$$\Delta\tau = \left[D + A\cos\left(\frac{2\pi(t - T_{\mathrm{P}})}{P}\right)\right]\mathrm{MF}, \quad (1)$$

where $A = \sum_{i=0}^{3} \alpha_i \varphi_{\mathrm{m}}^i$, $P = \sum_{i=0}^{3} \beta_i \varphi_m^i$, φ_{m} is the geomagnetic latitude of ionospheric Pierce point (IPP), t is the local time of the IPP, and $T_{\mathrm{p}} = 14$. MF represents the mapping function, which can be written as:

$$\mathrm{MF} = 1 + 16 \cdot (0.53 - \mathrm{El})^3, \quad (2)$$

where El is the elevation angle in semicircles.

NeQuick G model

NeQuick is an ionospheric model developed at the Aeronomy and Radiopropagation Laboratory of the Abdus Salam International Centre for Theoretical Physics in Trieste, Italy, and at the Institute for Geophysics, Astrophysics

and Meteorology of the University of Graz, Austria. NeQuick is a three-dimensional model that describes spatial and temporal variations of the ionospheric electron density. Slant TEC is computed by integrating TEC along the signal path (Radicella and Leitinger 2001; Nava et al. 2005; Leitinger et al. 2005; Nava et al. 2008). An original version of NeQuick 1 and an updated version of NeQuick 2 currently exist. The NeQuick Galileo version (NeQuick G) was developed by the European Space Agency and considered for the final Galileo receiver algorithm.

The NeQuick algorithm was originally developed to be used with monthly averaged solar flux F10.7 index. When used in real time GNSS ionospheric correction application, the monthly averaged F10.7 index is replaced by a daily input parameter, that is, Az. Az is defined as a second-degree polynomial of the receiver-modified dip latitude (MODIP) μ and is expressed in solar flux unit (sfu, 10^{-22} [W m^{-2} Hz^{-1}]).

$$\mathrm{Az}(\mu) = a_0 + a_1\mu + a_{21}\mu^2, \tag{3}$$

where the three coefficients a_0, a_1, and a_2 will be broadcasted through the Galileo navigation message. MODIP μ can reflect the comprehensive influences of magnetic field and the geographic latitude. MODIP is obtained using Eq. 4:

$$\tan\mu = \frac{I}{\sqrt{\cos\phi}}, \tag{4}$$

where I denotes geomagnetic dip and ϕ represents geographic latitude. I can be computed using IGRF (or DGRF) models.

NTCM-BC model

In 1995, a family of NTCM global or regional models has been proposed at DLR. These models are used for receiver and satellite bias estimation, TEC calibration, and background model in grid vertical delay computation (Jakowski 1996; Jakowski et al. 2011). In an empirical global NTCM model, 12 model coefficients are used to describe the TEC variation at all levels of solar activity. Recently, Hoque and Jakowski (2015) proposed the NTCM-BC as a broadcast ionosphere model for satellite navigation systems. The NTCM-BC model is given by:

$$\mathrm{TEC}^{\mathrm{vert}}_{\mathrm{NTCM-BC}} = F_1 \cdot F_2 \cdot F_3, \tag{5}$$

where F_1 describes local time (LT) dependency; F_2 represents the geomagnetic field dependency of TEC; and F_3 describes the two ionization crests at the low-latitude anomaly regions on both sides of the geomagnetic Equator. The LT dependency includes diurnal (V_{D}), semidiurnal (V_{SD}), and ter-diurnal (V_{TD}) harmonic components explicitly. cos χ*** and cos χ** in F_1 are used to reflect the solar elevation angle dependency of TEC.

$$F_1 = \cos\chi^{***} + \cos\chi^{**}(c_1\cos V_{\mathrm{D}} + c_2\cos V_{\mathrm{SD}} + c_3\sin V_{\mathrm{SD}} + c_4\cos V_{\mathrm{TD}} + c_5\sin V_{\mathrm{TD}}) \tag{6}$$

$$V_{\mathrm{D}} = 2\pi(\mathrm{LT} - 14)/24, \tag{7}$$

$$V_{\mathrm{SD}} = 2\pi\mathrm{LT}/12, \tag{8}$$

$$V_{\mathrm{SD}} = 2\pi\mathrm{LT}/8, \tag{9}$$

$$\cos\chi^{***} = \cos(\varphi - \delta) + 0.4 \tag{10}$$

$$\cos\chi^{**} = \cos(\varphi - \delta) - \frac{2}{\pi}\cdot\varphi\cdot\sin\delta, \tag{11}$$

where φ is the geographic latitude and δ is the declination of the Sun (all angles in radians).

$$F_2 = 1 + c_6\cos\varphi_{\mathrm{m}}, \tag{12}$$

where φ_{m} is geomagnetic latitude.

$$F_3 = c_7 + c_8\exp\left(-\frac{(\varphi_{\mathrm{m}} - \varphi_{\mathrm{c1}})^2}{2\sigma_{\mathrm{c1}}^2}\right) + c_9\exp\left(-\frac{(\varphi_{\mathrm{m}} - \varphi_{\mathrm{c2}})^2}{2\sigma_{\mathrm{c2}}^2}\right), \tag{13}$$

where $\varphi_{\mathrm{c1}} = 16°\mathrm{N}$ and $\varphi_{\mathrm{c2}} = 10°\mathrm{S}$, and they represent the positions of the northward and southward crests, respectively. The corresponding Gaussian half-widths are given by $\sigma_{\mathrm{c1}} = 12°$ and $\sigma_{\mathrm{c2}} = 13°$.

STEC along the satellite-to-receiver ray path can be calculated using an elevation-dependent mapping function $M(\varepsilon)$:

$$M(\varepsilon) = \left[1 - \left(\frac{R_{\mathrm{e}}\cos\varepsilon}{R_{\mathrm{e}} + h_{\mathrm{I}}}\right)^2\right]^{-1/2}, \tag{14}$$

where R_{e} is the Earth's mean radius in kilometers and ε is the elevation angle in radians. Based on a spherically thin-shell ionosphere assumption, the fixed shell height h_{I} is 400 km.

Data and method

In this article, GPS data from six IGS stations with different latitudes are used to analyze the TEC prediction and ionospheric delay correction performance of the models. All these IGS stations are near the 295°E longitude chain distributed in the North and South hemispheres. The information on the stations is provided in Table 1. The GPS data are available at ftp://cddis.gsfc.nasa.gov/gps/data/daily. Corresponding ephemeris files can also be downloaded from this website.

TEC from GIMs provided by the IGS final product are used as the reference value in this study. GIM files are in IONEX format (−87.5°:2.5°:87.5°N in latitude, −180°:5°:180°E in longitude, 2 h time resolution).

Table 1 Test stations coordinates

Station	City	Location	Geographic latitude (°)	Geographic longitude (°)	Geomagnetic latitude (°)	Height (m)
IQAL	Iqaluit	Canada	63.7560N	291.4895E	73.25N	91.70
HLFX	Halifax	Canada	44.6835N	296.3887E	54.04N	3.10
ABMF	Les Abymes	France	16.2623N	298.4725E	25.66N	−25.60
UNSA	Salta	Argentina	24.7275S	294.5924E	14.97S	1257.80
MGUE	Malargue	Argentina	35.7774S	290.6021E	25.92S	1553.70
PALM	Palmer Station	Antarctica	64.7751S	295.9489E	55.02S	31.24

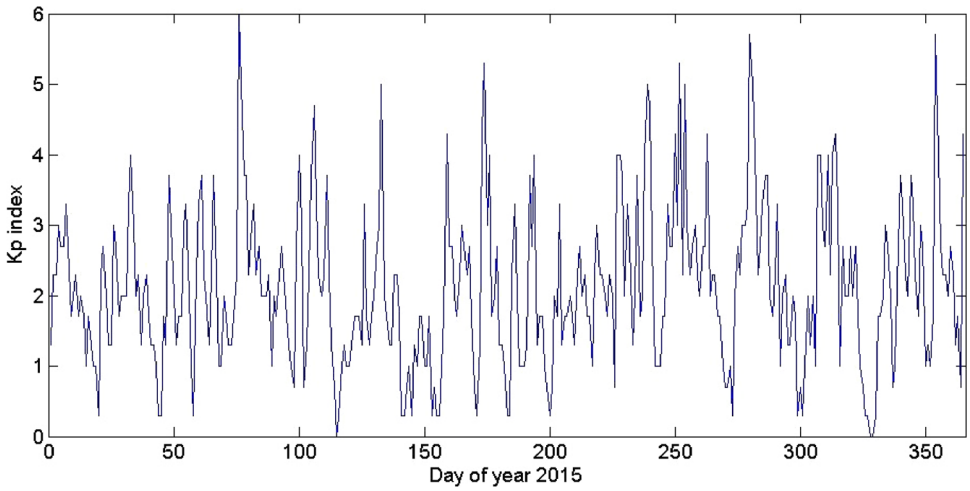


Fig. 1 Daily mean Kp index in the year of 2015

The one-day predicted ionosphere maps (C1PG file) of CODE are used to fit the coefficients of NTCM-BC model (García-Rigo 2011). CODE is the first IGS Associate Analysis Center (IAAC) that generate ionospheric predicted product. Their model is based on the extrapolation of the Spherical Harmonics coefficients (Schaer 1999). Their 1 and 2 days ahead VTEC maps (named C1PG and C2PG, respectively) have been released for years via internet. Final and predicted GIMs are available at ftp://cddis.gsfc.nasa.gov/gnss/products/ionex. The detailed descriptions of the coefficient fitting method of NTCM-BC model are shown in the works of Hoque and Jakowski (2015).

In this study, results of the three models are shown for several days in the years 2015 featured by different geomagnetic activity that, as known, affects ionospheric activity. The planetary magnetic activity index Kp is used to measure the general level of geomagnetic activity, and the Kp indices are available at http://www.ngdc.noaa.gov/stp/GEOMAG/kp_ap.html. A geomagnetic day is considered quiet when the Kp index satisfies the condition $0 \leq \mathrm{Kp} < 3.0$. In Fig. 1 the daily mean Kp index is plotted for the examination period.

The daily TEC deviation from the quiet days is also a basis for determination of the test days during quiet and disturbed ionosphere conditions in this study. We first take the GIMs monthly median value as the reference value. Then, the root-mean-square error (RMS) can be calculated for each day using formula as follows,

$$\mathrm{RMS} = \sqrt{\frac{\sum_{k=1}^{12}\sum_{j=1}^{73}\sum_{i=1}^{71}\left(\mathrm{VTEC}_{i,j,k} - \mathrm{VTEC}_{i,j,k}^{\mathrm{ref}}\right)^2}{N}} \tag{15}$$

where $N = 12 \times 73 \times 71$ is the number of VTEC values of 12 IONEX maps with 2-h time spacing. The grid number of an IONEX map is equal to $73 \times 71 = 5183$. RMS can reflect the deviation from the quiet days.

In Table 2, the selected analyzed Day Of Year (DOY) are classified for K_p value, related to geomagnetic activity. Figure 2 shows the RMS between monthly median GIMs and GIMs during the selected days.

As shown in Fig. 2, the RMS for each selected geomagnetically quiet day is around 2–3 TECU. For the six geomagnetically disturbed days, the RMS of GIMs are around 4–7 TECU.

Table 2 List of days used in this study with their minimum, maximum 3-h Kp index and daily mean Kp index

Geomagnetic activity	Date	DOY	Maximum 3-h Kp index	Minimum 3-h Kp index	Daily mean Kp index
Quiet	06/04/2015	155	1.0	0	0.3
	07/03/2015	184	0.7	0	0.3
	07/21/2015	201	1.3	0	0.7
	08/14/2015	226	1.0	0	0.7
	08/31/2015	243	1.7	0.3	1.0
	12/30/2015	364	1.3	0	0.7
Disturbed	03/17/2015	76	7.7	2.0	6.0
	06/08/2015	159	6.0	3.0	4.3
	08/27/2015	239	6.3	4.0	5.0
	09/09/2015	252	6.0	4.0	5.3
	10/08/2015	281	5.7	4.3	5.0
	10/13/2015	286	4.3	2.0	3.7

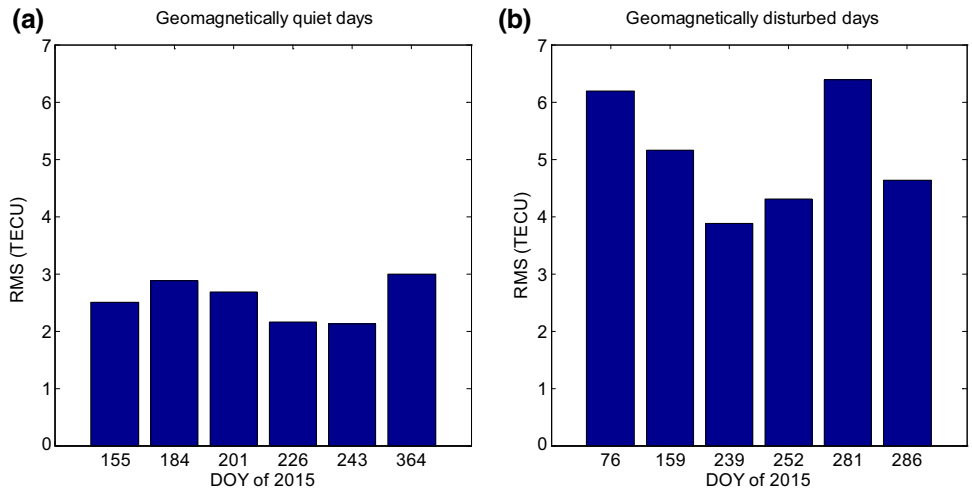


Fig. 2 RMS (in TECU) between monthly median GIMs and GIMs during selected: **a** geomagnetically quiet days; **b** geomagnetically disturbed days

For comparisons, the STEC along the satellite-to-receiver ray path are interpolated with IGS final GIMs. Using different ionospheric models, the STEC will also be calculated. From IGS final GIMs, STEC can be interpolated by employing the four-point formula (Schaer 1999).

$$\mathrm{STEC} = \mathrm{VTEC} \cdot M(\varepsilon) \tag{16}$$

$$\mathrm{VTEC} = (1-p)(1-q)E_{0,0} + p(1-q)E_{1,0} + q(1-p)E_{0,1} + pqE_{0,1}, \tag{17}$$

where VTEC is the TEC interpolated at the desired geographic longitude λ and latitude β; $E_{i,j}$ are the four nearest TEC grid values; p, q are multipliers within the interval [0,1]. p, q can be calculated as follows,

$$p = (\lambda - \lambda_0)/\Delta\lambda \tag{18}$$

$$q = (\beta - \beta_0)/\Delta\beta, \tag{19}$$

where (λ_0, β_0) are the geographic coordinates associated with $E_{0,0}$; $\Delta\lambda$, $\Delta\beta$ are the grid widths in longitude and latitude.

After obtaining the STEC along the satellite-to-receiver ray path, we use these STEC values to correct ionospheric time delay in the point positioning technique. The position results are then compared. GNSS point positioning method is based on C1 pseudorange and trilateration technique. The equation is as follows:

$$\rho - c\delta t_\mathrm{s} - d_\mathrm{T} - d_\mathrm{I} = \sqrt{(X^\mathrm{s} - X)^2 + (Y^\mathrm{s} - Y)^2 + (Z^\mathrm{s} - Z)^2} + c\delta t_\mathrm{u} + \varepsilon_\rho, \tag{20}$$

where ρ is the pseudorange measurement, $c\delta t_\mathrm{u}$ is the receiver clock offset scaled by the speed of light c, $(X^\mathrm{s}, Y^\mathrm{s}, Z^\mathrm{s})$ are satellite coordinates, and (X, Y, Z) are receiver coordinates. $c\delta t_\mathrm{s}$ is the satellite clock offset scaled by the speed of light, d_T is the tropospheric delay, d_I is the ionospheric delay, and ε_ρ contains other error sources. The least square method can be used to solve this point positioning problem when more than four satellites are observed.

From the IGS solution (Software/technique) INdependent EXchange Format (SINEX) file, the reference coordinates of the six IGS stations used in this article are extracted. Finally, the point positioning coordinates with different STEC corrections are compared with the reference values.

Results and discussion

Comparisons of reconstructed TEC maps with the ionospheric models

The GIMs were predicted by the three models mentioned above. Based on the eight coefficients from the GPS broadcast ephemeris, the Klobuchar model calculated the global TEC map. Using nonlinear fitting of TEC values from C1PG GIMs, NTCM-BC model reconstructed the global TEC map. With the three coefficients determining the Az variation with MODIP broadcasted to the user in the Galileo navigation message, the global TEC map can be outputted by the NeQuick G model.

We consider the TEC from the IGS final GIMs as the reference values and calculate the RMS of the TEC from different ionospheric models. All the GIMs grid data ($71 \times 73 \times 12$) are utilized to achieve the RMS value using formula (15).

For the C1PG GIMs file was used in NTCM-BC model in the analysis, its deviation from the final GIMs was also calculated. The RMS results are shown in Fig. 3.

From the RMS values in Fig. 3, we can conclude that NeQuick G and NTCM-BC models improved the GIMs reconstruction results compared with Klobuchar model on the test days. During the geomagnetically quiet days, the improvements are evident, and the performances of the NeQuick G model and NTCM-BC model were relatively

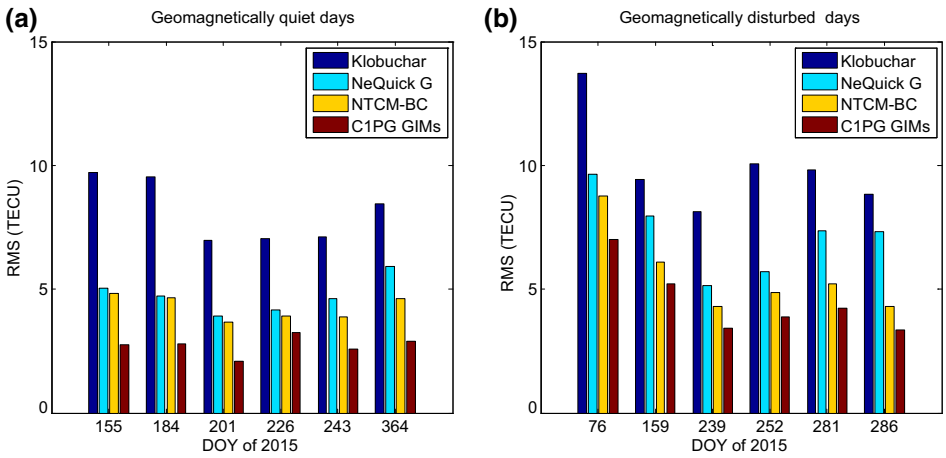


Fig. 3 RMS of GIMs obtained using different ionospheric models compared with IGS final GIMs during: **a** geomagnetically quiet days; **b** geomagnetically disturbed days

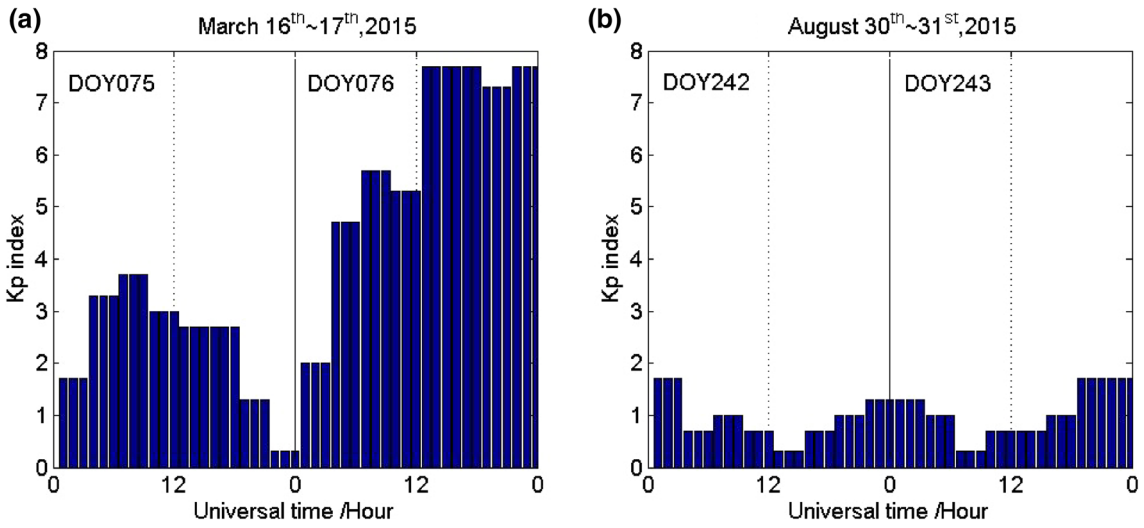
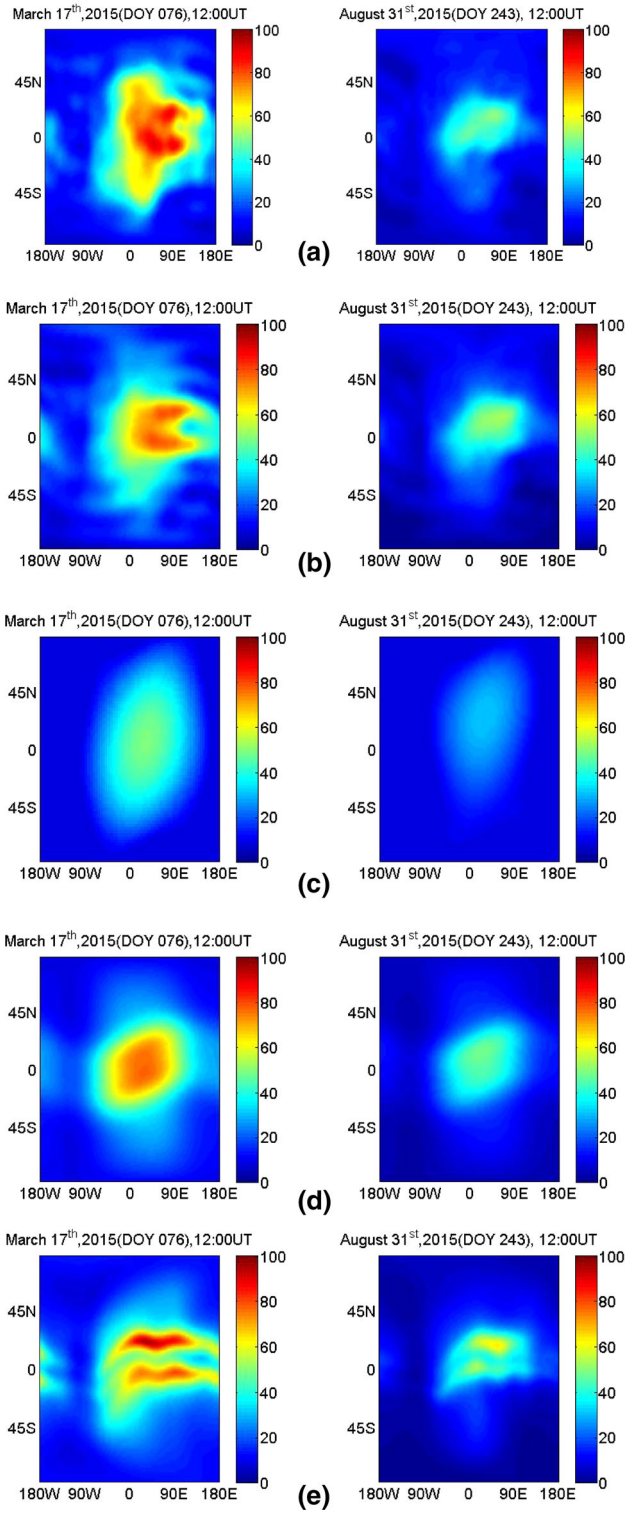


Fig. 4 Geomagnetic Kp indices during **a** March 16th to 17th, 2015 (DOY 075–076) and **b** August 30th to 31st, 2015 (DOY 242–243)



◀**Fig. 5** Global ionospheric TEC maps (in TECU) at 12:00UT from IGS GIMs (**a**), C1PG GIMs (**b**), and ionospheric models (**c**–**e**). **a** Global ionospheric TEC map at 12:00 UT on March 17, 2015 (DOY 076) (*left*) and August 31, 2015 (DOY 243) (*right*) from IGS final GIMs (unit: TECU). **b** Global ionospheric TEC map at 12:00UT on March 17, 2015 (DOY 076) (*left*) and August 31, 2015 (DOY 243) (*right*) from C1PG GIMs (unit: TECU). **c** Global ionospheric TEC map at 12:00 UT on March 17, 2015 (DOY 076) (*left*) and August 31, 2015 (DOY 243) (*right*) from Klobuchar model (unit: TECU). **d** Global ionospheric TEC map at 12:00 UT on March 17, 2015 (DOY 076) (*left*) and August 31, 2015 (DOY 243) (*right*) from NTCM-BC model (unit: TECU). **e** Global ionospheric TEC map at 12:00 UT on March 17, 2015 (DOY 076) (*left*) and August 31, 2015 (DOY 243) (*right*) from NeQuick G model (unit: TECU)

close to each other. From Fig. 3b, the NTCM-BC model achieved least RMS of GIMs on the geomagnetically disturbed days.

Furthermore, we select two days to show the detailed results, that is, March 17, 2015 (DOY 076) and August 31, 2015 (DOY 243). Figure 4 shows the geomagnetic Kp indices during the two days and the previous days. The geomagnetic conditions show more activity during DOY 076 than that during DOY 075. However, the geomagnetic activities during DOY 243 and DOY 242 were both very quiet.

Figure 5a shows the IGS final GIMs at 12:00 UT of the two days. As the geomagnetic activity levels of the two days are not the same, the TEC levels are also quite different. From Fig. 5a, b, the TEC values from C1PG GIMs on DOY 243/2015 are closely followed by the data from IGS final GIMs. However, the two kinds of GIMs disagree on DOY 076/2015. Therefore, the precision of the predicted TEC by C1PG GIMs files is correlated with the geomagnetic activity. Figure 5c shows the global TEC map at 12:00 UT obtained using Klobuchar model. Comparing Fig. 5c with 5a, evident differences between the IGS final GIMs and Klobuchar GIMs near the Equator are observed where the maximum ionospheric activity is expected. Furthermore, the global pattern of the ionospheric TEC described by Klobuchar model is relatively simple. Figure 5d shows the global TEC map at 12:00 UT calculated with NTCM-BC model. The NTCM-BC GIMs can represent the general spatial trend of the C1PG GIMs (Fig. 5b), whereas the detailed characteristics are not demonstrated, especially for the areas near the Equator. The GIMs from NeQuick G model at 12:00 UT are shown in Fig. 5e, which detail the variation of ionospheric TEC in areas near the Equator. However, the same cannot be said regarding the IGS GIMs.

The error statistic for the models results are shown in Table 3, and the histograms of the biases are shown in Figs. 6 and 7. Differences are observed between the one-day ahead C1PG GIMs and the IGS final GIMs, which are especially significant on the geomagnetic activity day (DOY 076).

Comparison of STEC from different models

The second performance analysis deals with differences between the STEC from IGS final GIMs and those

Table 3 Statistical comparisons of ionospheric models with IGS final GIMs in TECUs

Model	Average bias		RMS		Maximum bias		Minimum bias	
	DOY 076	DOY 243	DOY 076	DOY 243	DOY 076	DOY 243	DOY 076	DOY v243
C1PG	−0.90	−1.08	7.02	2.58	32.2	12.20	−52.2	−19.60
NeQuick G	−0.14	0.38	9.62	4.62	45.81	24.56	−57.69	−21.98
NTCM-BC	−1.02	−1.00	8.78	3.89	23.56	16.10	−61.91	−22.86
Klobuchar	−7.08	−1.10	13.72	7.11	23.07	14.83	−64.90	−37.25

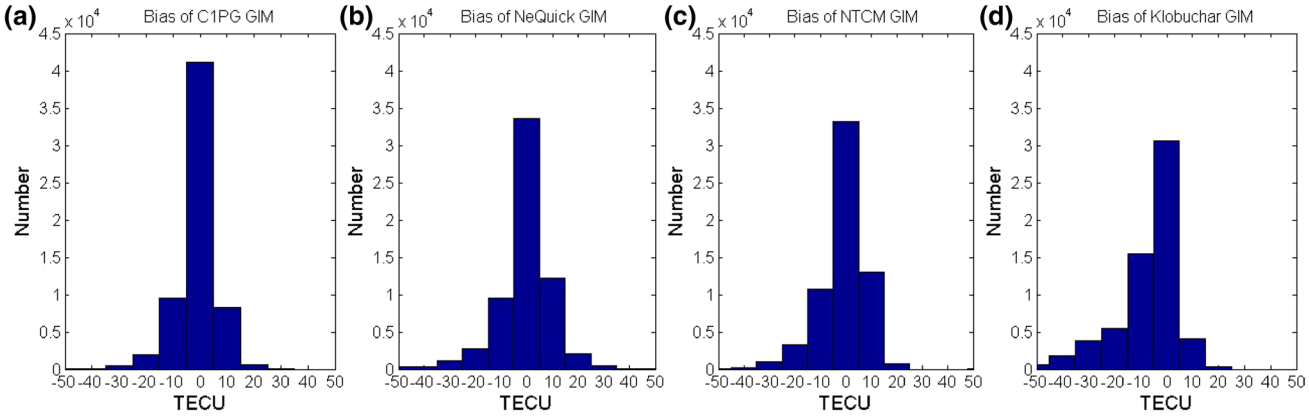


Fig. 6 Histogram of bias of modeled GIM TEC values compared with IGS final GIMs on March 17, 2015 (DOY 076)

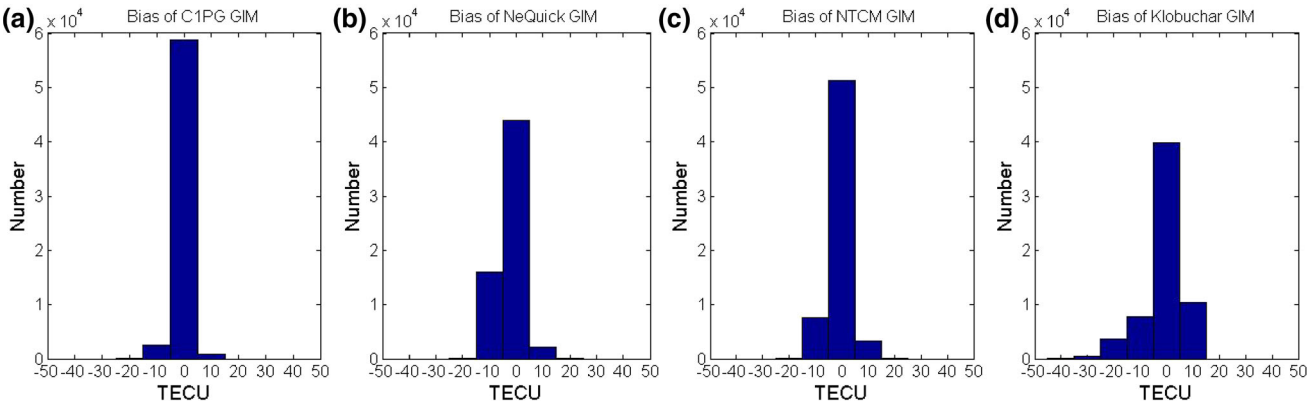


Fig. 7 Histogram of bias of modeled GIM TEC values compared with IGS final GIMs on August 31, 2015 (DOY 243)

computed with each ionospheric model. STEC at the six test stations are compared, and biases of the modeled STEC are made available.

Figure 8 shows the RMS of modeled STEC compared with STEC from IGS final GIMs. From Fig. 8a, b, we can see that the differences between RMS of STEC at high-latitude stations and that at the mid- and low-latitude stations are more significant on geomagnetically disturbed days. For the geomagnetically quiet days, the performances of the three models are quite different at high-latitude stations. While their RMS at high-latitude stations are close to each other during the geomagnetically quiet days. STEC from NeQuick G and NTCM-BC models tend to have less bias than those from Klobuchar model in most situations.

The STEC results of the six stations on March 17, 2015 (DOY 076) and August 31, 2015 (DOY 243) are shown in Figs. 9, 10, 11, 12, 13, and 14. In these figures, the biases on March 17, 2015 (DOY 076) are shown in the first column, and the biases on August 31, 2015 (DOY 243) are presented in the second column. The statistical RMS results of the biases are shown at the third column of each figure.

The STEC biases predicted by the three models on DOY 243 are less than those on DOY 076 as a result of the high geomagnetic level.

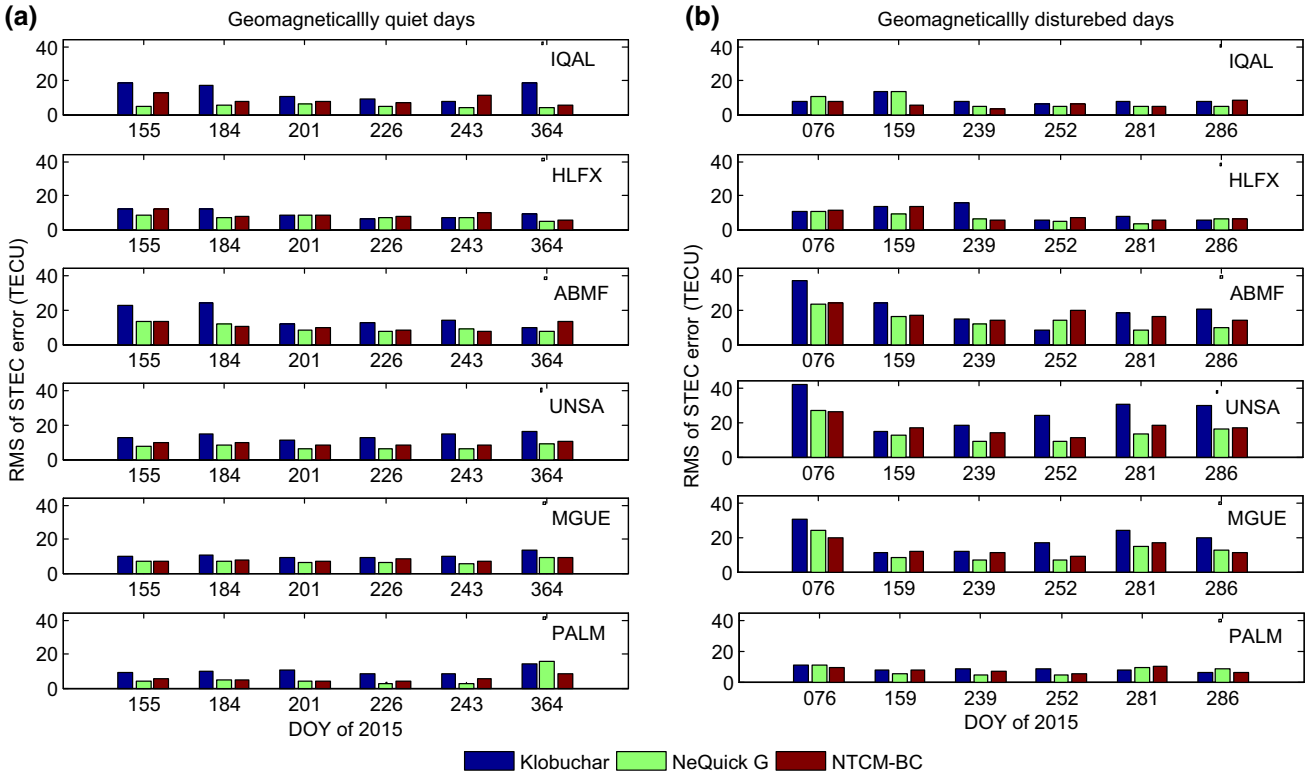


Fig. 8 RMS of STEC obtained using different ionospheric models compared with STEC from IGS final GIMs at all the test stations during: **a** geomagnetically quiet days; **b** geomagnetically disturbed days

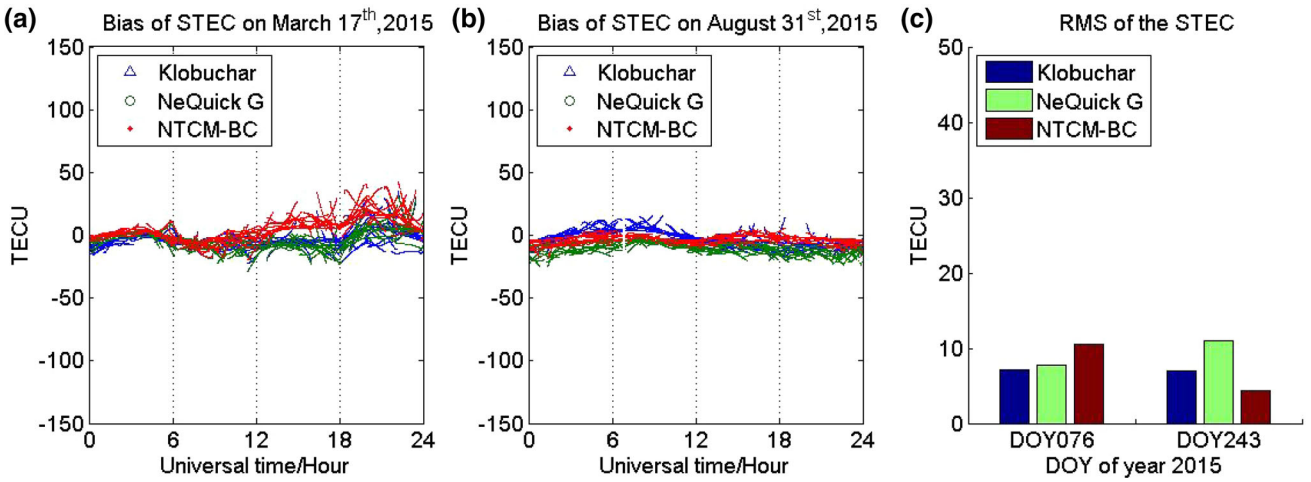


Fig. 9 STEC error statistics of ionospheric models compared with IGS final GIMs at IQAL station (geomagnetic latitude: 73.25°N): **a** STEC bias on March 17, 2015 (DOY 076); **b** STEC bias on August 31, 2015 (DOY 243); **c** RMS of STEC

On DOY 243, the RMS values of STEC are approximately 3–15 TECU. NeQuick G and NTCM-BC models have less STEC prediction bias and RMS than that of Klobuchar model, except for the Northern hemispheric stations IQAL and HLFX. At these two stations, the NeQuick G model produced relatively low STEC values the whole day. Therefore, the RMS value of STEC from NeQuick G model is larger than those of the two other models, which can be seen from Figs. 9 and 10. During this test period (DOY 243), the general trend characteristic of the global ionosphere was quite significant. All the three models appear to achieve relatively good STEC prediction results.

In the case of STEC prediction on DOY 76, the RMS values are approximately 7 $\sim$ 40 TECU. NeQuick G and NTEC-BC

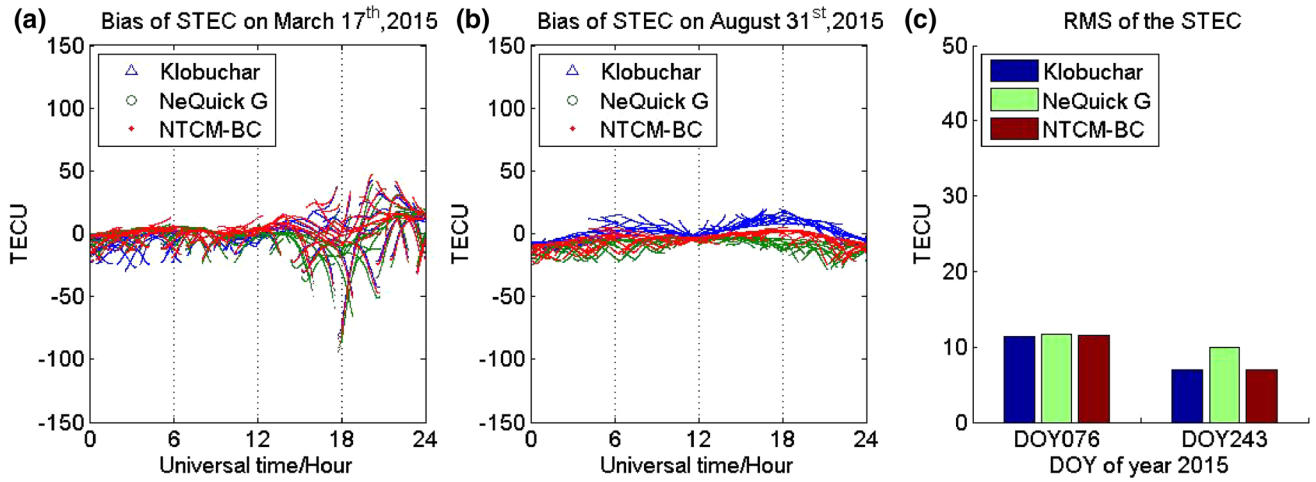


Fig. 10 STEC error statistics of ionospheric models compared with IGS final GIMs at HLFX station (geomagnetic latitude: 54.04°N): **a** STEC bias on March 17, 2015 (DOY 076); **b** STEC bias on August 31, 2015 (DOY 243); **c** RMS of STEC

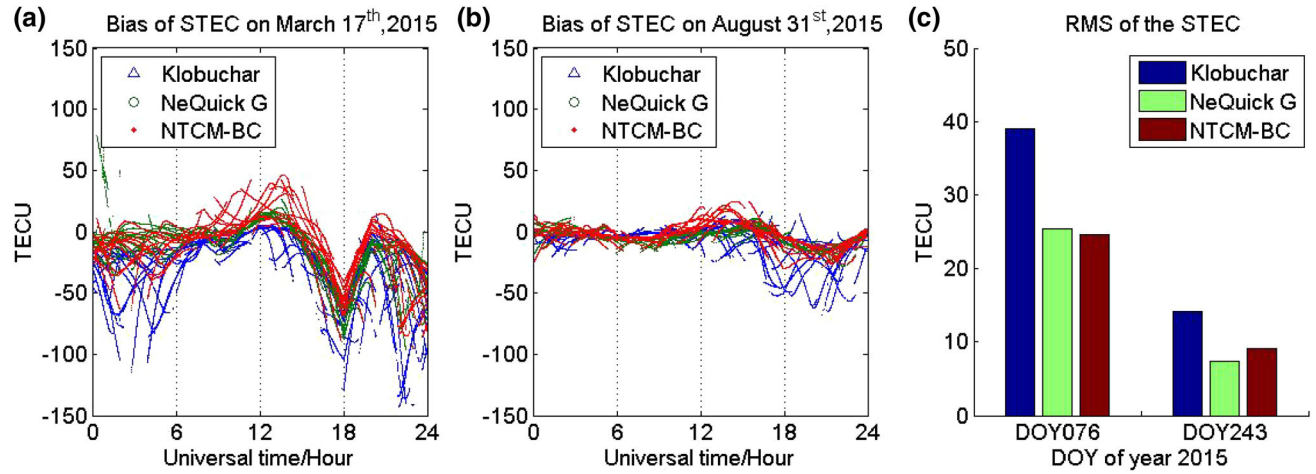


Fig. 11 STEC error statistics of ionospheric models compared with IGS final GIMs at ABMF station (geomagnetic latitude: 25.66°N): **a** STEC bias on March 17, 2015 (DOY 076); **b** STEC bias on August 31, 2015 (DOY 243); **c** RMS of STEC

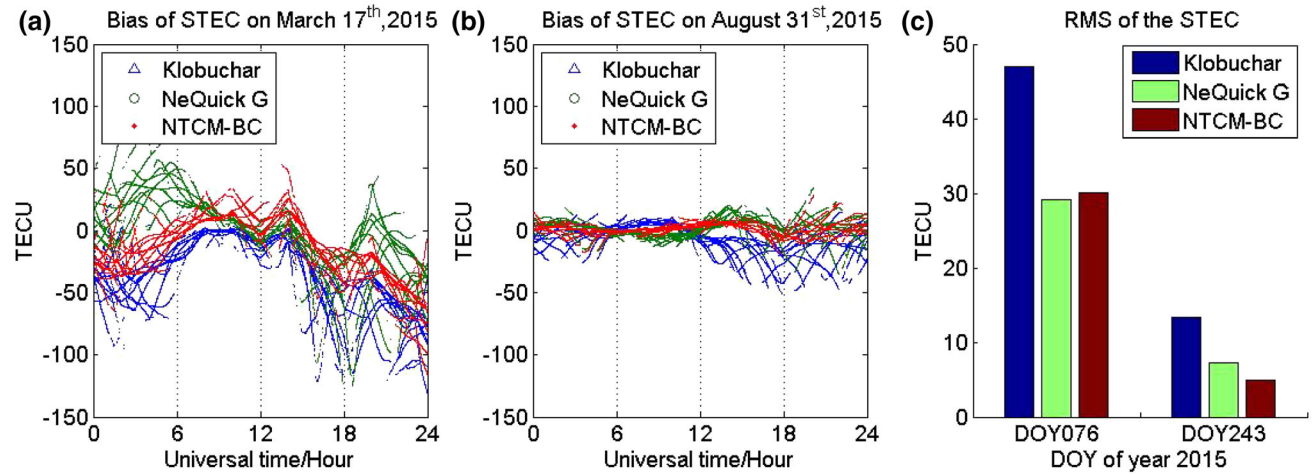


Fig. 12 STEC error statistics of ionospheric models compared with IGS final GIMs at UNSA station (geomagnetic latitude: 14.97°S): **a** STEC bias on March 17, 2015 (DOY 076); **b** STEC bias on August 31, 2015 (DOY 243); **c** RMS of STEC

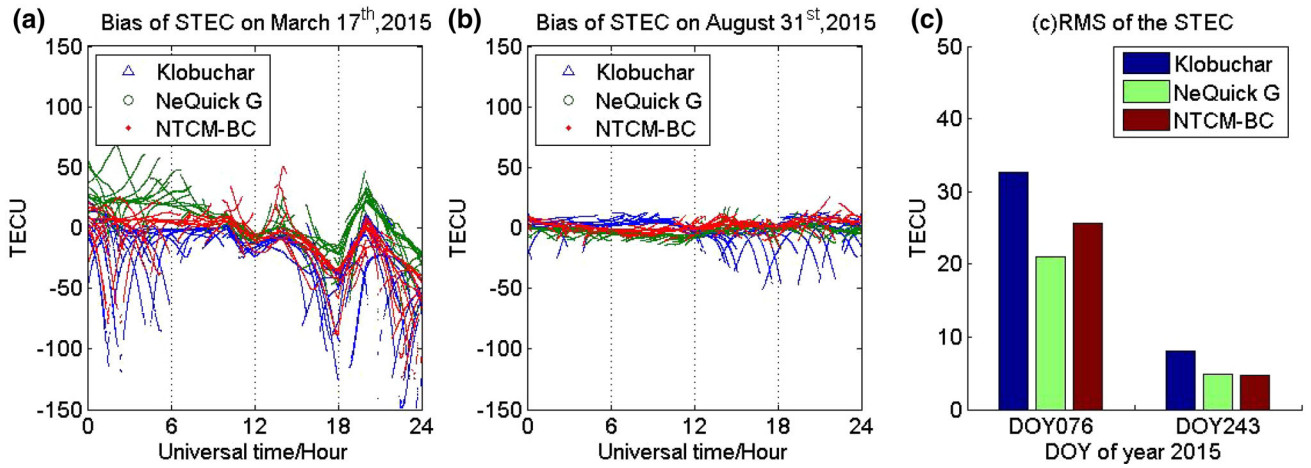


Fig. 13 STEC error statistics of ionospheric models compared with IGS final GIMs at MGUE station (geomagnetic latitude: 25.92°S): **a** STEC bias on March 17, 2015 (DOY 076); **b** STEC bias on August 31, 2015 (DOY 243); **c** RMS of STEC

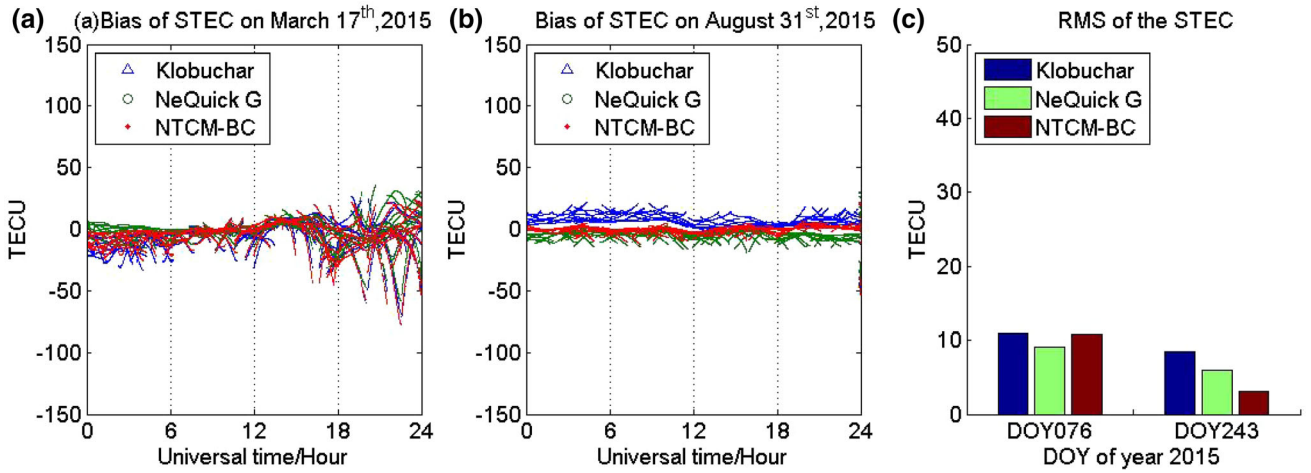


Fig. 14 STEC error statistics of ionospheric models compared with IGS final GIMs at PALM station (geomagnetic latitude: 55.02°S): **a** STEC bias on March 17, 2015 (DOY 076); **b** STEC bias on August 31, 2015 (DOY 243); **c** RMS of STEC

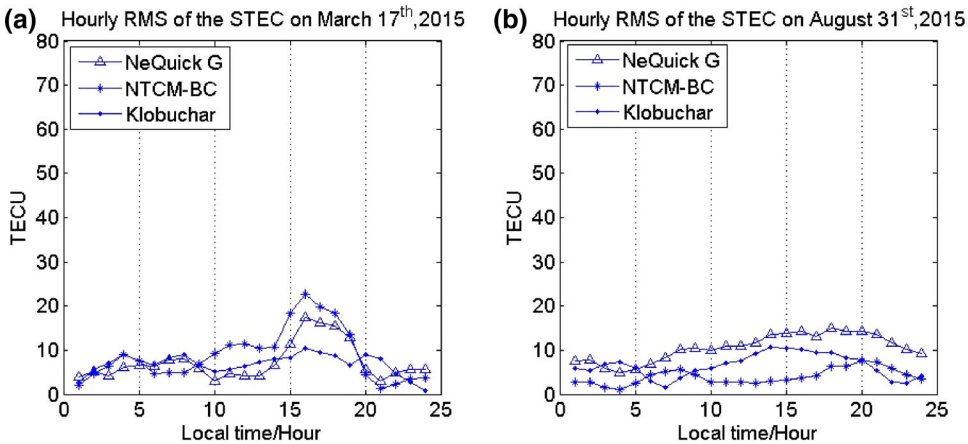


Fig. 15 Hourly RMS of STEC from ionospheric models compared with IGS final GIMs at IQAL station on: **a** March 17, 2015 (DOY 076); **b** August 31, 2015 (DOY 243)

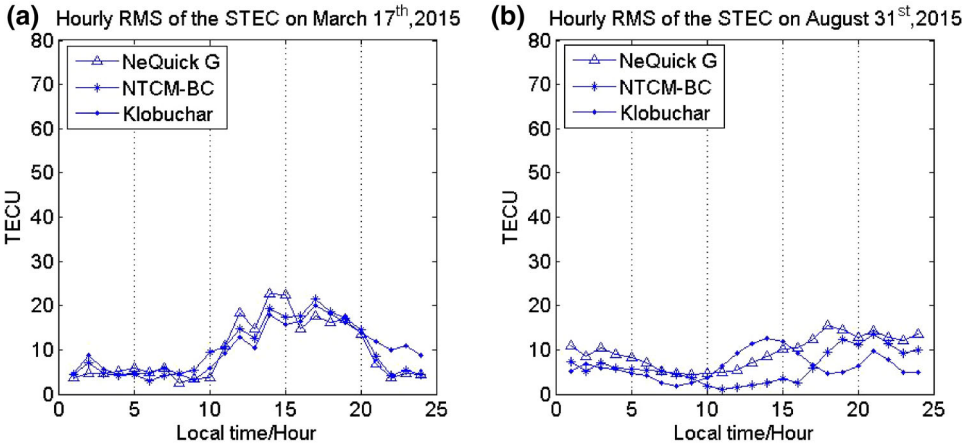


Fig. 16 Hourly RMS of STEC from ionospheric models compared with IGS final GIMs at HLFX station on: **a** March 17, 2015 (DOY 076); **b** August 31, 2015 (DOY 243)

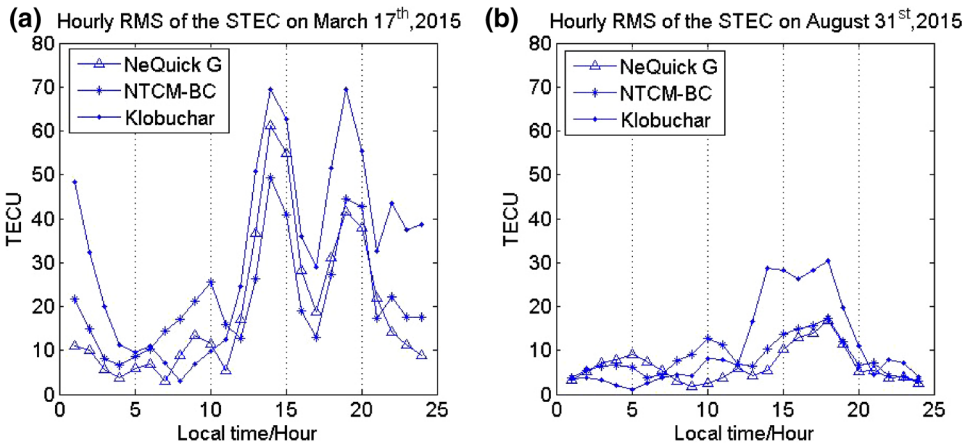


Fig. 17 Hourly RMS of STEC from ionospheric models compared with IGS final GIMs at ABMF station on: **a** March 17, 2015 (DOY 076); **b** August 31, 2015 (DOY 243)

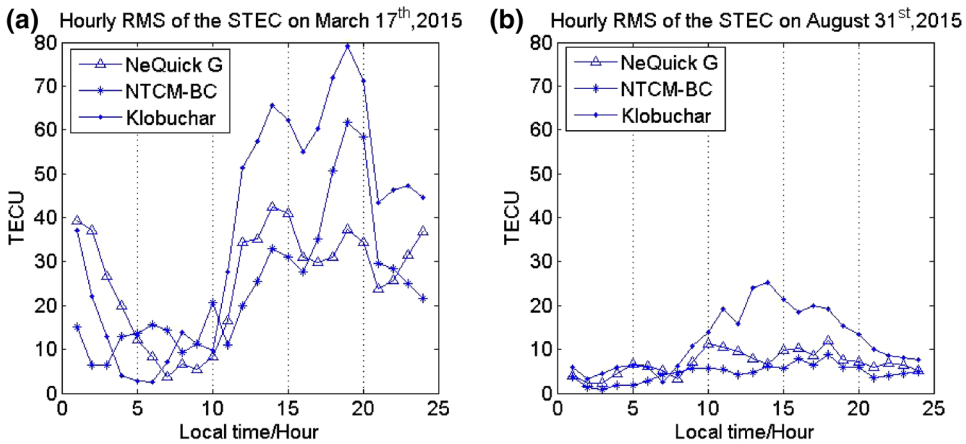


Fig. 18 Hourly RMS of STEC from ionospheric models compared with IGS final GIMs at UNSA station on: **a** March 17, 2015 (DOY 076); **b** August 31, 2015 (DOY 243)

model improve the STEC prediction results of the Klobuchar model at the mid- and low-latitude stations considerably. Except for IQAL stations, STEC are generally underestimated at daytime by the three models because of the large prediction error under the disturbed geomagnetic condition.

From the results of the STEC prediction performance on DOY 076 and DOY 243, we conclude that the STEC prediction performance of NTCM-BC was comparable to that of NeQuick G model.

As can be seen from Figs. 9, 10, 11, 12, 13, and 14, the biases tend to be larger during the second half of the day (in UT) during both days. The six stations are located along the 295°E meridian. The local time near this longitude chain is about 5 h late than universal time. Therefore, the

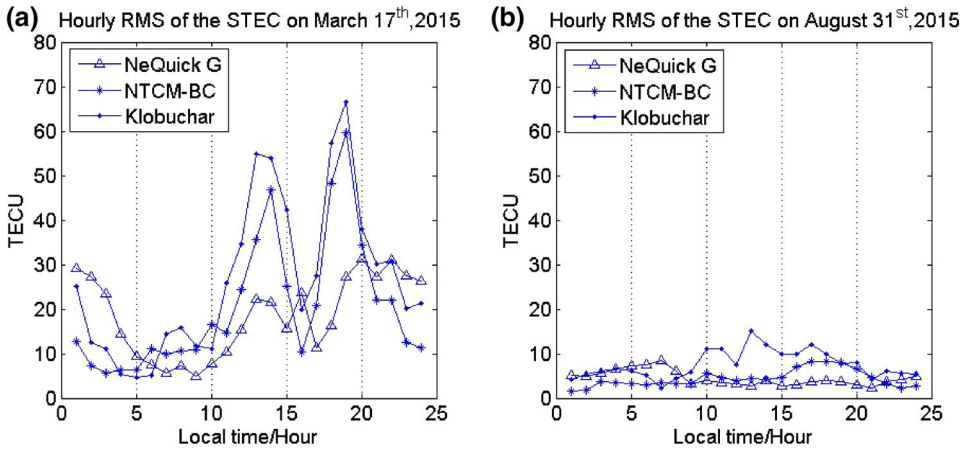


Fig. 19 Hourly RMS of STEC from ionospheric models compared with IGS final GIMs at MGUE station on: **a** March 17, 2015 (DOY 076); **b** August 31, 2015 (DOY 243)

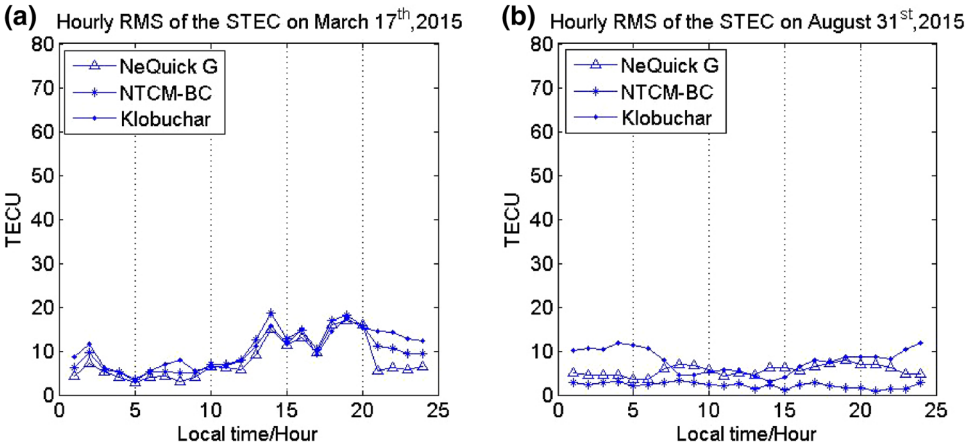


Fig. 20 Hourly RMS of STEC from ionospheric models compared with IGS final GIMs at PALM station on: **a** March 17, 2015 (DOY 076); **b** August 31, 2015 (DOY 243)

relatively large model biases mainly occurred during the daytime (0700–1900 LT).

We analyzed the RMS values of STEC at different local times, and the results are shown in Figs. 15, 16, 17, 18, 19 and 20.

Figures 15, 16, 17, 18, 19 and 20 imply that the STEC biases predicted by the three models change dramatically with local time.

Under quiet geomagnetic conditions, the ionospheric activity level on DOY 243 is relative low. The STEC RMS errors predicted by the NeQuick G model and the NTCM-BC model changed slightly with local time. However, the STEC RMS errors from Klobuchar model at stations AMBF, UNSA and MGUE were considerably larger during the period from 1000 LT to 2000 LT compared to other local times.

On DOY 076, the STEC RMS errors from 1000 LT to 2000 LT were obviously larger than that at other local times at all the stations for the three models. The NeQuick G model and the NTCM-BC model reduced the prediction errors dramatically. At 1900 LT, the RMS error from Klobuchar model reached 79.27TECU at UNSA station, while the RMS error from NeQuick G model was 37.37TECU. However, Figs. 9 and 10 showed that the NeQuick G model gave better predictions during daytime but worse predictions during nighttime at stations UNSA and MGEU compared to the other two models. We think that the daily effective ionization levels (Az) used in the NeQuick G model at these two stations maybe more in line with the situation during the daytime.

Ionosphere correction effect of different models on point positioning

The third performance analysis is carried out in the position domain. We calculate the errors by obtaining the difference between the actual position of the test stations and the ones computed with and without each ionospheric correction. Figure 21 shows the RMS of point positioning result using different models.

From these RMS values, the point positioning results without ionosphere correction show the largest error. Using ionosphere models, the RMS values decrease considerably, especially at mid- and low-latitude stations. With the ionospheric delay correction derived from IGS final GIMs,

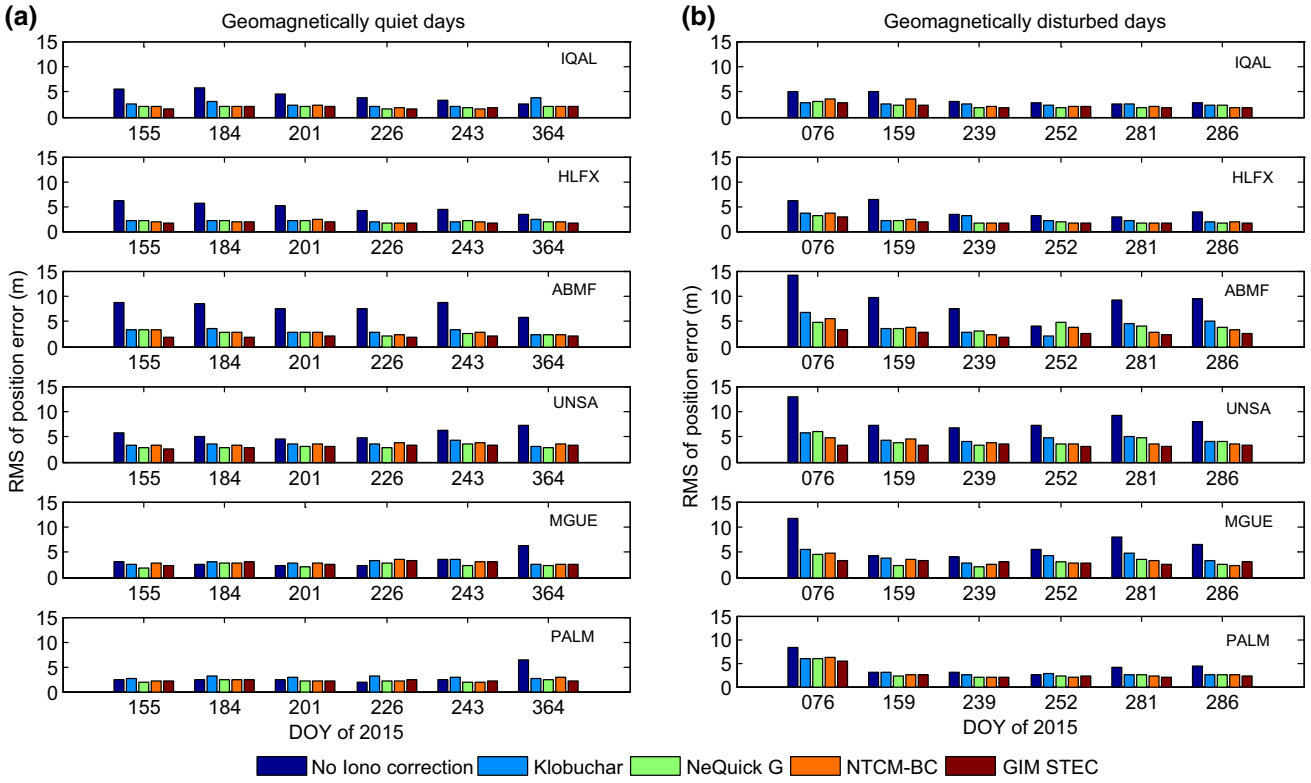


Fig. 21 RMS of point positioning errors obtained using STEC correction from different ionospheric models at all the test stations during: **a** geomagnetically quiet days; **b** geomagnetically disturbed days

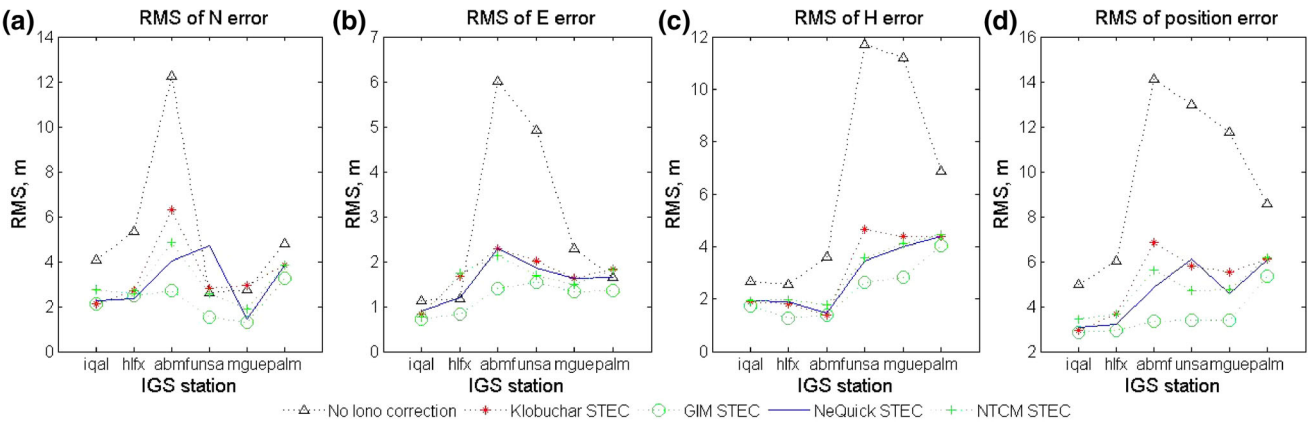


Fig. 22 RMS of coordinate errors in the North, East, and Up component at different stations on March 17, 2015 (DOY 076): **a** north error, **b** east error, **c** up error, and **d** position error

the RMS of position errors reaches minimum value at each station. In general, the performances of NTCM-BC and NeQuick G model are better than those of Klobuchar model. Moreover, the NTCM-BC model tends to behavior more stable than NeQuick G model.

The point positioning results of the six stations on March 17, 2015 (DOY 076) and August 31, 2015 (DOY 243) are shown in Figs. 22 and 23. The errors are examined in terms of RMS for the three components in local North-East-Up system.

Figure 22 shows the RMS of the coordinate errors in North, East, and Up components, as well as the position error at the six IGS stations on March 17, 2015 (DOY 076). In this figure, the six stations are arranged according to their latitude from left to right.

Figure 23 shows the RMS values of the point positioning results using different ionospheric correction models on August 31, 2015 (DOY 243).

During DOY 243, a geomagnetic quiet day, the point positioning error decreases significantly. Although the

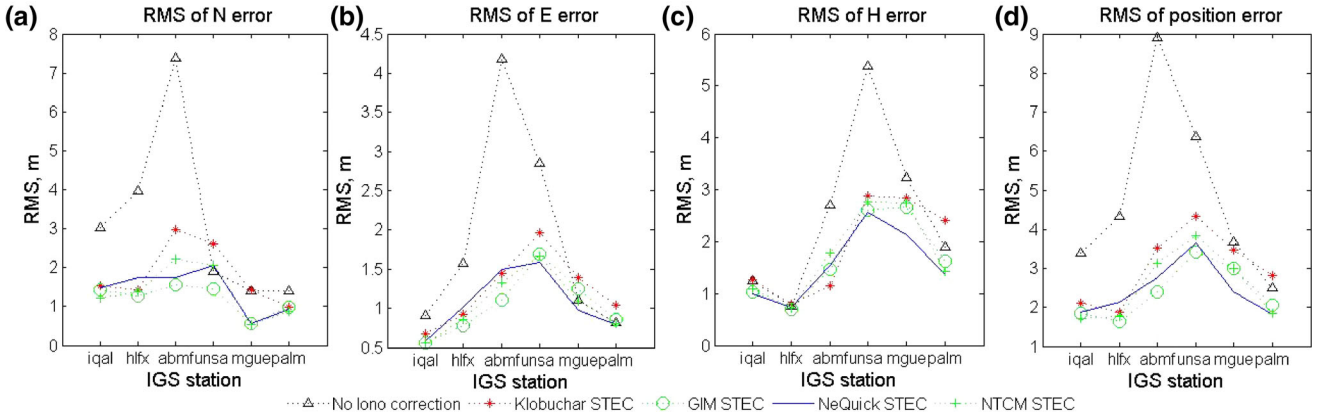


Fig. 23 RMS of coordinate errors in North, East, and Up components at different stations on August 31, 2015 (DOY 243): **a** north error, **b** east error, **c** up error, and **d** position error

performance of NTCM-BC and NeQuick G model at most stations are better than those of Klobuchar model, the point positioning RMS values using different ionosphere models are relatively close. Moreover, the point positioning results with ionospheric delay correction derived from IGS GIMs show relatively similar performances with NTCM-BC and NeQuick G model.

Conclusion

This study presents the performances of three ionospheric models associated with GNSS navigation message for single-frequency receiver user. The selected models include the Klobuchar, NTCM-BC, and NeQuick G models. The analysis is carried out by comparing the IGS final GIMs data, thereby providing independent and precise vertical TEC (at the level of a few TECU). From the results of the statistical analyses, we draw the following main conclusion.

1. The errors of GIMs, STEC, and point positioning using ionospheric models are higher during geomagnetically disturbed days than that during geomagnetically quiet days. As the geomagnetic activity levels of the two days are not the same, the precisions of the predicted ionospheric model parameters are also quite different. Therefore, the performances of the models in TEC prediction and ionospheric time delay correction for singe point positioning are closely related with the geomagnetic activity.
2. From the global ionospheric TEC results, both the NeQuick G and NTCM-BC models outperform the Klobuchar model during the test days.
3. In terms of STEC, NeQuick G and NTEC-BC model improve the STEC prediction results of the Klobuchar model observably at the mid- and low-latitude stations. The STEC prediction performance of NTCM-BC was comparable to that of NeQuick G model.
4. From the point positioning coordinate errors, the results are improved using ionospheric models, especially in mid- and low-latitude areas. On the geomagnetically quiet days, the differences among the results using the three models are not significant. In general, the performances of NTCM-BC and NeQuick G model are better than those of Klobuchar model during the geomagnetically disturbed days. Although the NTCM-BC model does not always achieve least RMS values of position errors, it tends to perform more stable than NeQuick G model during the test days.
5. The process of the NeQuick G model is relatively complicated compared with those of the Klobuchar and NTCM-BC models.
6. The NTCM-BC model can be constructed using nonlinear fitting of the predicted TEC data. Like the other two ionosphere models, NTCM-BC model can only represent the general trend of global TEC and cannot produce the detailed feature in some areas, such as the Equatorial anomaly area. Although the accuracy of NTCM-BC, in principle, is restricted because of its simple formulation, its use in GNSS navigation message has a certain advantage.

Acknowledgements This work is supported by the Fundamental Research Funds for the Central Universities (Grant Numbers 35832015084 and 35832015083) and the National Natural Science Foundation of China (Grant Numbers 41104025, 41574011 and 41401622).

References

Angrisano A, Gaglione S, Gioia C, Massaro M, Robustelli U (2013a) Assessment of NeQuick ionospheric model for Galileo single-

frequency users. Acta Geophys 61(6):1457–1476. doi:10.2478/s11600-013-0116-2

Angrisano A, Gaglione S, Gioia C, Massaro M, Troisi S (2013b) Benefit of the NeQuick Galileo version in GNSS single-point positioning. Int J Navig Observ 2013(3):32–36. doi:10.1155/2013/302947

Bidaine B, Lonchay M, Warnant R (2013) Galileo single frequency ionospheric correction: performances in terms of position. GPS Solut 17:63–73. doi:10.1007/s10291-012-0261-0

European Commission (2015), European GNSS (Galileo) open service—ionospheric correction algorithm for Galileo single frequency users, European Union

Farah A (2009) Comparison of GPS/Galileo single frequency ionospheric models with vertical TEC maps. Artif Satell 43(2):75–90. doi:10.2478/v10018-009-0008-5

Feltens J, Angling M, Jackson-Booth N, Jakowski N, Hoque M, Hernández-Pajares M, Aragón-Àngel A, Orús R, Zandbergen R (2011) Comparative testing of four ionospheric models driven with GPS measurements. Radio Sci 46:RS0D12. doi:10.1029/2010RS004584

Galileo Information Center (2006) Galileo open service signal in space interface control document, draft 0. OS SIS ICD, São José dos Campos

García-Rigo A, Monte E, Hernández-Pajares M, Juan JM, Sanz J, Aragón-Angel A, Salazar D (2011) Global prediction of the vertical total electron content of the ionosphere based on GPS data. Radio Sci 46:RS0D25. doi:10.1029/2010RS004643

Hochegger G, Nava B, Radicella S, Leitinger R (2000) A family of ionospheric models for different uses. Phys Chem Earth 25(4):307–310

Hoque MM, Jakowski N (2015) An alternative ionospheric correction model for global navigation satellite systems. J Geod 89:391–406. doi:10.1007/s00190-014-0783-z

Hoque MM, Jakowski N, Berdermann J (2015) An ionosphere broadcast model for next generation GNSS. ION GNSS + 2015, Tampa

Jakowski N (1996) TEC monitoring by using satellite positioning systems. In: Kohl H, Ruester R, Schlegel K (eds) Modern Ionos Sci. EGS, Katlenburg-Lindau, ProduServ GmbH Verlags service, Berlin

Jakowski N, Hoque MM, Mayer C (2011) A new global TEC model for estimating transionospheric radio wave propagation errors. J Geod 85(12):965–974. doi:10.1007/s00190-011-0455-1

Klobuchar JA (1987) Ionospheric time-delay algorithm for single frequency GPS users. IEEE Trans Aerosp Electron Syst 23(3):325–331

Klobuchar JA (1996) Ionospheric effects on GPS. In: Parkinson BW, Spilker JJ (eds) Global positioning system: theory and applications, vol I. American Institute of Aeronautics and Astronautics Inc., Washington, pp 485–515

Leitinger R, Zhang ML, Radicella SM (2005) An improved bottomside for the ionospheric electron density model NeQuick. Ann Geophys 48(3):525–534. doi:10.4401/ag-3217

Macalalad EP, Tsai L, Wu J, Liu C (2013) Application of the TaiWan Ionospheric Model to single-frequency ionospheric delay corrections for GPS positioning. GPS Solut 17:337–346. doi:10.1007/s10291-012-0282-8

Nava B, Coïsson P, Amarante GM, Azpilicueta F, Radicella SM (2005) A model assisted ionospheric electron density reconstruction method based on vertical TEC data ingestion. Ann Geophys 48(2):313–320. doi:10.4401/ag-3203

Nava B, Coïsson P, Radicella SM (2008) A new version of the NeQuick ionosphere electron density model. J Atmos Sol Terr Phys 70(15):1856–1862. doi:10.1016/j.jastp.2008.01.015

Orús R, Hernández-Pajares M, Juan JM, Sanz J, García-Fernández M (2002) Performance of different TEC models to provide GPS ionospheric corrections. J Atmos Sol Terr Phys 64:2055–2062. doi:10.1016/S1364-6826(02)00224-9

Radicella SM, Leitinger R (2001) The evolution of the DGR approach to model electron density profiles. Adv Space Res 27(1):35–40. doi:10.1016/S0273-1177(00)00138-1

Radicella SM, Nava B, Coïsson P (2008) Ionospheric models for GNSS single frequency range delay corrections. Fís Tierra 20:27–39

Schaer S (1999) Mapping and predicting the Earth's ionosphere using the Global Positioning System, Ph.D. thesis, Bern University, Bern

Wang N, Yuan Y, Li Z, Huo X (2016) Improvement of Klobuchar model for GNSS single-frequency ionospheric delay corrections. Adv Space Res 57(7):1555–1569. doi:10.1016/j.asr.2016.01.010

Yu X, She C, Zhen W, Bruno N, Liu D, Yue X, Ou M, Xu J (2015) Ionospheric correction based on ingestion of global ionospheric maps into the NeQuick 2 model. World J, Sci. doi:10.1155/2015/376702

Yuan Y, Ou J (2001) An improvement to ionospheric delay correction for single-frequency GPS users—the APR-I scheme. J. Geod. 75:331–336. doi:10.1007/s001900100182

Yuan Y, Huo X, Ou J, Zhang K, Chai Y, Wen D, Grenfell R (2008) Refining the Klobuchar ionospheric coefficients based on GPS observations. IEEE Trans Aerosp Electron Syst 44(4):1498–1510. doi:10.1109/TAES.2008.4667725