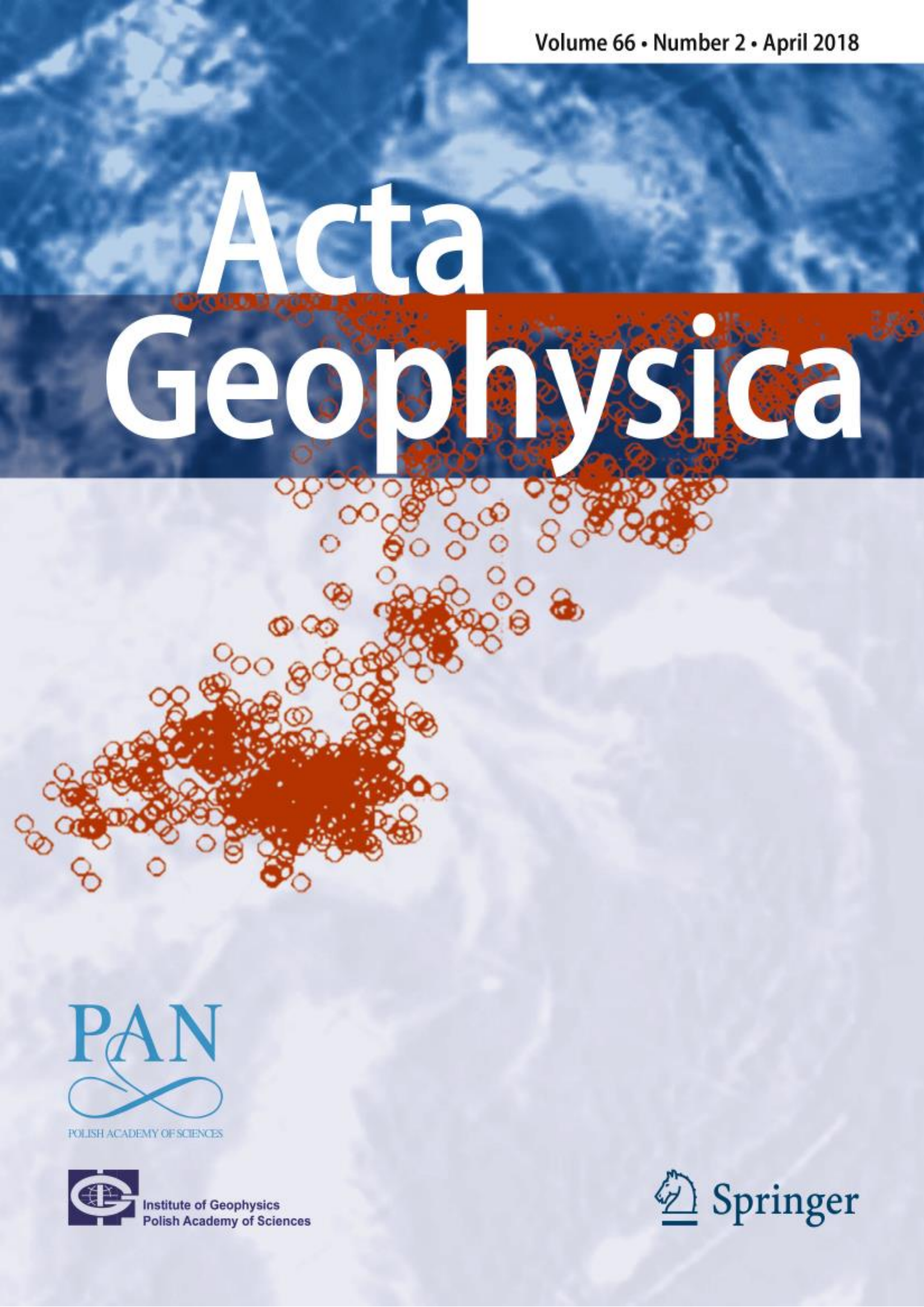
Volume 66 • Number 2 • April 2018

Acta Geophysica



PAN
POLISH ACADEMY OF SCIENCES

Institute of Geophysics
Polish Academy of Sciences

Springer

Acta Geophysica (2018) 66:121–130
https://doi.org/10.1007/s11600-018-0112-7

RESEARCH ARTICLE - SOLID EARTH SCIENCES

Simulation of the microtremor H/V spectrum based on the theory of surface wave propagation in a layered half-space

Zhen Zhang[1] · Xueliang Chen[1] · Mengtan Gao[1] · Zongchao Li[1] · Qianfeng Li[2]

Received: 17 July 2017 / Accepted: 17 January 2018 / Published online: 29 January 2018

Abstract
Subsurface velocity structures must be estimated to predict long-period ground motions and seismic hazards. Subsurface velocity structures can be constructed via an inversion of the horizontal-to-vertical (H/V) spectral ratio of microtremor (MHV) curves; thus, a method of simulating the MHV curves is key. In this study, we use the H/V spectral ratio of the surface wave (SHV) based on the surface wave propagation theory in a layered half-space to simulate the MHV curves at sites A and B of the Yuxi basin. Then, we attempt to analyze the features of the SHV curves. We find the H/V ratio of the microtremor loading source to be independent of the peak frequency of the SHV curve, but it has some relation to the amplitude of the SHV curve. Moreover, to reduce the error in subsurface velocity structures obtained by the MHV curves, we suggest that the SHV curves at near-peak frequencies should not be considered in the inversion, because the amplitude deviation is higher at the peak frequency of the MHV curve. In addition, the best frequency ranges for the inversion of the microtremor H/V spectrum are between the peak and trough frequencies of the microtremor H/V spectrum.

Keywords Seismic hazard · Velocity structure · Microtremor · Surface wave H/V spectrum · Microtremor loading source

Introduction

After the 1985 Mexico City earthquake, scientists realized that sedimentary layers amplify the amplitude and duration of long-period ground motions (Komatitsch et al. 2004; Shani-Kadmiel et al. 2012; Rong et al. 2016). Thus, the three-dimensional (3-D) velocity structures of the subsurface sedimentary layers must be estimated for the prediction of long-period ground motions (Gao et al. 2002; Chen et al. 2014).

Subsurface shallow velocity profiles can be determined from geotechnical or geophysical investigations. Compared with data obtained by drilling and seismic exploration, microtremor data are obtained by inexpensive techniques that do not require high logistical effort (Picozzi et al. 2005). Since the 1930s, Japanese scholars have been studying microtremors. To date, the methods of estimating subsurface velocity structures from microtremor array records have mainly fallen into two categories: (1) spatial autocorrelation analysis, which was initiated by Aki (1957) and (2) frequency-wavenumber spectral analysis, which was initiated by Capon (1969). Many studies have shown that the spatial autocorrelation analysis (Apostolidis et al. 2004; Claprood et al. 2011) and frequency-wavenumber spectral analysis (Satoh et al. 2001; Wu and Huang 2012) methods were effective for evaluating the subsurface S-wave velocity structures at different sites. However, since Nogoshi and Igarashi (1971) proposed the ratio between the Fourier spectral amplitudes of the horizontal and vertical components of microtremor recordings (MHV), the horizontal-to-vertical (H/V) curves of microtremors have become a common method for evaluation of the site effect (Bonnefoy-Claudet et al. 2006; Carcione et al. 2017). In addition, the H/V curves of microtremors have been widely used to evaluate subsurface velocity structures (Delgado et al. 2000; Dolenc 2005; Hobiger et al. 2009; Özalaybey et al. 2011). Studies have also evaluated subsurface S-wave velocity structures via joint inversions of the phase velocity dispersion and H/V ratio

✉ Xueliang Chen
chenxl@cea-igp.ac.cn

[1] Institute of Geophysics, China Earthquake Administration, Beijing, China

[2] General Construction Company of CCTEB Group Co., Ltd., Wuhan, China

curves from microtremor recordings (Arai and Tokimatsu 2005; Picozzi et al. 2005).

The key to estimating subsurface velocity structures by the inversion of MHV curves is the simulation of MHV curves. At present, many methods are used to simulate MHV curves (Lunedei and Malischewsky 2015). MHV curves can be reproduced by the finite-differential method (FDM) in a 3-D soil model (Rhie and Dreger 2009), although Guéguen et al. (2007) found that the fundamental peak frequency (HVfp) of the MHV curve based on the FDM does not conform to the peak frequency (RHVfp) of the fundamental-mode Rayleigh wave H/V spectrum in areas, where irregular subsurface structures occur. Moreover, Uebayashi et al. (2012) indicated that the HVfp in a 3-D velocity model always is higher than the RHVfp. The MHV peak shape broadens in areas with irregular subsurface structures (Uebayashi 2003).

In addition, since Harkrider (1964, 1970) proposed the theory of surface wave propagation in a 1-D layered half-space, the H/V spectra of the fundamental-mode Rayleigh waves have been used to simulate the MHV curves in 1-D velocity structures (Fäh et al. 2003). The peak frequencies of the H/V spectral ratio of fundamental-mode Rayleigh waves are similar to those of the microtremors at a given site (Lachetl and Bard 1994), and Malischewsky and Scherbaum (2004) found that higher similarity corresponded to a higher impedance contrast between the layer and the half-space. However, the amplitudes are dissimilar (Yamanaka et al. 1994), because higher mode Rayleigh waves, Love waves, and body waves are observed in microtremors. The subsurface S-wave velocity structures can be successfully evaluated by the inversion of the dispersion curves of microtremor vertical motions, which indicates that microtremors mainly consist of surface waves (Horike 1985). Moreover, because body waves attenuate more rapidly than surface waves, surface waves can predominate at distances greater than one wavelength of a Rayleigh or Love wave from the source (Tamura 1996). Hence, a more reasonable approach is to simulate MHV curves using surface waves rather than body waves. On the other hand, Malischewsky and Scherbaum (2004) studied the Love formula and H/V ratio (ellipticity) of Rayleigh waves, and found that the dependence of the ellipticity of Rayleigh wave on frequency is very sensitive on the material properties of the propagation medium. Bonnefoy-Claudet et al. (2008) found that the relative proportion of Love and Rayleigh waves in microtremors depends on the site conditions (the subsurface structure and the sources may have effects on the Love wave contribution) and especially on the impedance contrast. Certain sites present a low impedance contrast when Love waves dominate the wavefield at the H/V peak (Endrun 2011).

Studies have also focused on simulating Rayleigh and Love waves (Bonnefoy-Claudet et al. 2008; Endrun 2011). Arai and Tokimatsu (2004) proposed the H/V spectral ratio of the surface wave (SHV) curves of all modes of surface waves. The Rayleigh-to-Love wave amplitude ratio (*R/L*) in the horizontal components of microtremors is important for calculating the SHV. The *R/L* values can be estimated by analyzing the spatial autocorrelation function of the microtremor array records (Köhler et al. 2006, 2007; Bonnefoy-Claudet et al. 2008). The *R/L* value varies widely (from 10 to 90%) with the frequency (Köhler et al. 2007; Endrun 2011). Arai and Tokimatsu (2000) indicated that the *R/L* values occur in a range from 0.4 to 1.0 with changes in the sites and frequencies. However, the *R/L* values are assumed to be constant at all frequencies in the inversion of the microtremor H/V spectrum, because the *R/L* values cannot be evaluated by a single-station three-component seismometer if the subsurface seismic velocity model is unknown (Arai and Tokimatsu 2000).

Moreover, based on the model for the formalization of the full wavefield, Lunedei and Albarello (2010) indicated that the numerical approximation model consisting of the surface waves is appropriate if a source-free area with a radius of suitable length occurs around the receiver. Furthermore, based on the diffuse field theory (García-Jerez et al. 2013), the MHV can be simulated as the ratio of the imaginary parts of Green's function for the horizontal components to the vertical component when the exciting and receiving points coexist on the ground surface (Sánchez-Sesma et al. 2011). In addition, the inversion result of the H/V spectral ratio under the diffuse field assumption is reproduced well (García-Jerez et al. 2016; Piña-Flores et al. 2017).

In this study, the area studied is located in the Yuxi basin in Yuxi City of southwestern China, for which high-quality drilling data have been obtained and a precise subsurface S-wave velocity model has been developed. We simulate the MHV curves via the SHV. Then, we attempt to analyze the features of the SHV curves with changes in the *R/L* values. We also study the sensitivity of the SHV curve to verify the feasibility of using the SHV curves to invert the subsurface S-wave velocity structure.

Methods

Harkrider (1964) proposed the vertical and horizontal powers of the high-mode surface waves based on the surface wave propagation theory in a layered half-space. Then, assuming that the point sources are randomly distributed, the vertical and horizontal powers of all point sources are integrated to obtain the vertical and horizontal powers of multi-order surface waves (Harkrider 1964).

Arai and Tokimatsu (2000) indicated that the equation for SHV and *R/L* can be expressed as

$$\mathrm{SHV}(\omega) = \sqrt{\frac{P_{\mathrm{HS}}(\omega)}{P_{\mathrm{VS}}(\omega)}} = \sqrt{\frac{P_{\mathrm{HR}}(\omega) + P_{\mathrm{HL}}(\omega)}{P_{\mathrm{VR}}(\omega)}} \quad (1)$$

$$(R/L)(\omega) = \sqrt{\frac{P_{\mathrm{HR}}(\omega)}{P_{\mathrm{HL}}(\omega)}} \quad (2)$$

where the subscripts *R*, *L*, and *S* indicate Rayleigh, Love, and surface waves, respectively, and $P_{\mathrm{HR}}(\omega)$ and $P_{\mathrm{VR}}(\omega)$ are the horizontal and vertical powers of all mode Rayleigh wave from all point sources at a frequency ω, respectively. These powers can be expressed as

$$P_{\mathrm{HR}}(\omega) = \kappa L_v^2 \sum_{m=0}^{M} \left(\frac{A_{Rm}}{k_{Rm}}\right)^2 \left(\frac{\dot{u}}{\dot{w}}\right)_m^2 \left[1 + \left(\frac{\alpha^2}{2}\right)\left(\frac{\dot{u}}{\dot{w}}\right)_m^2\right] \quad (3)$$

$$P_{\mathrm{VR}}(\omega) = \kappa L_v^2 \sum_{m=0}^{M} \left(\frac{A_{Rm}}{k_{Rm}}\right)^2 \left[1 + \left(\frac{\alpha^2}{2}\right)\left(\frac{\dot{u}}{\dot{w}}\right)_m^2\right] \quad (4)$$

where *A* is the medium response factor (Harkrider 1964, 1970); m and *M* indicate surface modes; *k* is the wave number; $\dot{u}/\dot{w}$ is the H/V ratio of Rayleigh wave velocity on free surface (Haskell 1953); L_V is the vertical loading force of a microtremor; $\kappa = (2/h)e^{-4\pi h}$, *h* is the scattering damping ratio of soil; α is the H/V ratio of microtremor loading sources, and $\alpha = L_H/L_V$, and $P_{\mathrm{HL}}(\omega)$ is the horizontal power of all mode Love waves from all point sources at a frequency ω. This power can be expressed as

$$P_{\mathrm{HL}}(\omega) = \kappa L_v^2 \sum_{m=0}^{M} \left(\frac{\alpha^2}{2}\right) \left(\frac{A_{Lm}}{k_{Lm}}\right)^2. \quad (5)$$

Simulation of microtremor H/V spectral curves

Soil structure parameters and H/V spectral curves of microtremors

The Yuxi basin shown in Fig. 1 is located in the southern end of the fault zone, and it is a typical sedimentary basin in southwestern China. Yuxi City is located in the south-eastern part of the basin, and site B is closer to the city, while site A is far from the city. The S-wave velocity profiles are shown in Fig. 2 at sites A and B. Moreover, He et al. (2013) estimated the subsurface velocity structures of the Yuxi basin from the results of seismic exploration. The velocity structures below the depth of the borehole can be obtained from this velocity model (Table 1). The bedrock depths of sites A and B are 415 and 189 m, respectively.

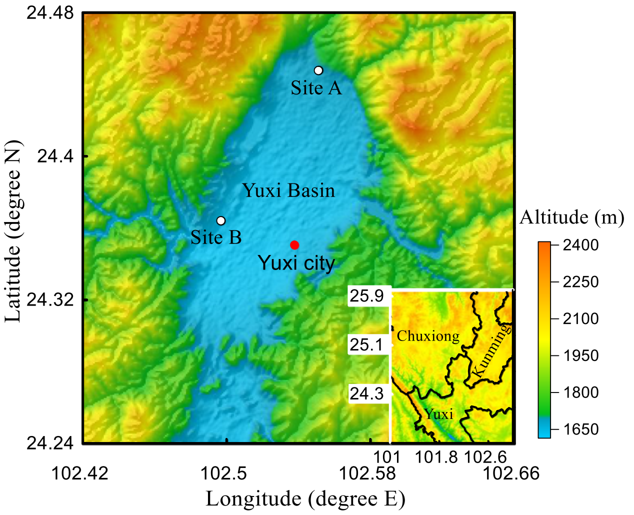


Fig. 1 Location/map of the Yuxi basin. The white dots represent single-station three-component microtremor observation sites. The red dot represents Yuxi City. The color scale bar shows the altitude

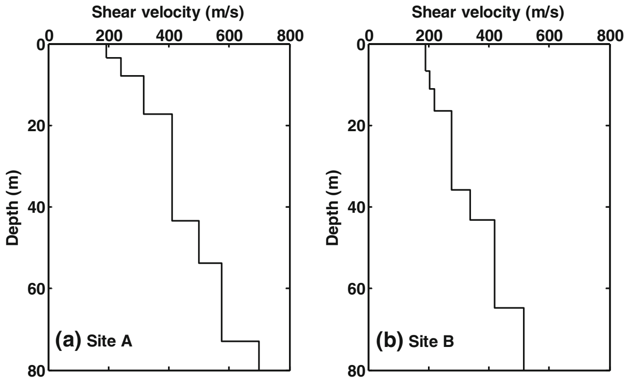


Fig. 2 Shallow subsurface S-wave velocity profile obtained from boreholes at sites A and B

Table 1 Velocity structures below the depth of the borehole based on the results of seismic exploration near site A

Depth (m)	V_S (m/s)	V_P (m/s)	ρ (t/m^3)
80–100	700	1534	1.9
100–185	800	1720	2.1
185–415	1200	2433	2.2
415–900	1700	2766	2.4
> 900	3000	5326	2.5

Microtremor signals are recorded by a single-station three-component seismometer with a sampling frequency of 8 ms at sites A and B. On June 3, 2012 at 14:26, 14:34, and 14:43, three sections of microtremors were recorded at site A. On May 27, 2012 at 18:02, 18:16, and 18:26, three sections of microtremors were recorded at site B. Each section of microtremor is 360.448 s. For each section of microtremor, 16.384 s at the head and the end each is removed, because these records contain interference by human factors. To reduce the measurement error, ten sets of data segments of 32.768 s each are used from each section of the microtremors. Hence, at each site, microtremor records of 983.04 s (30 sets of data segments) are used for the calculation of MHV curves. The MHV at frequency ω is defined as

$$\mathrm{MHV}(\omega) = \sqrt{\frac{P_{\mathrm{EW}}(\omega) + P_{\mathrm{NS}}(\omega)}{P_{\mathrm{UD}}(\omega)}} \tag{6}$$

where P_{EW} and P_{NS} are the Fourier power spectra of the two orthogonal horizontal motions and P_{UD} is the Fourier power spectrum of the vertical motion.

Measured H/V and simulated H/V curves

To calculate the SHV using Eq. 1, α values are required. The α values can be determined from Eq. 2, because the *R/L* values can be evaluated by microtremor array data. The relative proportion of Love waves in the horizontal components of microtremors varies widely (from 60 to 90%) with the frequency (Köhler et al. 2007; Endrun 2011). Arai and Tokimatsu (2000) indicated that the *R/L* values vary between 0.4 and 1.0 and have an average of approximately 0.7, with a period in the range of 0.1–5 s. We calculated the MHV and SHV curves for *R/L* values of 0.4, 0.7, and 1.0, and they are shown in Fig. 3. In the figure, the MHV curves are between the maximum and minimum curves of the SHV. Hence, a reasonable assumption is that the *R/L* values are in the range of 0.4–1.0 at those sites and frequencies.

The microtremor H/V curve can be used to evaluate the site effect and the subsurface velocity structures. The fundamental resonance frequency of the site is the key to evaluating the site effect and the subsurface velocity profile. The frequency window selected for the analysis of microtremors must contain the fundamental resonance frequency of the site (Picozzi et al. 2005; Sánchez-Sesma et al. 2011). The fundamental resonance frequencies of sites A and B are estimated to be 0.55 and 1.5 Hz, respectively. Accordingly, at sites A and B, we chose the H/V spectral curve in a frequency range of 0.1–5 Hz to analyze the features of SHV curves. Because the bedrock depth of site B (189 m) is less than the bedrock depth of site A (402 m), the fundamental resonance frequency of site B is greater than that of site A, as shown in Fig. 3.

The amplitude of the SHV curve decreases as the *R/L* value increases, as shown in Fig. 3. The amplitude deviation between the MHV and SHV curves for *R/L* value of *n* at the frequency ω can be expressed as

$$D_n(\omega) = |\mathrm{MHV}(\omega) - \mathrm{SHV}_{\mathrm{n}}(\omega)| \tag{7}$$

where *n* is the *R/L* value (*n* in a range from 0.4 to 1.0). The amplitude deviation between the SHV for *R/L* values of 0.4 and the SHV for *R/L* values of 1.0 is 4.44 at the peak frequency (0.5 Hz) at site A, and the amplitude deviations decrease gradually from 4.44 as distance from the peak frequency increases. The amplitude deviation between the SHV for *R/L* values of 0.4 and the SHV for *R/L* values of 1.0 is 2.39 at the peak frequency (1.4 Hz) at site B, and the amplitude deviations decrease gradually from 2.39 as the distance from the peak frequency increases. These findings suggest that the SHV curves are most sensitive to *R/L* values at the peak frequencies of the spectrum and the

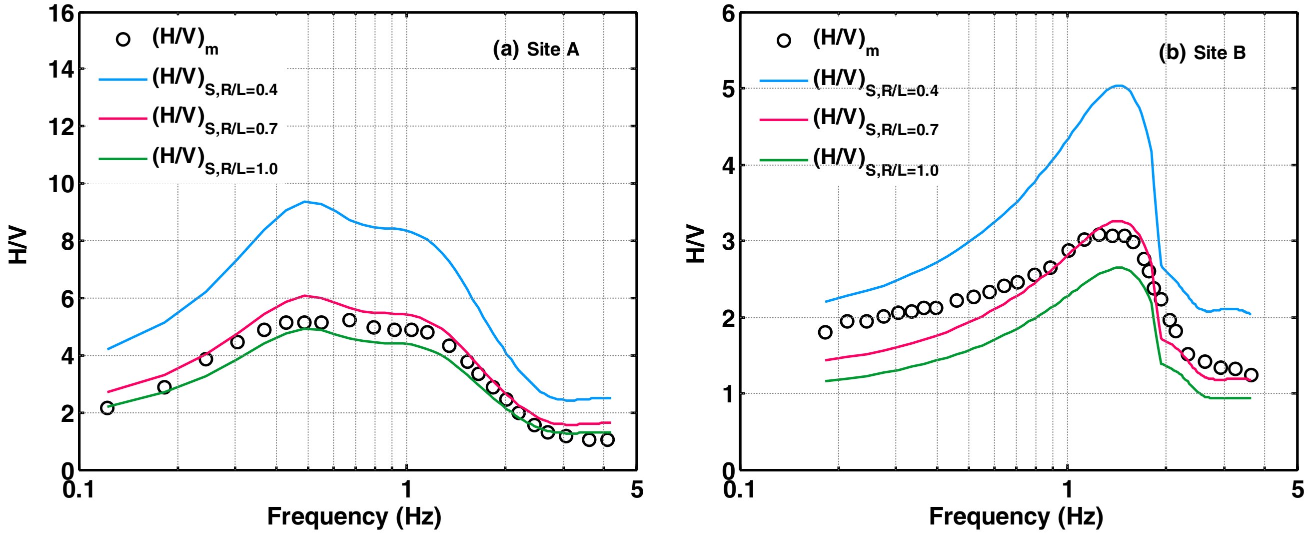


Fig. 3 Comparison of the measured H/V (open circles) with the simulated H/V (solid lines) obtained using a 1-D velocity structure model at sites A and B. The blue line, red line, and green line represent the H/V of surface waves for *R/L* values of 0.4, 0.7, and 1.0, respectively

sensitivity of the SHV curves to the *R/L* values decreases as the distance from the peak frequency increases.

To obtain more general results, we used two different soil models (a normal soil model and a low-velocity soil model) and calculated the SHV obtained using a 1-D velocity structure model (Fig. 4). We can reach the same conclusion from Fig. 4, in which the SHV curves are most sensitive to *R/L* values at the peak frequencies of the spectrum. Therefore, when the MHV curves are simulated by the SHV curves, where the *R/L* value is constant, the amplitude deviation between the MHV and SHV curves at the near-peak frequency is higher, as shown in Fig. 3. The degree of fitting between the SHV and MHV curves directly affects the accuracy of the velocity structures obtained by the inversion of the microtremor H/V spectrum. To reduce the error of the subsurface velocity structures obtained by the MHV curves, we suggest that the SHV curves at the near-peak frequencies should not be considered in the inversion of the microtremor H/V spectrum, because the amplitude deviation between the MHV and SHV curves is higher at the peak frequency of the MHV curves.

The peak frequencies of the SHV curves for *R/L* values of 0.4, 0.7, and 1.0 are similar to those of the MHV curves, although the amplitudes are not exactly consistent. Therefore, the peak frequencies of the SHV curves are independent of the *R/L* values and are not affected by the α value in the area studied. The amplitude of SHV for the constant *R/L* value is not consistent with that of the MHV curves, because the *R/L* value varies widely based on the site conditions and frequencies.

Figure 5 shows the amplitude deviations between the MHV and SHV curves for different *R/L* values and frequencies at sites A and B obtained using Eq. 7. The range of best *R/L* values at site B is larger than that at site A, as shown in Fig. 5. The fluctuations of the best *R/L* values in the high-frequency part (> 1 Hz) of the MHV curve are significantly higher than those in the low-frequency part (< 1 Hz) at site B, which is likely because the high-frequency part (> 1 Hz) of microtremor signal is primarily influenced by artificial noise (Seht and Wohlenberg 1999). These microtremor signals are recorded during the day, and site B is located close to the city. Human activities cause the points to fluctuate in a larger range at site B. As a result, the fluctuation range of the best *R/L* values is significantly larger in the high-frequency range.

When the *R/L* values are estimated by analyzing the spatial autocorrelation function of the microtremor array records, the *D* values are less than 0.2 (Arai and Tokimatsu 2000). Moreover, the inversion results of the MHV curves can ensure that the *D* values are less than 0.2 (Parolai et al. 2005; Picozzi et al. 2005). We define the best *R/L* value as the value that generates amplitude deviations between the SHV and MHV curves (*D* values) that are less than 0.2. We use the best *R/V* values obtained from Fig. 5 to simulate the MHV curves. When the *R/L* value is the best, the SHV curves simulate MHV curves very well, as shown in Fig. 6. We approximate the best *R/L* values as the true *R/L* values.

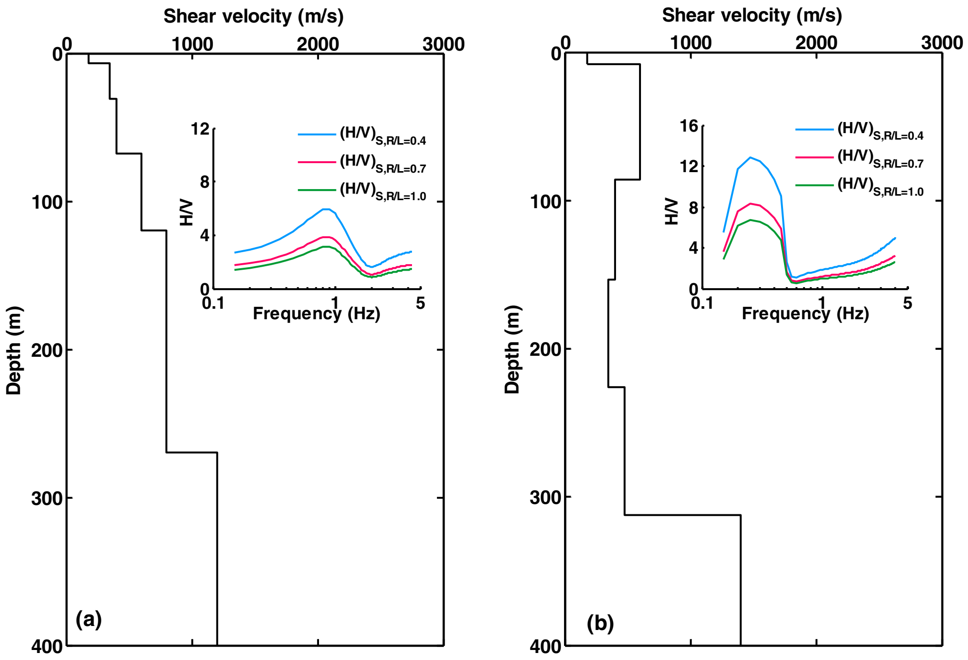


Fig. 4 S-wave velocity profile and H/V spectra of surface waves (solid lines) obtained using a 1-D velocity structure model with two different hypothetical models. The blue line, red line, and green line represent the H/V of surface waves when the *R/L* values are 0.4, 0.7, and 1.0, respectively

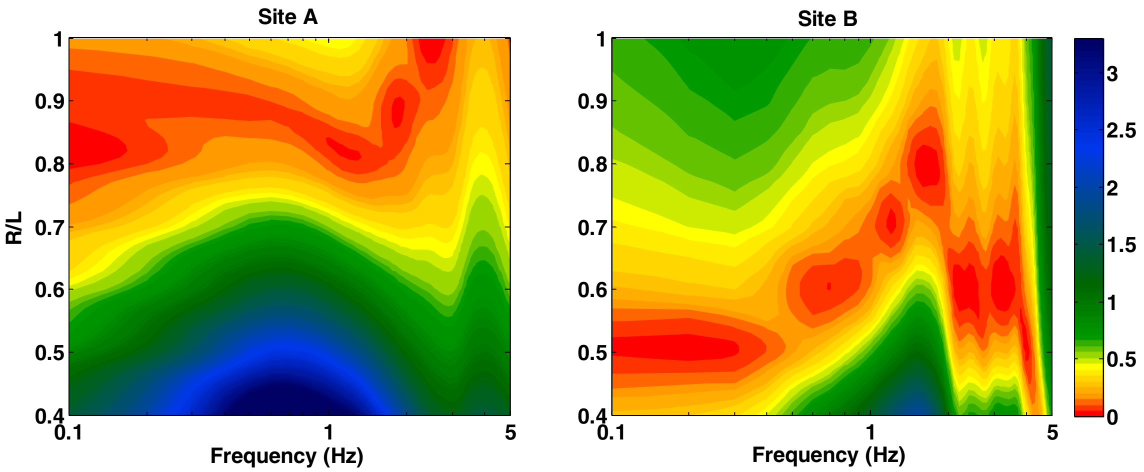


Fig. 5 Amplitude deviations between the MHV and SHV curves for different *R/L* values and frequencies at sites A and B obtained using Eq. 7. The color scale bar shows the amplitude deviations between the MHV and SHV curves

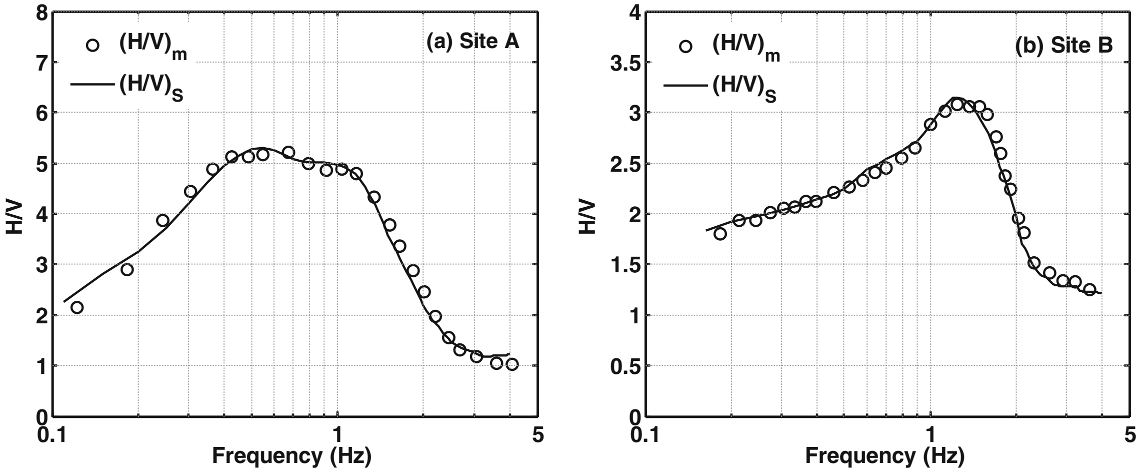


Fig. 6 Comparison of the measured H/V (open circles) with the simulated H/V (solid lines) of the best *R/L* values at sites A and B

Sensitivity analysis of the surface wave H/V spectrum

To verify the feasibility of using MHV curves to estimate subsurface velocity structures, the sensitivities of the SHV curves must be analyzed. The sensitivities of the SHV curves (Arai and Tokimatsu 2000), which are the absolute values of the dimensionless partial derivative of SHV for any parameter P of the soil models, can be expressed as

$$S_{\mathrm{ji}}^{p}(\omega) = \left| \frac{P}{\mathrm{SHV}(\omega)} \frac{\partial \mathrm{SHV}(\omega)}{\partial P} \right|_{P=P_{ji}} \qquad (8)$$

where S_{ji}^{p} is the sensitivity of the SHV and P_{ji} may represent the thickness H_j, the P-wave velocity V_{Pj}, the S-wave velocity V_{Sj}, and the density ρ_j in the *j*th layer of the soil model.

The sensitivities (*S*) of the SHV curves for the soil model parameters at site A that were obtained using Eq. 8 are shown as Fig. 7. The S-wave velocity and thickness of the soil layer have a great influence on the SHV curve, whereas the density and P-wave velocity of the soil layer have a lesser influence on the SHV curve. Therefore, the inversion of MHV curves can be best achieved by adjusting either the S-wave velocity, thickness or both these parameters at each soil layer.

The maximum *S* values ($S_{\max}$) of the SHV curves for the soil layer parameters in a frequency range of 0.1–5 Hz are shown in Fig. 8. At site A, the $S_{\max}$ values are mainly concentrated in a frequency range of 1.1–2.4 Hz, which is between the peak and trough frequencies of the H/V spectrum of microtremors recorded at site A. At site B, the $S_{\max}$ values are concentrated in a frequency range of 1.4–2.4 Hz, which is between the peak and trough frequencies of the H/V spectrum of microtremors recorded at site B. In addition, Fig. 7 shows that the *S* values at frequencies less than 1 Hz are far less than those at frequencies greater than

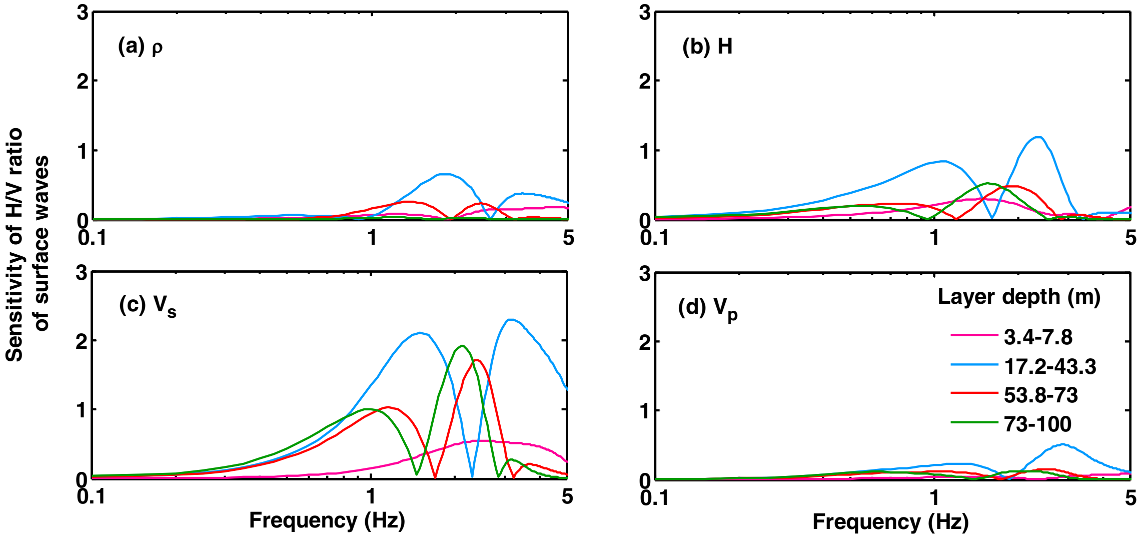


Fig. 7 Sensitivity value S of the surface wave H/V spectrum for the soil model parameters at site A obtained using Eq. 8. **a–d** *S* Values for the density, thickness, and S- and P-wave velocity of the soil layers

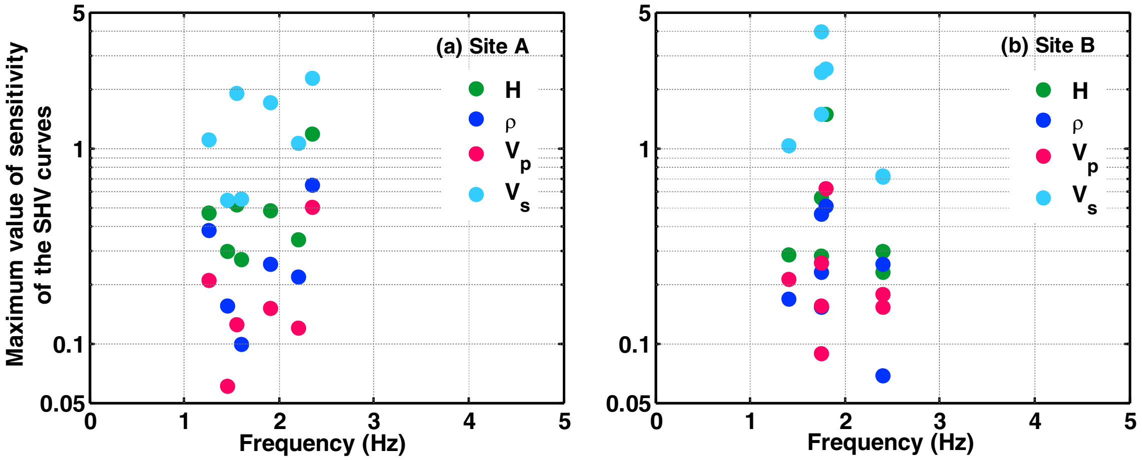


Fig. 8 Maximum sensitivity value of the SHV curves for shallow soil model parameters

1 Hz at site A. The same trends can be obtained from the soil model at site B. Therefore, the spectrum curves are sensitive to the parameters in a frequency range that is located between the peak and trough frequencies of the microtremor H/V spectrum. Therefore, this frequency range is the most suitable for inversion.

Discussion

We simulated MHV curves using the surface wave propagation theory in a layered half-space. In Fig. 3, we compare the MHV and SHV for *R/L* values of 0.4, 0.7 and 1. The misfit is clearly different for both *R/L* curves compared to the measured H/V. In this paper, we find the best *R/L* values by simulating MHV curves for different values and by evaluating the misfit with respect to the measured MHV. Figure 5 shows that the best *R/L* values are stable in the range of 0.8–1.0 with changes in frequency at site A, and the best *R/L* values are stable in the range of 0.4–0.9 with changes in frequency at site B. Hence, the *R/L* values vary widely based on the site condition and frequency. Bonnefoy-Claudet et al. (2008) and Endrun (2011) calculated *R/L* values via a frequency-dependent analysis of the main propagation and polarization based on microtremor array measurements and found that the relative proportion of Love and Rayleigh waves in the microtremors varies widely with frequency and site conditions and the *R/L* values are stable in the ranges of 0.2–1.0. At sites A and B, the best *R/L* values we get are also in this range. However, the *R/L* value cannot be evaluated from microtremor data recorded by a single-station three-component seismometer if the velocity model is unknown. Further study is required to determine a method of identifying the true *R/L* values to estimate the subsurface velocity structure using the H/V spectrum of microtremors recorded by a single-station three-component seismometer.

Figures 3 and 4 show that the SHV curves are the most sensitive to *R/L* values at the peak frequencies of the spectrum. To reduce the error of subsurface velocity structures obtained from the MHV curves, the SHV curves at the near-peak frequencies should not be considered in the inversion of the microtremor H/V spectrum. Because the location of the peak frequency depends on the subsurface structure, it varies naturally from low to high frequencies for different sites. Therefore, we cannot provide a specific frequency range that contains the peak frequency that is not used to estimate the velocity structures.

Finally, we analyze the sensitivity of the SHV curves. Figure 7 confirms that the S-wave velocity and the thickness of the soil layer has a great influence on the SHV curve. By calculating the partial derivatives of the H/V spectrum of Rayleigh waves, Tsuboi and Saito (1983) indicated that the H/V curves are most sensitive to the S-wave velocity. Furthermore, the MHV curves are sensitive to the parameter in a frequency range that is located between the peak and trough frequencies of the SHV, as shown in Fig. 7. In the process of applying the mode summation method and a finite difference technique to investigate the spectral ratio, Fäh et al. (2001) similarly found that the MHV curves are controlled by the ellipticity of the fundamental-mode Rayleigh wave in the frequency band between the fundamental frequency of resonance of the sediments and the first minimum of the average H/V ratio. Although the peak frequencies of the H/V spectral ratio of fundamental-mode Rayleigh waves are similar to that of the microtremors in general, there are some sites with a low impedance contrast, where the Love waves dominate the wavefield at the H/V peak. Hence, a more reasonable approach is to express the sensitive frequency range of MHV using surface waves rather than Rayleigh waves.

Conclusions

In this study, we used the SHV equation based on the surface wave propagation theory in a layered half-space to simulate the MHV curves of sites A and B in the Yuxi basin. We found that the peak frequency of the SHV curve is independent of the α value, although the amplitude of the SHV curve has some relation with the α value. Furthermore, we found that the SHV curves are most sensitive to *R/L* values at the peak frequencies of the spectrum. Therefore, to reduce the error of subsurface velocity structures obtained by the MHV curves, we suggest that the SHV curves at the near-peak frequencies should not be considered in the inversion of the microtremor H/V spectrum, because the amplitude deviation between the MHV and SHV curves is higher at the peak frequency of the MHV curves.

In addition, to verify the feasibility of the inversion of MHV curves, we analyzed the features of the sensitivity of SHV curves. We confirmed that the inversion of MHV curves can be best achieved by adjusting either the S-wave velocity, thickness, or both these parameters at each soil layer. Moreover, the best frequency range for the inversion of the microtremor H/V spectrum is between the peak and trough frequencies of the microtremor H/V spectrum.

Acknowledgements This research work was supported by the National Natural Science Foundation of China (Nos. 51278470, 51678537) and the Key Laboratory of Seismic Observation and Geophysical Imaging Project. The microtremor records were obtained from the China Seismic Array at http://www.chinaarray.org. The authors thank the anonymous referees for their comments, which contributed to improving the work.

Compliance with ethical standards

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

References

Aki K (1957) Space and time spectra of stationary stochastic waves, with special reference to microtremors. Bull Earthq Res Inst Tokyo Univ 35:415–456

Apostolidis P, Raptakis D, Roumelioti Z, Pitilakis K (2004) Determination of S-wave velocity structure using microtremors and SPAC method applied in Thessaloniki (Greece). Soil Dyn Earthq Eng 24(1):49–67. https://doi.org/10.1016/j.soildyn.2003.09.001

Arai H, Tokimatsu K (2000) Effects of Rayleigh and Love waves on microtremor H/V spectra. In: the 12th world conference on earthquake engineering, Auckland, New Zealand, 30 January–4 February 2000

Arai H, Tokimatsu K (2004) S-wave velocity profiling by inversion of microtremor H/V spectrum. Bull Seismol Soc Am 94(1):53–63

Arai H, Tokimatsu K (2005) S-wave velocity profiling by joint inversion of microtremor dispersion curve and horizontal-to-vertical (H/V) spectrum. Bull Seismol Soc Am 95(5):1766–1778. https://doi.org/10.1785/0120040243

Bonnefoy-Claudet S, Cornou C, Bard PY, Cotton F, Moczo P, Kristek J, Fäh D (2006) H/V ratio: a tool for site effects evaluation. Results from 1-D noise simulations. Geophys J Int 167(2):827–837

Bonnefoy-Claudet S, Köhler A, Cornou C, Wathelet M, Bard PY (2008) Effects of Love waves on microtremor H/V ratio. Bull Seismol Soc Am 98(1):288–300

Capon J (1969) High-resolution frequency-wavenumber spectrum analysis. Proc IEEE 57(8):1408–1418

Carcione JM, Picotti S, Francese R, Giorgi M, Pettenati F (2017) Effect of soil and bedrock anelasticity on the S-wave amplification function. Geophys J Int 208:424–431

Chen XL, Gao MT, Yu YX, Chen K, Li TF (2014) Applicability of topographic slope method for seismic site soil classification of Yuxi-Jiangchuan-Tonghai-Basin in China. Earthq Eng Eng Dyn S1:146–152 **(in Chinese)**

Claprood M, Asten MW, Kristek J (2011) Using the SPAC microtremor method to identify 2D effects and evaluate 1D shear-wave velocity profile in valleys. Bull Seismol Soc Am 101 (101):826–847. https://doi.org/10.1785/0120090232

Delgado J, Casado CL, Giner J, Estevez A, Cuenca A, Molina S (2000) Microtremors as a geophysical exploration tool: applications and limitations. Pure Appl Geophys 157(9):1445–1462

Dolenc D (2005) Microseisms observations in the Santa Clara Valley, California. Bull Seismol Soc Am 95(3):1137–1149

Endrun B (2011) Love wave contribution to the ambient vibration H/V amplitude peak observed with array measurements. J Seismol 15(3):443–472

Fäh D, Kind F, Giardini D (2001) A theoretical investigation of average H/V ratios. Geophys J Int 145(2):535–549

Fäh D, Kind F, Giardini D (2003) Inversion of local s-wave velocity structures from average H/V ratios, and their use for the estimation of site-effects. J Seismol 7(4):449–467. https://doi.org/10.1023/B:JOSE.0000005712.86058.42

Gao MT, Yu YX, Zhang XM, Wu J, Hu P, Ding YH (2002) Three-dimensional finite-difference simulations of ground motions in the Beijing area. Earthq Res China 18(4):356–364 **(in Chinese)**

García-Jerez A, Luzón F, Sánchez-Sesma FJ, Lunedei E, Albarello D, Santoyo MA, Almendros J (2013) Diffuse elastic wavefield within a simple crustal model. Some consequences for low and high frequencies. J Geophys Res-Solid Earth 118(10):5577–5595

García-Jerez A, Piña-Flores J, Sánchez-Sesma FJ, Luzón F, Perton M (2016) A computer code for forward calculation and inversion of the H/V spectral ratio under the diffuse field assumption. Comput Geosci 97:67–78

Guéguen P, Cornou C, Garambois S, Banton J (2007) On the limitation of the h/v spectral ratio using seismic noise as an exploration tool: application to the grenoble valley (france), a small apex ratio basin. Pure Appl Geophys 164(1):115–134. https://doi.org/10.1007/s00024-006-0151-x

Harkrider DG (1964) Surface waves in multilayered elastic media, part 1. Bull Seismol Soc Am 54(2):627–679

Harkrider DG (1970) Surface waves in multilayered elastic media, part 2. Bull Seismol Soc Am 60(6):1937–1987

Haskell NA (1953) The dispersion of surface waves on multilayered media. Bull Seismol Soc Am 43(1):17–34

He ZQ, An HS, Shen K, Lu LY, Hu G, Ye TL (2013) Detection of Puduhe fault in Yuxi basin of Yunnan by seismic reflection method. Acta Seismol Sin 35(6):836–847 **(in Chinese)**

Hobiger M, Bard PY, Cornou C, Bihan NL (2009) Single station determination of Rayleigh wave ellipticity by using the random decrement technique (RayDec). Geophys Res Lett 36(14): L14303

Horike M (1985) Inversion of phase velocity of long-period microtremors to the S-wave-velocity structure down to the basement in urbanized areas. J Phys Earth 33(2):59–96

Köhler A, Ohrnberger, M., Scherbaum, F (2006) The relative fraction of Rayleigh and Love waves in ambient vibration wavefields at different European sites. In: Third international symposium on the effects of surface geology on seismic motion Grenoble, France, 30 August–1 September 2006

Köhler A, Ohrnberger M, Scherbaum F, Wathelet M, Cornou C (2007) Assessing the reliability of the modified three-component spatial autocorrelation technique. Geophys J Int 168(2):779–796

Komatitsch D, Liu Q, Tromp J, Süss P, Stidham C, Shaw JH (2004) Simulations of ground motion in the Los Angeles basin based upon the spectral-element method. Bull Seismol Soc Am 94 (1):187–206

Lachetl C, Bard PY (1994) Numerical and theoretical investigations on the possibilities and limitations of Nakamura's technique. Earth Planets Space 42(5):377–397. https://doi.org/10.4294/jpe1952.42.377

Lunedei E, Albarello D (2010) Theoretical HVSR curves from full wavefield modelling of ambient vibrations in a weakly dissipative layered Earth. Geophys J Int 181(2):1342–1342

Lunedei E, Malischewsky P (2015) A review and some new issues on the theory of the H/V technique for ambient vibrations. Perspectives on european earthquake engineering and seismology. Springer, New York, pp 371–394

Malischewsky PG, Scherbaum F (2004) Love's formula and H/V-ratio (ellipticity) of Rayleigh waves. Wave Motion 40:57–67

Nogoshi M, Igarashi T (1971) On the amplitude characteristics of microtremor (part 2). J Seismol Soc Jpn 24(1):26–40

Özalaybey S, Zor E, Ergintav S, Tapırdamaz MC (2011) Investigation of 3-d basin structures in the İzmit bay area (turkey) by single-station microtremor and gravimetric methods. Geophys J Int 186 (2):883–894

Parolai S, Picozzi M, Richwalski SM, Milkereit C (2005) Joint inversion of phase velocity dispersion and H/V ratio curves from seismic noise recordings using a genetic algorithm, considering higher modes. Geophys Res Lett 32(1):67–100. https://doi.org/10.1029/2004GL021115

Picozzi M, Parolai S, Richwalski SM (2005) Joint inversion of H/V ratios and dispersion curves from seismic noise: estimating the S-wave velocity of bedrock. Geophys Res Lett 32(11):339–357. https://doi.org/10.1029/2005GL022878

Piña-Flores J, Perton M, Garcíajerez A, Carmona E, Luzon F, Molina-Villegas JC, S´anchez-Sesma FJ (2017) The inversion of spectral ratio H/V in a layered system using the diffuse field assumption. Geophys J Int 208:577–588

Rhie J, Dreger D (2009) A simple method for simulating microseism H/V spectral ratio in 3D structure. Geosci J 13(4):401–406. https://doi.org/10.1007/s12303-009-0036-y

Rong MS, Wang ZM, Woolery EW, Lyu YJ, Li XJ, Li SY (2016) Nonlinear site response from the strong ground-motion recordings in western China. Soil Dyn Earthq Eng 82(4):99–110

Sánchez-Sesma FJ, Rodríguez M, Iturrarán-Viveros U, Luzón F, Campillo M, Margerin L, García-Jerez A, Suarez M, Santoyo MA, Rodríguez-Castellanos A (2011) A theory for microtremor H/V spectral ratio: application for a layered medium. Geophys J Int 186(1):221–225. https://doi.org/10.1111/j.1365-246X.2011.05064.x

Satoh T, Kawase H, Iwata T, Higashi S, Sato T, Irikura K, Huang HC (2001) S-wave velocity structure of the Taichung basin, Taiwan, estimated from array and single-station records of microtremors. Bull Seismol Soc Am 91(5):1267–1282

Seht IV, Wohlenberg J (1999) Microtremor measurements used to map thickness of soft sediments. Bull Seismol Soc Am 89 (1):250–259

Shani-Kadmiel S, Tsesarsky M, Louie JN, Gvirtzman Z (2012) Simulation of seismic-wave propagation through geometrically complex basins: the dead sea basin. Bull Seismol Soc Am 102 (4):1729–1739

Tamura S (1996) Comparison of body and Rayleigh wave displacements generated by a vertical point force on a layered elastic medium. In the 11th World conference on earthquake engineering, Acapulco, Mexico, 23–28 June 1996

Tsuboi S, Saito M (1983) Partial derivatives of Rayleigh wave particle motion. Earth Planets Space 31(2):103–113

Uebayashi H (2003) Extrapolation of irregular subsurface structures using the horizontal-to-vertical spectral ratio of long-period microtremors. Bull Seismol Soc Am 93(2):570–582

Uebayashi H, Kawabe H, Kamae K (2012) Reproduction of microseism H/V spectral features using a three-dimensional complex topographical model of the sediment-bedrock, interface in the Osaka sedimentary basin. Geophys J Int 189(2):1060–1074. https://doi.org/10.1111/j.1365-246X.2012.05408.x

Wu CF, Huang HC (2012) Estimation of shallow S -wave velocity structure in the Puli basin, Taiwan, using array measurements of microtremors. Earth Planets Space 64(5):389–403. https://doi.org/10.5047/eps.2011.12.002

Yamanaka H, Takemura M, Ishida H, Niwa M (1994) Characteristics of long-period microtremors and their applicability in exploration of deep sedimentary layers. Bull Seismol Soc Am 84(6):1831–1841

Acta Geophysica (2018) 66:131–139
https://doi.org/10.1007/s11600-018-0125-2

RESEARCH ARTICLE - SOLID EARTH SCIENCES

Electric resistivity distribution in the Earth's crust and upper mantle for the southern East European Platform and Crimea from area-wide 2D models

Igor M. Logvinov[1] · **Viktor N. Tarasov**[1]

Received: 8 September 2017 / Accepted: 2 March 2018 / Published online: 16 March 2018

Abstract

Previously obtained magnetotelluric 2D models for 30 profiles made it possible to create an overview model of electric resistivity for the territory between 28°E and 36°E and between 44.5°N and 52.5°N. It allows us to distinguish a number of low resistivity objects (LRO) with resistivities lower than 100 Ω m the Earth's crust and mantle. Two regional conductivity anomalies are traced. The Kirovograd conductivity anomaly extends south to the Crimea mountains. A new regional conductivity anomaly (Konkskaya) can be distinguished along the southern slope of the Ukrainian Shield from 29° to 34°E. In addition, many local LROs have been identified. According to the modeling results, the local low resistivity objects on the East European Platform appear along fault zones activated during last 5–7 M years and the model suggests their relation to known zones of graphitization and polymetallic ore deposits. Local LROs in the Dnieper–Donets Basin correlate with the main oil and natural gas fields in this area. The depth of the anomalous objects amounts to 5–22 km. This is consistent with the hypotheses that hydrocarbon deposits are related to generation and transport zones of carbon-bearing fluids.

Keywords Conductivity · Earth's crust and upper mantle · East European Platform · Scythian plate · Crimea · Mineral resources

Introduction

By means of one-dimensional (1D) and two-dimensional (2D) inversions of magnetotelluric (MT) data, resistivity models for certain areas of central Ukraine have been obtained in a number of earlier studies (Gordienko et al. 2005, 2006; Kirovograd 2013; Varentsov et al. 2013; Logvinov 2015; Logvinov et al. 2016). The quasi-3D model based on the unimodal thin-sheet approximation (Kovacikova et al. 2016) made it possible to analyses of conductivity distribution in the upper crust. An important result from these studies is the discovery of the Kirovograd anomaly of electrical conductivity (KirAC) in the central part of the Ukrainian Shield (Rokityansky et al. 1969). Its main anomalous object in the resistivity model where ρ is less than 40 Ω m occurs at 10–40 km depth. Studies have shown that KirAC apparently includes two parallel objects of low resistivity (LROs) with $\rho < 100$ Ω m that extend from both sides of the Ingulo-Krivorozhsko-Krupetskaia zone (Figs. 1B, 3, 4).

In this paper, we want to give an overview of the spacial distribution of LROs which has been modeled in the mentioned studies. For this purpose, an analysis of 2D inversion results has been performed along a set of profiles (Fig. 1B).

Tectonically, the study area belongs to the southern East European Platform (EEP) including the slope of the Voronezh Massif, Dnieper–Donetsk Depression (DDD), Ukrainian Shield (USh), Southern Ukrainian Monocline, part of Scythian Plate and Mountainous Crimea (Fig. 1).

✉ Igor M. Logvinov
anna_log@ukr.net

[1] S.I. Subbotin Name Institute of Geophysics, National Academy of Sciences of Ukraine, Kyiv, Academician Palladin, av. 32, Kiev 03680, Ukraine

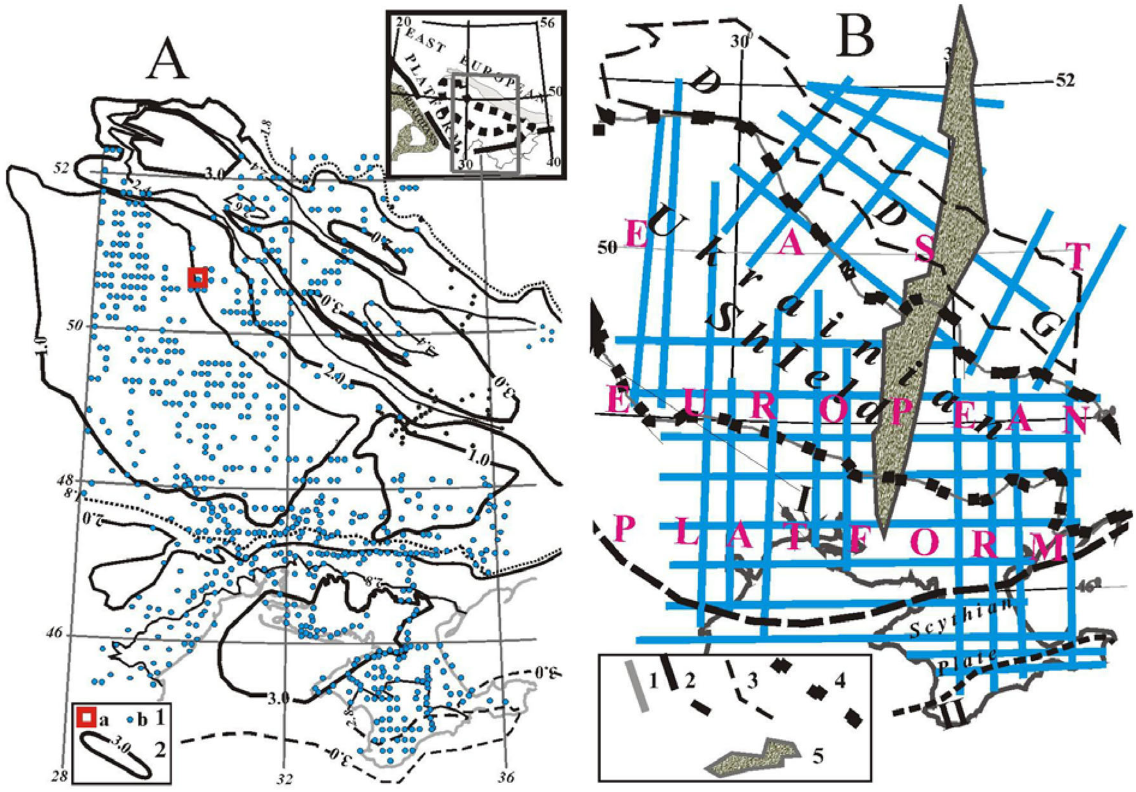


Fig. 1 **A** Observation sites (1): a—KIV observatory, b—of MT measuring. 2—values of the total longitudinal conductivity of sediments (from Logvinov 2015; Logvinov et al. 2016). **B** Profile location for 2D inversion on the tectonic map (from Gintov 2005; Galetskij and Shevchenko 2006; National Atlas of Ukraine 2007; Kirovograd 2013). 1—2D inversion profiles; boundaries: 2—EEP; 3—Dnieper-Donetsk Graben (DDG); 4—USh; 5—Ingulo-Krivorozhsko-Krupetskaia zone (IKKZ). I—Southern Ukrainian monocline, II—Mountainous Crimea

Materials and methods

The initial data are MT curves (MTS) and vertical magnetic transfer functions (VMTF) obtained by industrial organizations and the Institute of Geophysics National Academy of Sciences of Ukraine (IGF). These data are discussed in detail elsewhere (Ingerov et al. 1999; Gordienko et al. 2005, 2006; Varentsov et al. 2013; Logvinov et al. 2016). Most of industrial MT curves were obtained by digital first-generation apparatus. They were aligned to measuring lines and registered in the period range from 1–16 to 900 s.

There are a total of 800 observation sites in the study area. For 450 of them were obtained VMTF values. The density of observation points on the study area is irregular (Fig. 1A). This is mostly due to powerful industrial interferences caused by electric railways, pipelines and large industry agglomerations. The choice of the direction of the modeling profiles in the study of such a large area was based on taking into account regional surface in homogeneities. The greatest conductivity of sedimentary deposits is noted for the territory of the Dnieper–Donets Basin and young sedimentary structures along the boundary of the Eastern European Platform. Therefore, along the network of profiles perpendicular to the strike of these structures, the TE mode was matched by data along the strike of the structures. At the same time, previous studies have been revealed the Kirovograd anomaly of conductivity (KirAC), which has a meridional strike, and its underlying conditions lie at depths of more than 10 km. KirAC spatially agrees with the propagation of rocks of the Early Proterozoic age. In the study area, similar meridional zones of the Early Proterozoic age are identified along 30°E and 36°E. Therefore, to detect anomalously conducting meridional strike objects, the inversion was carried out along the latitudinal profiles and the TE mode corresponded to the latitudinal component of the MTS curves.

Inversion was performed along 30 profiles whose length depended on the amount of observation sites available for each profile. Data at 10–42 sites and from 8 to 13 periods in the period range between 1 and 6400 s were selected from different profiles. The total number of sites was 685. When choosing an interpretation profile, the following requirements were taken into account: distance between the points is not more than 20 km; the total number of interpretation parameters for the selected period range is not less than 50%.

The analysis of experimental data at all the profiles shown in Fig. 1B was made using the REBOCC 2D inversion code (Siripunvaraporn and Egbert 2000). In this study, the misfit between modeled and experimental data was considered satisfactory at the levels of 0.1 for magnetic transfer function, 10° for the impedance phase, and 30% for the apparent resistivity. In the result of processing the variations of the MT field, errors in the determination of each parameter are given. In the 2D simulation program, a triple value of this error is entered in the "ERROR_RESPONSE" section. Initial models are based on geoelectric parameters of sediments from geologic (National Atlas of Ukraine 2007) and geoelectric information (Fig. 2) (Logvinov 2015; Logvinov et al. 2016) available for the separate profiles and each region. The "normal" section was based on 1D inversion results obtained at KIV observatory (Semenov et al. 2008). The northern segments of the meridional profiles were analyzed using geoelectric parameters obtained in the DDD. The geoelectric parameters from the Southern Ukrainian Monocline were taken for the southern portions of the meridional profiles.

Inverted experimental data included impedance estimates for TM and TE modes and VMTF values (TP). First calculations were carried out separately for TE, TP. At the final inversion stage, the model with the smallest RMS for these separate mode solutions was chosen as the start model for the joint TE + TM + TP inversion.

Results and discussion

The modeling resulted in geoelectric cross-sections with a depth range from 1–2 to 75 km. Deeper sections can be determined only for a limited area because the profiles with transfer functions at periods longer than 2500 s do not cover to the whole study region. In modeling, the horizontal grid size changed from 3 to 18 km depending on the distance between observation sites on the profile while the vertical one was constant for all profiles (1–4 km above 48 km depth, 8–15 km from 48 to 90 km depth).

Based on the resistivity cross-sections along the profiles, a three-dimensional matrix was created that included the coordinates of each site on the profile and the ρ value for it. The step along the surface for the matrix was chosen to be the same for the whole area equal to 6 km. Earlier 2D forward modeling for the DDD revealed that rocks with resistivity less than 5 Ω m are associated with the Mesozoic–Cenozoic deposits whose thickness is not greater than 3–4 km (Fig. 2) and the LROs occur below these deposits (Logvinov 2015). Therefore, Fig. 3 shows slices (starting with 5 km depth) for some of the most characteristic depths in the crust and upper mantle.

Generally, the resistivity of crystalline crust and upper mantle rocks on the EEP is not less than 1000 Ω m. In this range, there are inclusions of more conductive rocks, namely rocks of increased porosity with mineralized water, rocks enriched in ore minerals and graphite, fluids and partially molten rocks. In this study, areas with ρ values

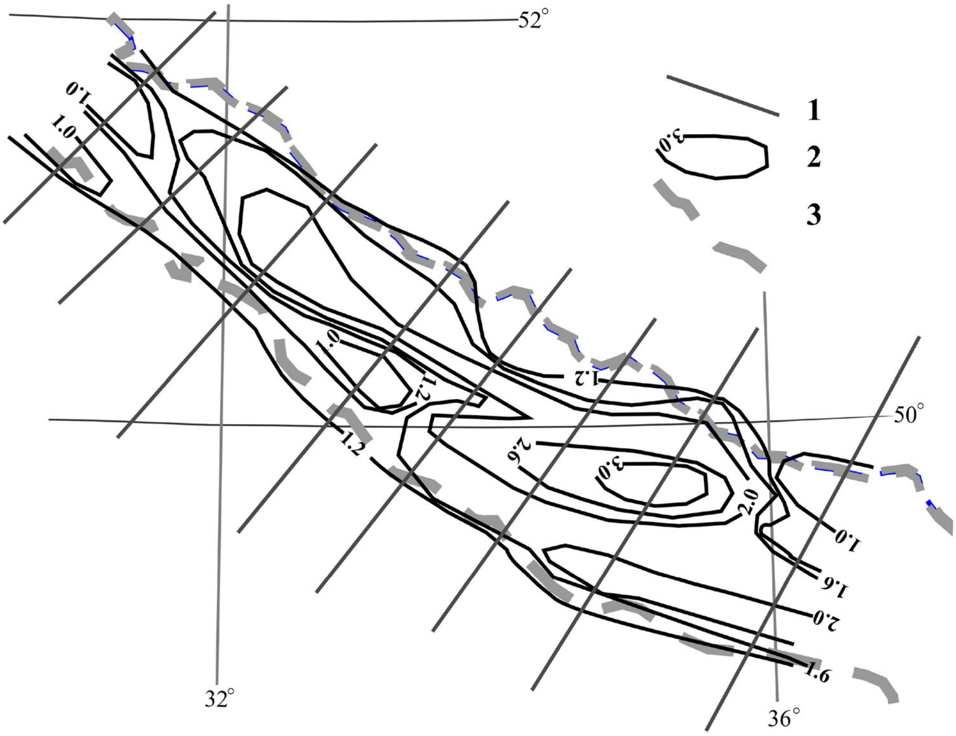


Fig. 2 Schematic presentation isolines of sedimentary thickness (km) with $\rho < 5$ Ω m (1) in the upper sedimentary cover of the DDG (Logvinov 2002); interpretation profiles—2; DDG border—3

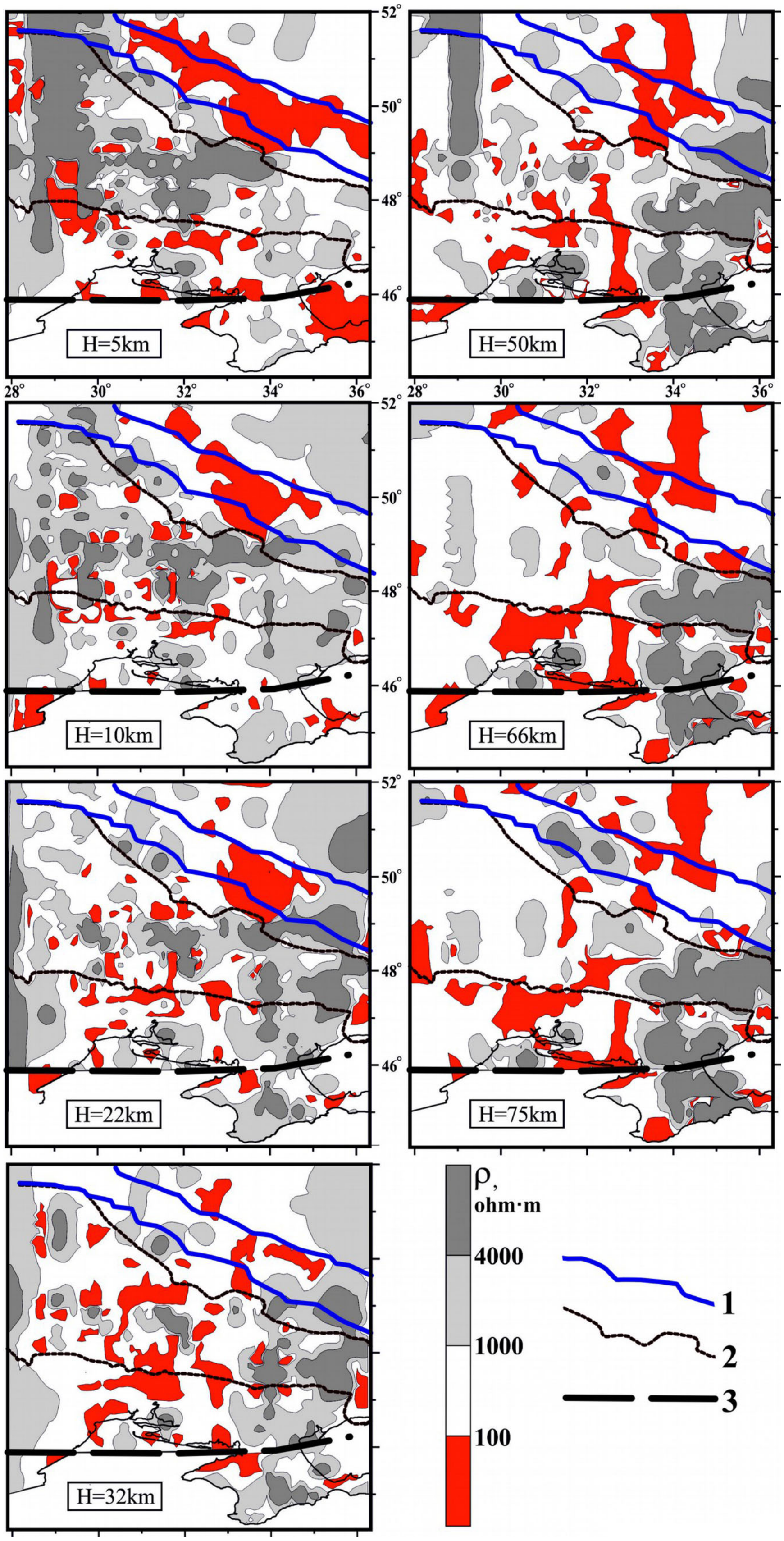


Fig. 3 Contours of electric resistivity derived from 2D models for different depths H. Tectonic boundaries: 1—DDG; 2—USh; 3—EEP

less than 100 Ω m are referred to as LROs. The MT methods are known to poorly discriminate rocks of high resistivity; consequently, the change in high ρ values is shown on the slices only roughly (100–4000 Ω m and more than 4000 Ω m).

The position of the LROs points out their local character because none of the LROs forms a continuous zone. To estimate the integral conductivity of local objects, the parameter *G* (Rokityansky 1982) was used:

$$\mathrm{Gi} = \mathrm{Qi}/\rho\mathrm{i},$$

where Qi is the cross-section area of the conductive object (in m^2), and ρi is its resistivity (in Ω m). In case of complex multi-segment structures, the overall $G = \sum \mathrm{Gi}$. The number of LROs with high integrated conductance *G* (most ρ values < 40 Ω m) allows to distinguish two depth intervals of their preferred occurrence. The first interval has its base at a depth of about 7 km, the second one ranges from about 30 to 75 km.

In the first interval, LROs with a high G are concentrated within the DDD, along the southern flank of the USh from 29° to 33°E and along the boundary of the EEP. LROs, which can be attributed to Mesozoic–Paleozoic sedimentary rocks and Riphean black shale, are consistent with the thickness of these deposits (Fig. 6—sedimentary thickness map) occupy the same area to a depth of 16–24 km. The objects of meridional strike appear at great depths. The LROs associated with the EEP boundary can be traced up to 75 km depth, with a decreasing in dimension and G at depths greater than 7 km. In the depth interval of 5–22 km, the LROs along the southern flank of the USh seem to represent a continuous system that can be only confirmed by a denser network of modeling profiles. At depths from 7 to 30 km, the decrease is clearly seen in the G of the LROs over the entire area, although their number and spatial distribution remain the same.

The second interval occurs in the lower crust deeper than 30 km. The area of the LROs with high conductivity increases noticeably in the central USh from 29° to 33°E and from 47° to 50°N. According to previous studies, the KirAC has a submeridional strike and is located between 32° and 34°E. At depths shallower than 20 km, the LROs of this area are unlikely to be related to the KirAC. Starting with a depth of about 20 km between 44.5° and 52°N, a chain of LROs is clearly traced, the integral conductivity of which varies little in the depth range of 30–60 km. The absence of a continuous linkage between LROs in the 48°–49°N interval can be explained by large distances between the latitudinal profiles for which 2D inversions were performed.

At mantle depths, the LROs in the central part of the USh practically vanish. There remain objects that spatially coincide with the KirAC, DDD, the southern flank of the USh from 29° to 36°E and the southern EEP boundary. At all depths above 42 km (close to the base of the Earth's crust), the ρ background values vary between 100 and 4000 Ω m. At depths up to 7 km, there are three regions where the ρ values even exceed 4000 Ω m. The western region can be traced from the Pripyat Trough in the north to the southern slope of the USh between 29° and 30°E. The latitudinal region just to the south of 49°N stretches from the western boundary of the study area to the Inguletsko-Krivorogzhsko-Krupetskoi zone. The third region is located on the territory of the Ukrainian shield east of the IKKZ. At the depth up to 7 km all these zones can be fragmentarily traced. At depths, greater than 20 km regions of high resistivity are already observed mainly to the east of the IKKZ with the exception of the western region. They are determined up to 75 km depth from the DDD to Crimea. Spatially, the area east of the IKKZ belongs to the block of Archaean rocks.

For further analysis, only regional peculiarities of geoelectric parameters were used. The LROs with an area less than 10 × 10 km are not shown on the slice maps. The aim of this analysis was to show the possible relationship between the separate LROs and regional tectonic units (Fig. 4), metallogeny (Fig. 5) and oil and gas presence (Fig. 6).

The migration pathways of fluids with ore elements that can make up deposits are associated with fault zones. Figure 4 shows fault zones active state in the last 5–7 million years (Verkhovtseff 2006; National Atlas of Ukraine 2007). Reactivation should be accompanied by a change in physical parameters in such zones in comparison with the host medium, and the concentration of ore elements leads to a local decrease in ρ values.

The KirAC is the main regional geoelectric structure that crosses the entire study area from the south to the north. It is spatially related to the IKKZ, which is of greatest metallogenic importance in Ukraine. Here are located giant deposits of iron ores, manifestations of graphite, gold, silver, zinc, copper, lead, as well as uranium, scandium, rare metals, etc. (Fig. 6). Most of the LROs forming the KirAC occur along the main Krivorozhsko–Kremenchugskii deep fault zone (Fig. 5). The length of the zone is about 900 km. Although on the tectonic maps, the IKKZ is delineated only up to the middle of the Southern-Ukrainian monocline. Figure 5 shows that active faults consistent with the zone can be traced further to the south even to the Black Sea. As the experimental data of the MT field are replenished, the geoelectric parameters of the KirAC are refined, which can be seen from publications for different years (Gordienko et al. 2005, 2006; Kirovograd 2013; Varentsov et al. 2013; Logvinov 2015; Logvinov et al. 2016).

The results of the modeling allow us to delineate one more regional high-conductive zone named Konskaya

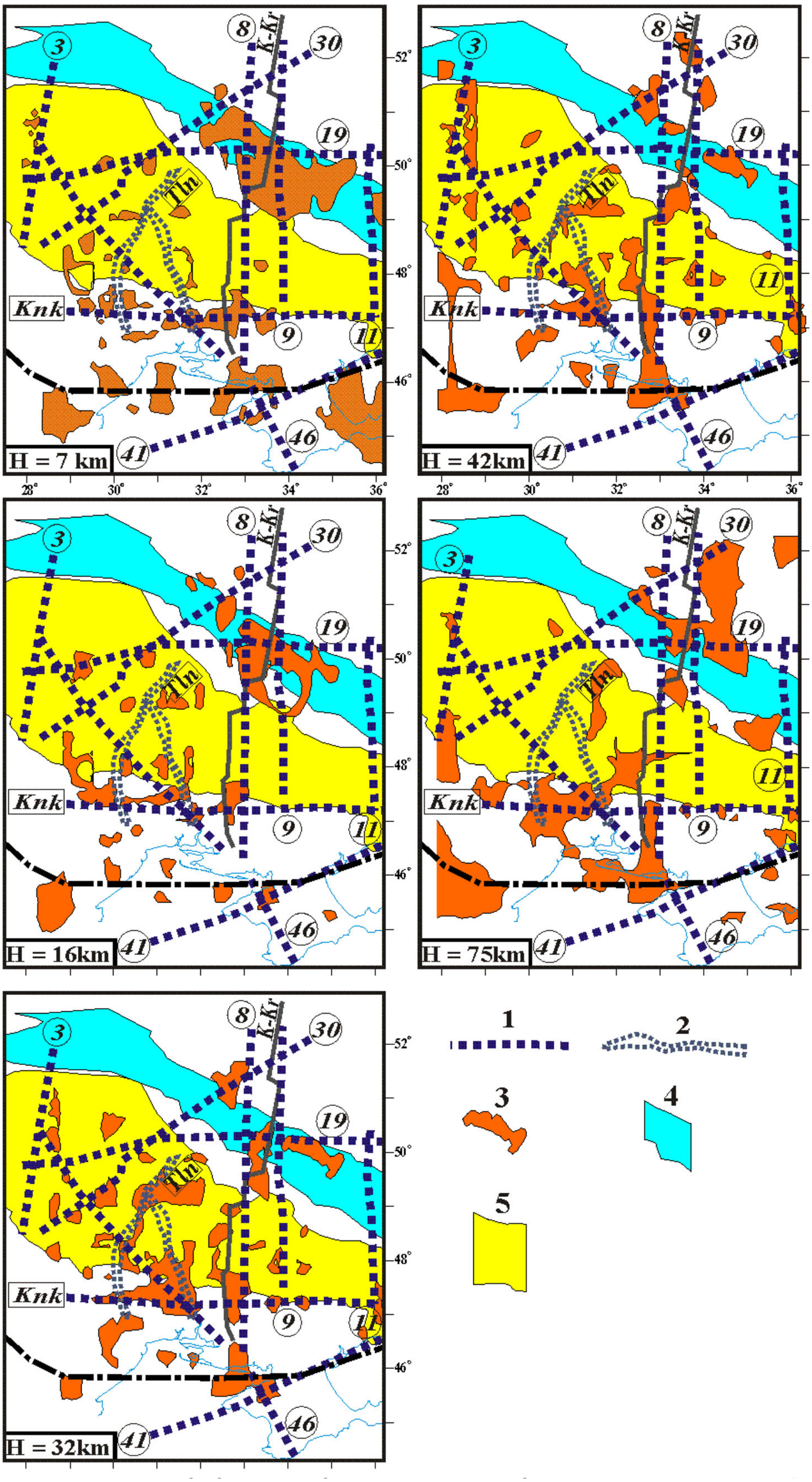


Fig. 4 Regional fault zones reactivated during the past 5–7 mln years: 1—to (Verkhovtseff 2006) Olevsko-Muravanskii—3, Novgorodsko-Seversko—Novo-Kakhovskii—8, Druzhba-Ordzhonikidzevskii—9, Dergach-Gulyaypol—11, Yavoriv-Volchanskiy—19, Mogilev-Podolsk-Shostkinsky—30, Sivash-Karkinitskiy—41, Kamen-Kashirsko-Yalta—46; 2—to (National Atlas of Ukraine 2007; Kirovograd 2013) Knk—Konkskii, Tln—Talnovskii, K–Kr (Krivorogsko-Kremenchugskii. Contours: of geoelectric objects (3), 4—DDG, 5—USh. See Fig. 1B for tectonic setting

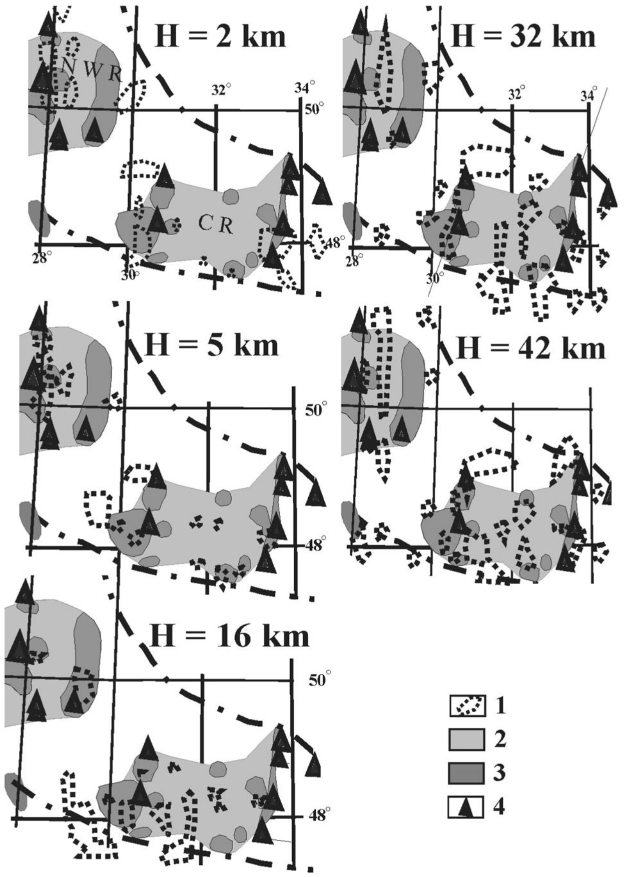


Fig. 5 Contours of LROs (1). Graphitization areas (Yatsenko 1998): regions (2) and districts (3). Districts of polymetallic ore deposits—4 (Galetskij and Shevchenko 2006)

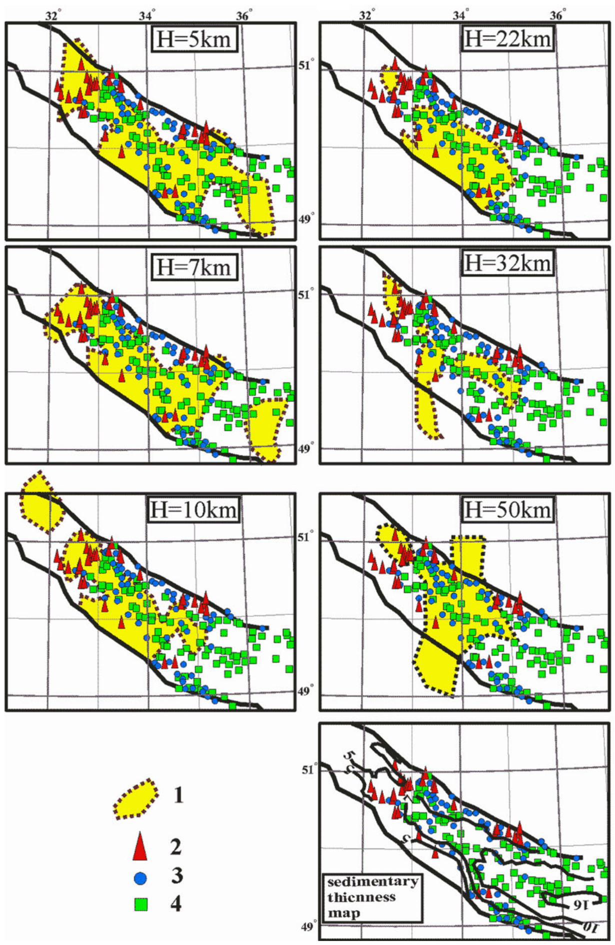


Fig. 6 Contours of LROs (1 − ρ values less than 40 Ω m). Areas of deposits: 2—oil, 3—oil–gas-and-gas condensate, 4—gas and gas condensate (Ivanyuty et al. 1998)

(Fig. 4). It is located along the southern slope of the USh, almost coinciding with the position of the Konskii fault, which has reactivated in the last 3 million years (Verkhovtseff 2006). The LROs forming this zone from 29° to 34°E are distinguished from 2 to 75 km depth. The maximum integral conductivity of the LROs is observed at ca. 30 km depths. Similar conductivity anomalies are also found in other margins of the East European Platform (Ernst et al. 2008) and in South American Platform (Bologna et al. 2014).

The LROs associated with the southern boundary of the EEC are traced up to 75 km depth, decreasing in size and integral conductivity at depths greater than 7 km. Taking into consideration the thickness of the sedimentary rocks which are spatially consistent with the LROs, most of the zones seem to be caused by the conductivity of the sediments. The vast LROs in the western study area and in the junction zone of the EEP and Crimea are most likely caused by the reactivation of the separate areas of the Scythian plate.

Along the western boundary of the study area between 28° and 29°E, a LROs chain is observed close to the Olevsko-Muravanskii fault, which has reactivated in the last 3 million years (Verkhovtseff 2006). The LROs are revealed at two depth intervals of 2–5 and 32–50 km, respectively. The chain of the LROs can be traced along Talnovskiy and the faults adjacent to it from the east, which delineate the Golovanevskaia Precambrian zone of the mantle origin. These LROs are mapped in this zone through the entire crust.

Thus, most of the LROs clearly correlate with fault zones of ancient origin which have been reactivated over the last 3 million years.

"The Ukrainian shield is exceptionally enriched in graphite and is estimated in this context as a unique province all of the World" (Yatsenko 1998). On the USh, graphitization areas are revealed (Fig. 5) in rock types from greenschist to granulite facies of metamorphism (Yatsenko 1998). In local areas of deposits and manifestations of graphite, its concentration reaches 15–20%. The graphite concentration within some areas can be estimated to be one percent that corresponds to a value of ρ of about 10 Ω m

(Gordienko et al. 2005). Within the north-western and central graphitization regions, the average mineral content is estimated to be 0.5–1% that roughly corresponds to ρ = 100 Ω m. The distribution of the graphitization zones with depth is unclear.

The magnitude of the erosional cut of USh rocks varies from 13 to 18 km while in the IKKZ it reaches up to about 30 km in the western part of the central region. This indicates that graphitized rocks in neighboring areas are formed at depths that differ by 10 or more kilometers. Graphitisation documented at shallow depth can also occur at considerable depths.

Most of the LROs are found in the central region that may be due to a smaller number of interpretable profiles in the NW region. The LRO dimensions in the graphitization areas are small and similar to those of the graphite-bearing areas (Fig. 5). The largest graphite deposit in Europe (Zaval'evskoe) is located in the southwest corner of the central region within known LROs. The LROs area and their integral conductivity decrease in the depth interval of 2–7 km. Below 10 km, their parameters increase, reaching the maximum values in the lower crust (30–40 km depth). Starting with a depth of about 5 km up to the mantle surface the LROs chain is clearly traceable along the perimeter of the central region contouring the central graphite-bearing region.

Figure 5 also shows the location of numerous polymetal deposits (Ni, Cu, Zn, Pb, etc.). As visible they spatially coincide with the graphitization areas. It allows us to suggest a relation between LROs and graphitization and to consider some LROs as ore regions.

Magnetotelluric research has been used in oil exploration for a long time; some references on this work can be obtained in articles by Keller (1968, 1969) and Caldwell (1969). A connection with oil and gas content and well-conducting object in the territory of the North German Basin is noted in Hoffmann et al. 2008. In Fig. 6, a comparison is presented between the geoelectrical results and oil and gas presence in the DDG. At depths up to 16–22 km, the LROs area in the DDD is almost completely consistent with the distribution of the main oil and gas fields. The vanishing of the LROs also agrees with the western boundary of the oil- and gas-bearing area. The absence of interpretation profiles and development of salt-dome tectonics between the 34° and 36°E do not allow us to extend this conclusion to the entire DDD. The oil fields appear to occur more frequently in the marginal zones of the LROs.

The results obtained allow for the following conclusion. Naturally, regional studies cannot be used to search for oil and gas fields. We can see the agreement of the geoelectric parameters with the oil-and gas-bearing area corresponding to the abiogenic hypothesis in the depth interval of 5–24 km. Geoelectrical data confirm the hypothesis of the association of hydrocarbon fields with zones of generation and transporting of hydrocarbon-bearing fluids which consequently cause the anomalous conductivity (Sherwood Lollar et al. 2002; Geoffrey 2006; Sephton and Hazen 2013).

Despite poor observations, network and ambiguity in the interpretation of geoelectrical data, the spatial coincidence of the LROs with the tectonic units, metallogeny, and oil and gas presence makes it possible to apply deep geoelectric data to recommend prospective areas for deposits of different mineral resources. For this purpose, more detailed geoelectrical studies and combination with other geological–geophysical information are necessary.

Acknowledgements Thanks to Prof. O.M. Rusakov for stimulating the study, discussion of the results and editing the manuscript. We thank the reviewers—their comments have significantly improved the presentation of the results of the research.

References

Bologna MS, Padilha AL, Pádua MB, Vitorello Í, Chamalaun FH (2014) Paraguay-araguaia belt conductivity anomaly: a fundamental tectonic boundary in South American Platform imaged by electromagnetic induction surveys. Geochem Geophys Geosyst 15(3):509–515. https://doi.org/10.1002/2013gc004970

Caldwell RL (1969) Electrical methods of exploration in the Soviet Union, in unconventional methods in exploration for petroleum and natural gas. Southern Methodist University, Dallas, pp 87–104

Ernst T, Brasse H, Cerv V, Hoffmann N, Jankowski J, Jozwiak W, Kreutzmann A, Neska A, Palshin N, Pedersen LB, Smirnov M, Sokolova E, Smirnov M (2008) Electromagnetic images of the deep structure of the Trans-European Suture Zone beneath Polish Pomerania. Geophys Res Lett 35(15):L15307. https://doi.org/10.1029/2008GL034610

Galetskij LS, Shevchenko TP (2006) New understanding of the structure and ore crust Ukraine. Geofyz J 28(5):57–65 **(in Ukrainian)**

Geoffrey PGLASBY (2006) Abiogenic origin of hydrocarbons: an historical overview. Resour Geol 56(1):83–96. https://doi.org/10.1111/j.1751-3928.2006.tb00271.x

Gintov OB (2005) Field tectonophysics and its application in the study of deformations of the earth's crust of Ukraine. Phoenix, Kiyv, p 572 **(in Russian)**

Gordienko VV, Gordienko IV, Zavorodnyaya OV, Logvinov M, Tarasov VN, Usenko OV, Kovacikova S (2005) Ukranian shield (geophysica, deep processing). Korvin Press, Kiev, p 210 **(in Russian)**

Gordienko VV, Gordienko IV, Zavgorodnyaya OV, Kovachikova S, Logvinov IM, Pec J, Tarasov VN, Usenko OV (2006) Dnieper-Donetsk Basin (Geophysics, deep processes). Korvin Press, Kiev, p 144 **(in Russian)**

Hoffmann N, Hengesbach L, Friedrichs B, Brink H-J (2008) The contribution of magnetotellurics to an improved understanding of the North German Basin—review and new results. Zeitschrift der Deutschen Gesellschaft für Geowissenschaften 159(4):591–606. https://doi.org/10.1127/1860-1804/2008/0159-0591)

Ingerov AI, Rokityansky II, Tregubenko VI (1999) Forty years of MTS studies in the Ukraine. Earth Planet Space 51:1127–1133

Ivanyuty MM, et al. eds.(1998) Atlas of oil and gas in Ukraine, vol I–III. Ukrainian Oil and Gas Academy, Lviv. **(in Ukrainian)**

Keller GV (1968) Electrical prospecting for oil: quarterly of the Colorado school of mines, vol 63, 2nd edn, p 267

Keller GV (1969) Electromagnetics may be the key to direct oil finding: world oil, Dec, pp 85–88

Kirovograd ore district (2013) Deep structure. Tectonophysical analysis. Deposits of ore minerals. Starostenko VI, Gintov OB, editor. Simple people. Kyiv. p 500. **(in Russian)**

Kovacikova S, Logvinov I, Nazarevych A, Nazarevych L, Pek J, Tarasov V, Kalenda P (2016) Seismic activity and deep conductivity structure of the Eastern Carpathians. Studia Geophysica et Geodaetica 60(2):280–296

Logvinov IM (2002) Geoelectric parameters of sedimentary of the Dnieper-Donets basin according to the data of two-dimensional modeling of MTS results. Geofiz J 24(6):79–92 **(in Russian)**

Logvinov IM (2015) Deep geoelectrical structure of the central and Western Ukraine. Acta Geophys 63(5):1216–1230. https://doi.org/10.1515/acgeo-2015-0049

Logvinov IM, Tarasov VN, Tregubenko VI (2016) Geoelectric model of the junction zones of the Scythian plate with the East European Platform and the mountainous Crimea. Dokladi NAS Ukraine 7:78–85. https://doi.org/10.15407/dopovidi2016.07.078 **(in Russian)**

National Atlas of Ukraine (2007) Cartographia. State Scientific Production Enterprise, Kyiv, p 440 **(in Ukrainian)**

Rokityansky II (1982) Geoelectromagnetic investigation of the earth's crust and mantle. Springer, Berlin. https://doi.org/10.1007/978-3-642-61801-7

Rokityansky II, Logvinov IM, Luginina NA (1969) Magnitovariational profiling on the Ukrainian shield Izv, vol 3. The USSR Academy of Sciences. Fizika of Earth, Moskva, pp 100–111 **(in Russian)**

Semenov VYu, Pek J, Adam A, Jozwiak W, Ladanyvskyy B, Logvinov I, Pushkarev P, Vozar J (2008) Electrical structure of the upper mantle beneath Central Europe: results of the CEMES project. Acta Geophys 56(4):957–981

Sephton Mark A, Hazen Robert M (2013) On the origins of deep hydrocarbons. Rev Mineral Geochem 75:449–465. https://doi.org/10.2138/rmg.2013.75.14

Sherwood Lollar B, Westgate TD, Ward JD, Slater GF, Lacrampe-Couloume G (2002) Abiogenic formation of alkanes in the Earth's crust as a minor source for global hydrocarbon reservoirs. Nature 446(6880):522–524. https://doi.org/10.1038/416522a

Siripunvaraporn W, Egbert G (2000) An efficient data-subspace inversion method for 2-D magnetotelluric data. Geophysics 65(3):791–803

Varentsov IVM, Gordienko VV, Gordienko IV, Zavgorodnyaya OV, Kovachikova S, Logvinov IM, Tarasov VN, Tregubenko VI (2013) The slope of the Voronezh crystalline massif (geophysics, deep processes). Logos, Kiev, p 112 **(in Russian)**

Verkhovtseff V (2006) Newest vertical crustal movements in Ukraine, their relationship with linear and circular structures. In: Vyzhva SA (ed) Power earth, its geological and environmental displays, scientific and practical use. Kyiv National University, Kyiv, pp 129–137 (**in Ukrainian**)

Yatsenko VG (1998) Regularities of the spatial arrangement of graphite manifestations on the Ukrainian shield. Aspects of mineralogy in Ukraine. GNC ROS, Kiyv, pp 254–270 **(in Russian)**

Acta Geophysica (2018) 66:141–152
https://doi.org/10.1007/s11600-018-0123-4

RESEARCH ARTICLE - SOLID EARTH SCIENCES

Evaluation of strength and failure of brittle rock containing initial cracks under lithospheric conditions

Xiaozhao Li[1,2] · **Chengzhi Qi**[1] · **Zhushan Shao**[2] · **Chao Ma**[1]

Received: 5 October 2017 / Accepted: 14 February 2018 / Published online: 19 February 2018

Abstract

Natural brittle rock contains numerous randomly distributed microcracks. Crack initiation, growth, and coalescence play a predominant role in evaluation for the strength and failure of brittle rocks. A new analytical method is proposed to predict the strength and failure of brittle rocks containing initial microcracks. The formulation of this method is based on an improved wing crack model and a suggested micro–macro relation. In this improved wing crack model, the parameter of crack angle is especially introduced as a variable, and the analytical stress–crack relation considering crack angle effect is obtained. Coupling the proposed stress–crack relation and the suggested micro–macro relation describing the relation between crack growth and axial strain, the stress–strain constitutive relation is obtained to predict the rock strength and failure. Considering different initial microcrack sizes, friction coefficients and confining pressures, effects of crack angle on tensile wedge force acting on initial crack interface are studied, and effects of crack angle on stress–strain constitutive relation of rocks are also analyzed. The strength and crack initiation stress under different crack angles are discussed, and the value of most disadvantaged angle triggering crack initiation and rock failure is founded. The analytical results are similar to the published study results. Rationality of this proposed analytical method is verified.

Keywords Brittle rocks · Most disadvantaged crack angle · Friction · Strength · Failure

Introduction

In brittle rock, existence of microcracks greatly influences its mechanical behavior. The process of microcrack extension plays an important role in strength and failure of rock under lithospheric conditions. The compressive loading conditions acting on the rocks can be approximately regarded as a form of lithospheric conditions. Thus, the laboratory study of mechanical behaviors in rocks under compressive loadings has an important meaning for evaluating the geomechanics under lithospheric conditions. The rock fracture induced by microcrack growth under compression was widely studied by acoustic emission monitoring (Cox and Meredith 1993; Martin 1997; Caia et al. 2004; Chang and Lee 2004; Diederichs et al. 2004; Damjanac and Fairhurst 2010; Lan et al. 2010; Zhou et al. 2017). Published experimental results showed that the stress–strain curves for a compressed brittle rock could be divided into six phases: crack closure, linear elastic deformation, crack initiation, stable crack growth, unstable crack growth corresponding to the permanent deformation, failure and postpeak behavior (Martin and Chandler 1994). The microscopic fracture processes of granite were studied by volumetric strain and crack volume (Brace et al. 1966). The deformation of brittle rock was divided into two major components: elastic deformation and brittle deformation caused by axial cracking (Lajtai 1998).

Due to the complex distribution of numerous initial microcracks in brittle rocks, effects of microcrack on strength and failure of brittle rock hardly are investigated

✉ Xiaozhao Li
lixiaozhao@bucea.edu.cn

✉ Chengzhi Qi
qichengzhi65@163.com

[1] School of Civil and Transportation Engineering, Beijing Advanced Innovation Center for Future Urban Design, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

[2] School of Civil Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China

directly. Thus, effect of crack was experimentally investigated by use of brittle specimen which contains minor pre-existing open crack (see Fig. 1). Mechanism of microcrack initiation, propagation and coalescence during fracture process under uniaxial or biaxial compression has been widely investigated (Mavko and Nur 1978; Hoek and

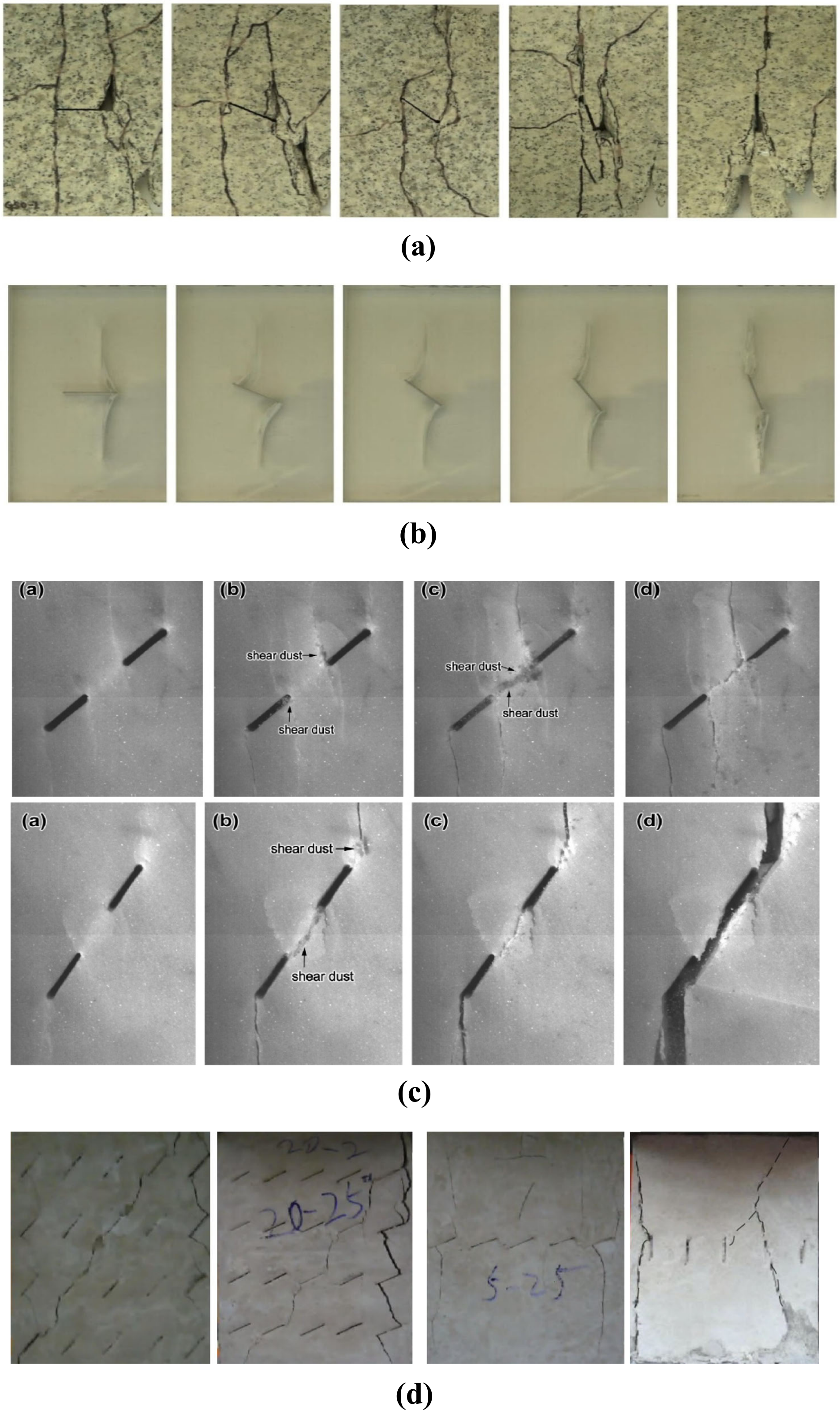


Fig. 1 Compressed cracking process under different crack inclinations for brittle materials of **a** granite (Lee and Jeon 2011), **b** PMMA (Lee and Jeon 2011), **c** marble (Wong and Li 2013), **d** concrete (Cao et al. 2016)

Bieniawski 1984; Bobet and Einstein 1998; Germanovich et al. 1994; Clayton 2010; Yang and Jing 2011; Morgan al. 2013; Silva and Einstein 2013; Wu et al. 2017; Fan et al. 2017). Lajtai (1971, 1974) studied the initiation and propagation of pre-existing minor flaws and their effects on compressed brittle rock, and examined the effect of initial flaw inclination angle on fracture initiation from tensile stress concentrations of a narrow crack. Using pre-existing open straight flaw with different inclination angles in rocks, effects of flaw inclination angle on crack initiation, propagation and coalescence were studied using of numerical and experimental method (Sagong and Bobet 2002; Lee and Jeon 2011; Li and Wong 2012; Wong and Li 2013). Their results showed that flaw inclination angles have great influence on mechanical properties of rock. However, the above studies of crack coalescence in rock materials were mainly focused on the specimen containing open flaws. Park and Bobet (2010) experimentally studied crack propagation and coalescence from frictional discontinuities in pre-cracked specimens of brittle rocks. The fracturing processes in open and closed flaws are similar. But, friction along the cracks increases the stress of crack initiation and coalescence. Combining similar material testing and discrete element method, the peak strength, failure characteristics, and crack coalescence pattern of rock-like materials with multi-cracks with similar orientation were studied, and effect of the inclination and density of cracks on mechanical properties and failure mechanism of the rock mass were analyzed (Cao et al. 2016; Vergara et al. 2016).

To understand the mechanism of strength and failure of brittle rock containing microcracks, many numerical models were suggested. Wong et al. (2006) presented statistical model which is established using of wing crack model and Weibull parameters. Singh and Digby (1989) suggested a constitutive damage model that describes the anisotropic response of brittle solid during cracking under applied stresses. Zhou et al. (2008) presented a micromechanical model of material containing intermittent cracks and studied the effect of crack angle on stress–strain relationship. Ashby and Hallam (1986) and Ashby and Sammis (1990) established a theoretical model that is developed for the growth and interaction of cracks in brittle solids under compressive stress states. This model presented a micromechanical explanation for strength of brittle rock. However, the crack angle, in Ashby model, was assumed as a certain value (i.e., $\varphi = 45°$and $\varphi = [\tan - 1(1/\mu)]/2$) in which microcrack is subjected maximum shear stress. Based on Ashby model, the effects of compressive loads and strain rates on the rock failure from microcrack growth were studied (Bhat et al. 2011, 2012; Brantut et al. 2012).

For the above analytical, numerical and experimental studies, it can be seen that the microcrack growth have great significance for mechanical behaviors under compression in brittle rocks. However, the analytical method to explore the mechanism of strength and failure considering effect of crack angle is rarely proposed. In this paper, employing an improved micromechanical wing crack model considering effect of crack angle, an analytical investigation of strength and failure for brittle rock containing initial cracks under compression will be presented. Effect of crack angle, initial crack size, confining pressure and friction coefficient on crack initiation stress, strength, stress–crack relation, and stress–strain relation will be studied in detail.

Description of analytical method

A wing crack model of rocks with same inclination and size of initial cracks is shown in Fig. 2 (Ashby and Sammis 1990). The failure of strain localization is approximately studied by the global failure of brittle rocks (Fig. 3). Subjected to the remote compressive stress σ1 and σ3, the stress intensity factor of mode-I crack (i.e., KI) in this model is (Ashby and Sammis 1990):

$$K_{\mathrm{I}} = \frac{F_{\mathrm{w}}}{[\pi(l+\beta a)]^{\frac{3}{2}}} + \frac{2}{\pi}(\sigma_3 + \sigma_3^i)\sqrt{\pi l}, \tag{1}$$

where a represents the size of initial crack, l represents the length of wing crack, and β is a correction factor to adjust the value of K_{I} at $l = 0$. F_{w} is wedging force acting on initial crack surface. $\sigma i3$ is internal stress acting between wing cracks, which has an important meaning for evaluation of interaction and coalescence between cracks to

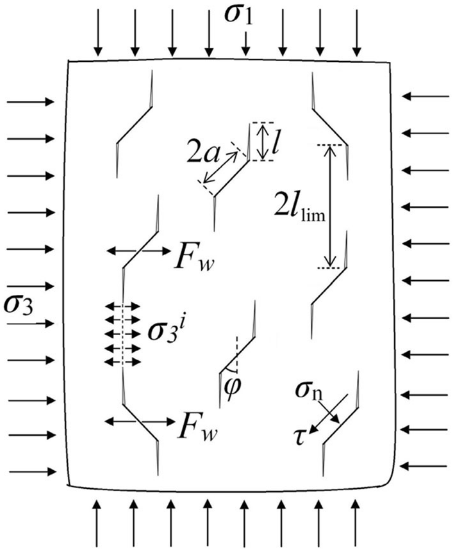


Fig. 2 Schematic diagram for mechanical model of brittle rocks induced by wing crack growth under compressive loadings

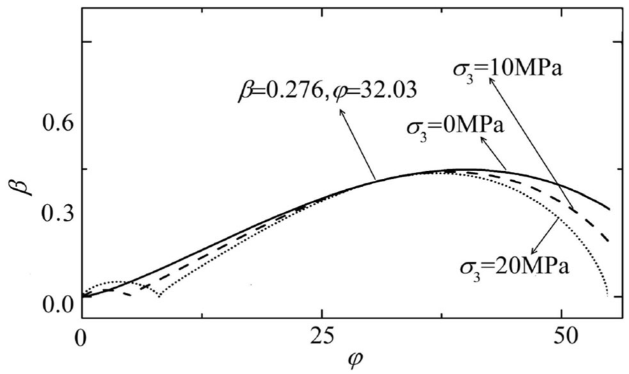


Fig. 3 Prediction of parameters β and φ

predict the failure of rocks. This internal stress is determined as:

$$\sigma_3^i = \frac{F_w}{S - \pi(l + \alpha a)^2}, \tag{2}$$

where $\alpha = \cos\varphi$, φ is the angle between the initial crack and the axial stress $\sigma 1$ (i.e., crack angle). In this model, the individual crack is assumed to occupy averagely a spherical volume. This spherical volume is $1/\mathrm{NV} = 4\pi r3/3$, where NV is the initial crack number in unit volume. Based on this assumption, the average area S of the individual crack is:

$$S = \pi^{1/3}\left(\frac{3}{4N_V}\right)^{2/3}. \tag{3}$$

Following the definition of Ashby and Sammis (1990), the wedging force $F_w = -(\mathrm{A}1\sigma 1\text{–}\mathrm{A}2\sigma 3)\mathrm{a}2$, where $\mathrm{A}1 = \pi(\beta/3)1/2[(1 + \mu 2)1/2\text{–}\mu]$, $\mathrm{A}2 = \mathrm{A}1[(1 + \mu 2)1/2 + \mu]/[(1 + \mu 2)1/2\text{–}\mu]$, which is derived by some assumptions (e.g., $\tan 2\varphi = 1/\mu$) and simplifications. But, the effect of crack angle cannot be considered. In this paper, the assumption and simplification are neglected, and F_w is expressed as:

$$F_w = (\tau + \mu\sigma_n)\pi a^2 \sin\varphi, \tag{4}$$

where

$$\tau = \frac{\sigma_3 - \sigma_1}{2}\sin 2\varphi, \tag{5}$$

$$\sigma_n = \frac{\sigma_1 + \sigma_3}{2} + \frac{\sigma_3 - \sigma_1}{2}\cos 2\varphi, \tag{6}$$

where μ is friction coefficient acting between initial crack interfaces. τ and σn are the shear stress and the normal stress acting on crack plane, respectively. It is clearly illustrated in the results and discussions below that the effect of crack angle is important to the strength and failure behaviors when the assumption and simplification of crack angle are neglected. Compared with the influence of KI on the wing cracking, the influences of KII and KIII are negligible in this model (Bhat et al. 2011; Brantut et al. 2012).

Furthermore, in Eq. (2), $\pi(l + \alpha a)2$ is a projected area of initial crack containing two wing cracks to a plane. This plane is perpendicular to direction of confining pressure $\sigma 3$. Furthermore, Eq. (3) shows an average area of the individual crack. Assuming that this average area of individual crack of Eq. (3) is equal to the projected area of initial crack containing two wing cracks [i.e., $S = \pi(l + \alpha a)2$], a limitation length for coalescence of two adjacent wing cracks can be solved as:

$$l_{\lim} = \left(D_o^{-1/3}\alpha^{-1} - 1\right)\alpha a, \tag{7}$$

where D_o is initial damage of model. In Ashby and Sammis' (1990) study, $D_o = 4\pi\mathrm{NV}(\alpha a)3/3$. However, to consider the effect of crack angle on strength and failure in this proposed analytical method, the initial damage is replaced with the expression of $D_o = \mathrm{NV}a3$ (Budiansky and O'Connel 1976).

Crack in rocks could extend at $K_I = K_{IC}$, where K_{IC} is fracture toughness. Therefore, based on Eq. (1), a stress–crack relation (i.e., the equation of wing crack growth under compressive loadings) is obtained as:

$$\sigma_1(l) = \frac{K_{IC} - \sigma_3\left(A_1\pi a^2\eta\sin\varphi/2 + 2\sqrt{l/\pi}\right)}{A_2\pi a^2\eta\sin\varphi/2}, \tag{8}$$

where

$$A_1 = \sin 2\varphi + \mu + \mu\cos 2\varphi, \tag{9}$$

$$A_2 = \mu - \mu\cos 2\varphi - \sin 2\varphi, \tag{10}$$

$$\eta = [\pi(l + \beta a)]^{-3/2} + 2\sqrt{l/\pi}[S - \pi(l + \alpha a)^2]^{-1}. \tag{11}$$

This stress–crack relation has a maximum value of axial stress (which is corresponding to the rock compressive strength of brittle rocks), which will be explained in "Results" and "Discussion".

Based on Weibull (1951) distribution, the damage defined by the axial strain ε under compression of natural brittle rocks is (Chen et al. 1997):

$$D = 1 - \exp[-(\varepsilon/\varepsilon_o)^m], \tag{12}$$

where m and ε_o are constants of material.

Following the study by Budiansky and O'Connel (1976), the damage is obtained as:

$$D = N_V(l + a)^3. \tag{13}$$

Assuming the value of damage from Eq. (12) is equal to the value of damage from Eq. (13), the correlation of crack length and macroscopic axial strain can be derived as:

$$\varepsilon = \varepsilon_o \left[-\ln \left[1 - \frac{D_o (l + a)^3}{a^3} \right] \right]^{1/m}. \tag{14}$$

Substituting Eq. (14) in Eq. (8), the constitutive relation between axial stress and axial strain under compression is:

$$\sigma_1(\varepsilon) = \frac{K_{IC} - \sigma_3 (A_1 \pi a^2 J_1 \sin \varphi / 2 + 2\sqrt{J_2/\pi})}{A_2 \pi a^2 J_1 \sin \varphi / 2}, \tag{15}$$

where

$$J_1 = [\pi(J_2 + \beta a)]^{-3/2} + \frac{2\sqrt{J_2/\pi}}{S - \pi(J_2 + \alpha a)^2} \tag{16}$$

$$J_2 = a\left\{ D_o^{-1/3} [1 - \exp[-(\varepsilon/\varepsilon_o)^m]]^{1/3} - 1 \right\}. \tag{17}$$

Inserting Eq. (7) into Eq. (14), the axial failure strain caused by crack coalescence is predicted as:

$$\varepsilon_f = \varepsilon_o \left(-\ln \left[1 - \frac{(l_{\lim} + a)^3 D_o}{a^3} \right] \right)^{\frac{1}{m}}. \tag{18}$$

Equation (18) provides a failure criterion of rocks in the proposed analytical method.

Results

In this study, the experimental data of Jinping marble in China is used to validate the proposed analytical method (see Table 1). Some micro-parameters in model are hardly measured directly by experiment, and the specific values of micro-parameters are obtained by comparing the experimental and theoretical data, which provide an important help for applying the proposed analytical model to real rocks. The specific method for selection of model parameters is as follows.

A stress of crack initiation under different confining pressures is determined as (Ashby and Hallam 1986):

Table 1 Specific parameters of Jinping marble specimens in this research

Parameters	Values	References
Crack angle φ	32.03°	This study
Correction factor β	0.276	This study
Material constant m	1	This study
Fracture toughness K_{IC}	1.61 MPa·m1/2	Wan et al. (2010)
Material constant ε_o	0.0147	This study
Friction coefficient μ	0.51	This study
Initial crack size a	3.1 mm	This study
Initial damage D_o	0.048	This study

$$\sigma_{1c} = \frac{\sqrt{1 + \mu^2} + \mu}{\sqrt{1 + \mu^2} - \mu} \sigma_3 - \frac{\sqrt{3}}{\sqrt{1 + \mu^2} - \mu} \frac{K_{IC}}{\sqrt{\pi a}}. \tag{19}$$

A linear relation between confining pressure $\sigma3$ and axial crack initiation stress $\sigma1c$ is $\sigma1c = 2.67\sigma3 + 46$ MPa obtained by the conventional compressive test (Wang et al. 2008). Comparing this linear experimental relation with Eq. (19), both parameters a and μ are obtained, respectively (see Table 1).

Crack angle was assumed to be $\varphi = 45°$ (Bhat et al. 2011; Brantut et al. 2012; Ashby and Sammis 1990). The correction factor β was assumed to be a constant $\beta = 0.32$ (Bhat et al. (2011). The purpose of parameter β is to correct the value of K_I at $l = 0$ in Eq. (1), and the relationship between stress sate and crack length for Eq. (8) is derived from Eq. (1), it can be concluded that the crack initiation stress also is influenced by parameter β. In this paper, based on Eq. (8), the parameter β is solved as:

$$\beta = \left[\sqrt{a/\pi} (A_2 \sigma_1 (l{=}0) + \sigma_3 A_1) \sin \varphi \Big/ (2K_{IC}) \right]^{2/3}. \tag{20}$$

Combining Eq. (20) and empirical relation $\sigma1c = 2.67\sigma3 + 46$ MPa $= \sigma1$ $(1 = 0)$, the parameter β and crack angle φ both are obtained in Fig. 3 (i.e., $\varphi = 32.03°$, $\beta = 0.276$). It is noted that the crack angle is assumed as an average value for numerous and randomly distributed initial microcrack angles in natural rock, which provides an assistance in applying the proposed analytical approach to the prediction of strength and failure in rocks.

As shown in Fig. 4, the stress–strain curves are obtained. The applied axial stress increases first and then decreases with the increasing axial strain. The strength (i.e., peak stress) increases with the increasing confining pressure. The analytical results are similar to the experimental results (Wang et al. 2008). The difference between analytical and experimental results is from the assumption in which the failure of strain localization is equivalently studied by use of the global failure in the analytical method.

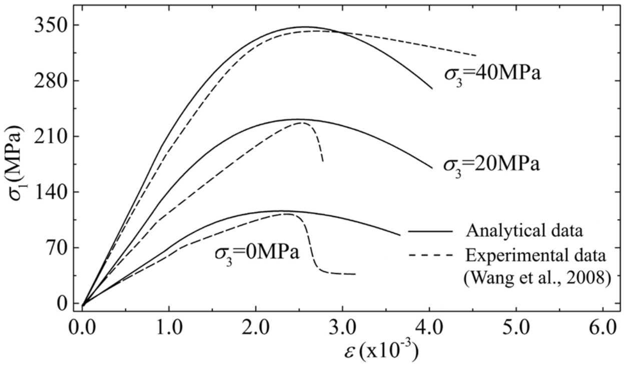


Fig. 4 Effect of confining pressure on stress–strain relation

Discussions

Effect of crack angle on wedge force

As shown in Fig. 5a, for friction coefficient $\mu = 0$, wedge force increases first and reaches a peak value, then decreases with increment of crack angle. This phenomenon is similar to the relationship between maximum tensile stress concentration around minor open elliptical crack boundary and crack orientation (Lajtai 1971). This open elliptical crack can be approximately regarded as the microcrack at friction coefficient equaling to zero in this study. It also can be seen that, when the crack angle is parallel (0°) or perpendicular (90°) to axial stress, wedge force will keep zero for any stress states ($\sigma1$, $\sigma3$), this result agrees with that of maximum tensile stress concentration around minor open elliptical crack boundary and crack orientation in Griffith fracture theory. For friction coefficient $\mu = 0$, the value of crack angle at peak point is always equal to 54.7°. It indicates that, without or ignoring effect of friction, the crack angle at peak point would not be affected by axial stress. And the wedge forces are always tensile force (positive) at $\mu = 0$, which means that the crack type is Mode I crack. Furthermore, for a given crack angle, wedge force increases with the increment of axial stress.

It can be seen from Fig. 5b that, for a given friction coefficient (i.e., $\mu = 0.8$), the crack angle at peak point increases with increment of axial stress (i.e., 30.9° at $\sigma1 = -500$ Mpa and 28.8° at $\sigma1 = -300$ Mpa). When the axial stress is small enough (i.e., $\sigma1 = 100$ Mpa), the wedge force will always be compressive force (negative), it means that the crack would not grow anyway. Only if axial stress exceeds a certain value, it could result in tensile wedge force, and the crack is possible to grow. Furthermore, the wedge force always is equal to zero when the crack angle is parallel (0°) to axial stress. Because of friction effect, the wedge force will be compressive stress when the crack angle exceeds a certain value. Only if the value of crack angle is in the limited range around peak point, wedge force could become tensile force, and the crack is possible to grow.

It can be concluded from Fig. 5 that the friction coefficient plays a significant role in crack growth under compression.

As shown in Fig. 6a, wedge force decreases with increment of confining pressure for a given crack angle. As shown in Fig. 6b, c, the range of crack angle which could lead to tensile wedge force decreases with increment of confining pressure, and the crack angle at peak point decreases with increment of confining pressure.

Comparing Fig. 6a with b, the wedge force decreases with increment of friction coefficient for given confining pressure and crack angle, and the crack angle at peak point decreases with increment of friction coefficient for given confining pressure. As shown in Fig. 6a, the crack angle at peak point is always equal to 54.7°, it means that confining pressure does not affect the crack angle at peak point when friction coefficient is equal to zero. When the friction coefficient is equal to 0.8, as shown in Fig. 6(b), the crack angles at peak point are 26.6°, 31.7° and 33.2° for different confining pressure $\sigma3 = 30$ Mpa, 10 Mpa, 0 Mpa, respectively, it means that the crack angle at peak point is influenced by the confining pressure in the case of friction coefficient not equal to zero.

Comparing Fig. 6b with c, the wedge force increases with increment of initial microcrack size at certain crack angle and confining pressure. The crack angle at peak point is same for different initial crack sizes (a = 0.5 mm and

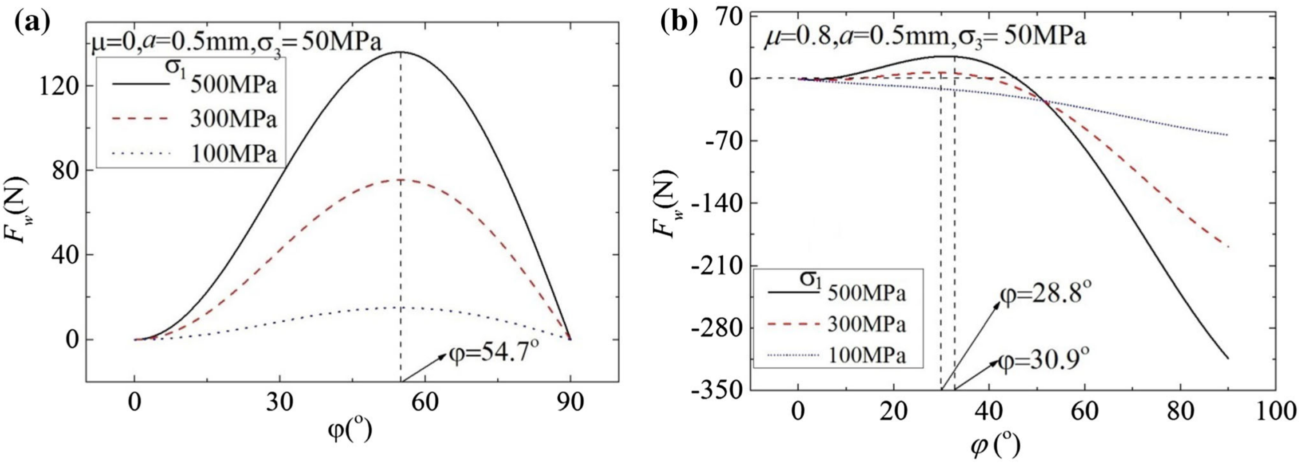


Fig. 5 Effects of axial stress on relation between wedge force F_w and crack angle φ with initial microcrack size $a = 0.5$ mm and confining pressure $\sigma3 = 50$ MPa. **a** $\mu = 0$; **b** $\mu = 0.8$. *Note* tensile wedge force is positive and compressive wedge force is negative

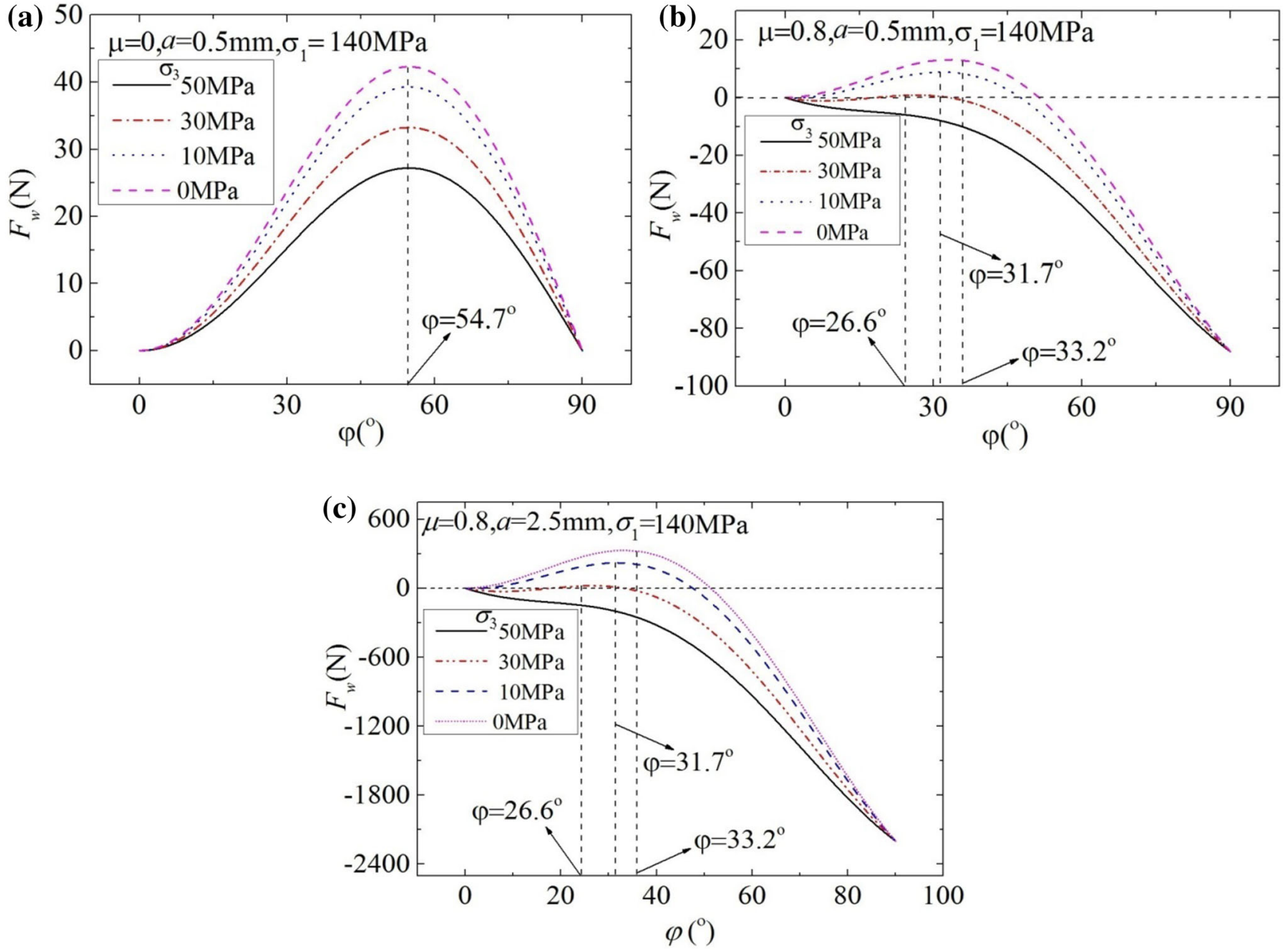


Fig. 6 Effects of confining pressure on relation between wedge force and crack angle. **a** $\sigma1 = 140$ MPa, $\mu = 0$, and $a = 0.5$ mm; **b** $\sigma1 = 140$ MPa, $\mu = 0.8$, and $a = 0.5$ mm; **c** $\sigma1 = 140$ MPa, $\mu = 0.8$, and $a = 2.5$ mm

2.5 mm), it means that initial microcrack size does not affect the crack angle at peak point.

In summary, crack angle at peak point is affected by friction coefficient μ and stress state ($\sigma1$, $\sigma3$) when friction coefficient is not equal to zero, and it will be a certain value ($\varphi = 54.7°$) when friction coefficient equal to zero. Moreover, the crack angle at peak point is not influenced by microcrack size. The initial microcrack would not grow under any compressive stress states when wedge force is zero or compressive force, which could provide a certain help for judging the failure of brittle rock.

In the case of confining pressure not equal to zero, analytical expression of crack angle at peak point is hardly obtained. To discuss the effect of the stress state ($\sigma1$, $\sigma3$), the crack angle at peak point with different confining pressure, initial microcrack size and friction coefficient is numerically calculated and graphically presented above. However, in the case of confining pressure $\sigma3 = 0$, a peak point of the relation between wedge force and crack angle in Eq. (4) can be obtained. This peak point should satisfy the condition $\partial F_{\mathrm{w}}/\partial\varphi = 0$. So, crack angle $\varphi F_{\mathrm{wmax}}$ (i.e., the maximum wedge force angle) at peak point can be solved as:

$$\varphi_{F_{\mathrm{wmax}}} = \tan^{-1}\left[\frac{1}{2}\left(\sqrt{8+9\mu^2} - 3\mu\right)\right]. \tag{21}$$

Furthermore, $\varphi(\tau + \mu\sigma n)$max is defined as an angle when the effective shear stress (i.e., $\tau + \mu\sigma n$) on failure plane of rocks equals a maximum value, i.e., $\varphi(\tau + \mu\sigma n)$max = (tan − 11/μ)/2 (Martin and Chandler 1994). It is noteworthy that the wedge force acts on the crack interface and the effective shear stress acts on shear failure plane. The maximum wedge force angle $\varphi F_{\mathrm{wmax}}$ is different from the critical failure plane angle $\varphi(\tau + \mu\sigma n)$max. φ$\varphi(\tau + \mu\sigma n)$max is a macroscopic angle and $\varphi F_{\mathrm{wmax}}$ is a micromechanical angle. Based on the two definitions of wedge force in crack interface and effective shear stress in failure plane, the relationship between these two angles is given in Fig. 7. It can be seen that, for any given friction coefficient, the maximum wedge force crack

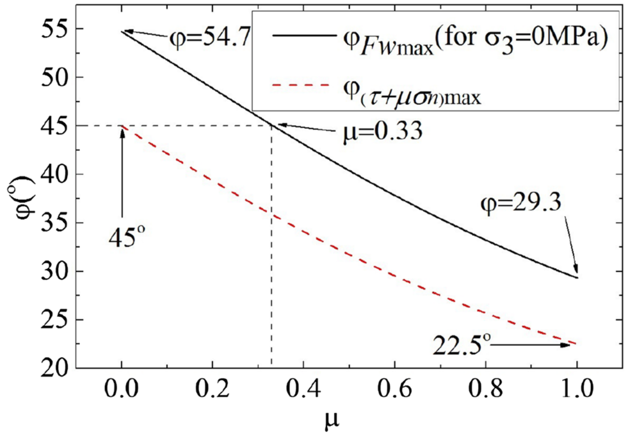


Fig. 7 Comparison of maximum wedge force crack angle $\varphi F_{\mathrm{wmax}}$ and critical failure plane angle $\varphi(\tau + \mu\sigma n)\mathrm{max}$

angle $\varphi F_{\mathrm{wmax}}$ is always larger than the critical failure plane angle $\varphi(\tau + \mu\sigma n)\mathrm{max}$. This phenomenon can be explained from the difference of definitions of friction coefficient for crack interface and macroscopic failure plane in brittle rock (Baud et al. 2014).

Effect of crack angle on stress–strain relation

Effect of crack angle on stress–strain relation under different friction coefficients is shown in Fig. 8a–c. Comparing Fig. 8a–c, it is observed that the number of curve for stress–strain decreases with increment of friction coefficient, which shows that the range of crack angle causing rock failure is decreasing. The most disadvantageous crack angles (i.e., corresponding to the smallest peak stress) are 55°, 45°, 35° in Fig. 8a–c, respectively, which show that the most disadvantageous crack angle decreases with the increasing friction coefficient. Furthermore, the most disadvantageous crack angle at $\mu = 0$ is 55° or so, it agrees well with the experimental result of Ashby and Hallam (1986) and numerical result of Zhou et al. (2008) in which the minor open flaw is considered.

Comparing Figs. 8 and 9, it is observed that peak stress decreases with increment of initial crack size at the given crack angle. Comparing Figs. 10 and 8c, peak stress increases with increment of confining pressure at the given crack angle. Figure 11 shows the relation between crack length and axial stress which is corresponding to the stress–strain relation of Fig. 10. The study results of effect of model parameters on the relation between crack length and axial stress are similar to the results of effect of model parameters on stress–strain relation.

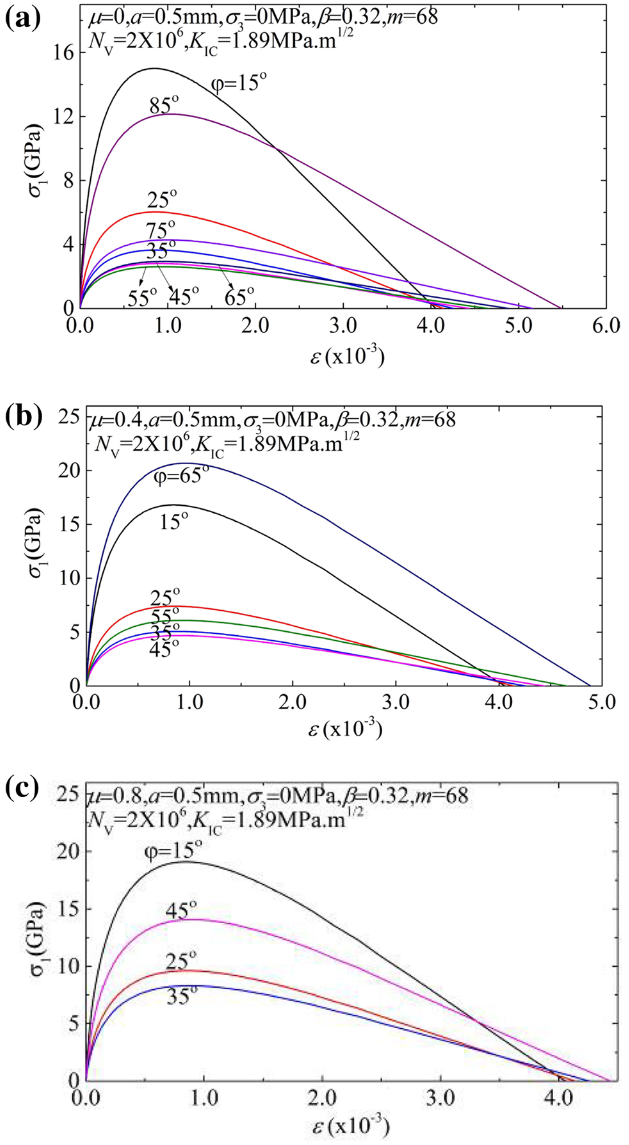


Fig. 8 Effects of crack angle on relationship between axial stress and strain ($a = 0.5$ mm, $\sigma 3 = 0$ MPa)

Effect of crack angle on crack length at peak stress

Figures 12 and 13 show the relation between crack angle and peak crack length (i.e., the crack length corresponding to peak stress) under different initial crack sizes and confining pressures. It is observed that peak crack length increases with the increasing crack angle. As shown in Fig. 12, peak crack length increases with decrement of initial crack size at given crack angle. As shown in Fig. 13, peak crack length increases with increment of confining pressure. Furthermore, the analytical results show that the friction coefficient does not influence the peak crack length. Thus, the effect of friction coefficient on peak crack length is not shown graphically in this study.

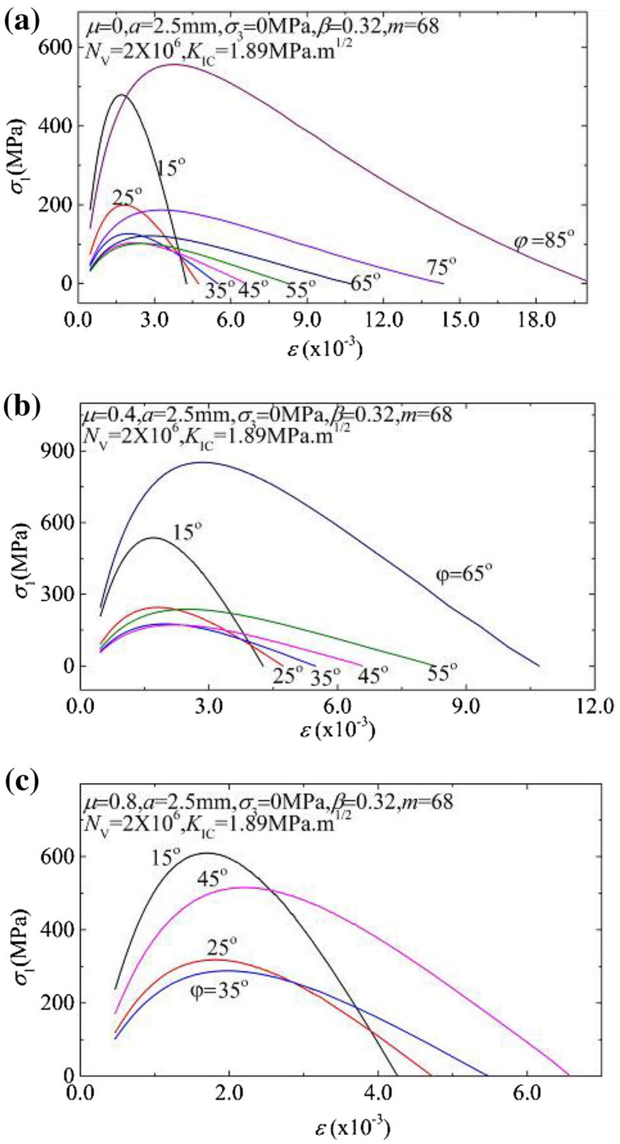


Fig. 9 Effects of crack angle on relationship between axial stress and strain ($a = 2.5$ mm, $\sigma3 = 0$ MPa)

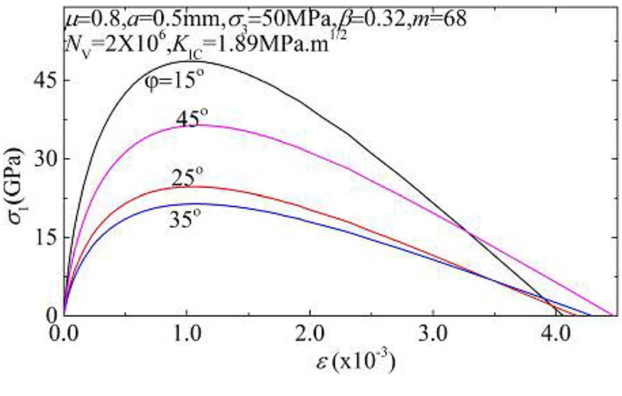


Fig. 10 Effects of crack angle on relationship between axial stress and strain (a = 0.5 mm, $\sigma3 = 50$ MPa)

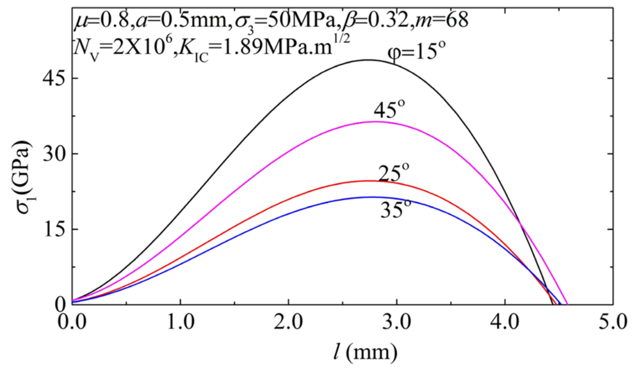


Fig. 11 Effects of crack angle on relationship between axial stress and crack growth

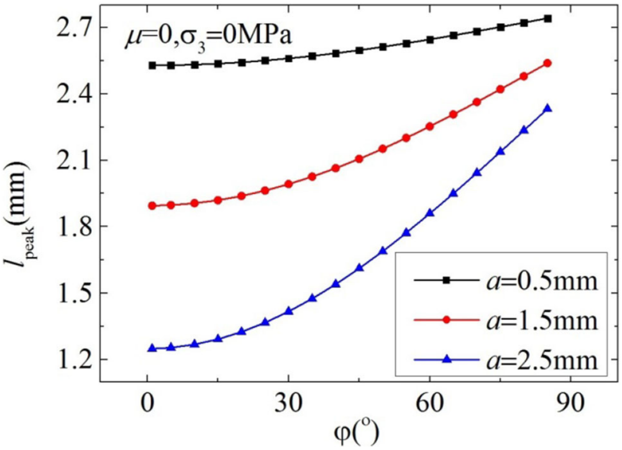


Fig. 12 Effects of initial crack size on relationship peak crack length and crack angle

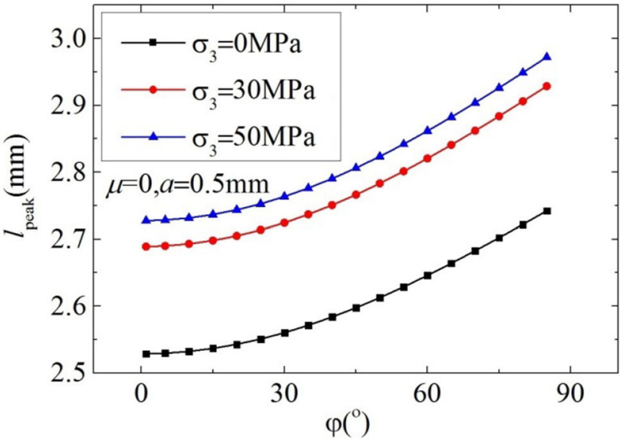


Fig. 13 Effects of confining pressure on relationship peak crack length and crack angle

Effects of crack angle on strength and crack initiation stress

Based on the above illustration of stress–strain relation and stress–crack relation in different crack angles, the effects of crack angle on crack initiation stress and peak stress will be studied below. Due to the random distribution of initial microcracks in brittle rocks, the effects of crack angle on mechanical properties of brittle rock almost cannot be investigated directly. However, the effects of the minor pre-existing flaw angle on mechanical behaviors of brittle rocks were widely investigated in compressive loadings (Lajtai 1974; Ashby and Hallam 1986; Sagong and Bobet 2002; Zhou et al. 2008; Park and Bobet 2010; Lee and Jeon 2011; Zhang and Wong 2012, 2013; Li and Wong 2012; Wong and Li 2013; Cao et al. 2016; Vergara et al. 2016). Thus, the analytical results will approximately be compared with the experimental result below.

As shown Figs. 14, 15 and 16, peak stress (i.e., rock strength) decreases and accesses a minimum value, and increases with increment of crack angle. This analytical phenomenon is in good agreement with the published study results in Fig. 17 (Lee and Jeon 2011; Zhang and Wong 2012, 2013; Vergara et al. 2016). Furthermore, crack initiation stress also decreases and accesses a minimum value, and increases with increment of crack angle, which is similar to the published study results of relation between crack initiation stress and flaw inclination angle under compression in Fig. 17 (Lee and Jeon 2011; Zhang and Wong 2012, 2013; Wong and Li 2013). The range of crack angle (at which crack growth can occur in compression) decreases with increment of friction coefficient. The crack initiation stress and peak stress both increase with increment of friction coefficient for given crack angle, it agrees well with the experiment results in which friction between closed flaw interfaces is considered (Park and Bobet 2010).

As shown in Fig. 15, the crack initiation stress and peak stress (i.e., rock strength) both decrease with increment of initial microcrack size for given crack angle. As shown in Fig. 16, crack initiation stress and peak stress both increase with increment of confining pressure for given crack angle, which agrees well with the study results of Vergara et al. (2016).

Conclusions

An analytical method is proposed to predict the strength and failure of brittle rocks. The formulation of this method is based on a wing crack model considering effect of crack angle and a micro–macro relation describing correlation of crack length and strain. Effects of microcrack angle on wedge force, stress–crack relation, and stress–strain relation with different initial microcrack sizes, friction coefficients and confining pressures are investigated. Rationality of the analytical study is verified by a comparison of the analytical results and the published experimental results

Wedge force increases first and reaches a peak value, and then decreases with increment of crack angle. Wedge force increases with increment of initial crack size or axial stress, and decrement of friction coefficient or confining pressure. Furthermore, a maximum wedge force angle is founded. This maximum wedge force angle is irrelevant to stress state when friction coefficient equals to zero, and it is relative to stress state when friction coefficient does not equal to zero. The maximum wedge force angle increases with increment of axial stress, and decrement of confining pressure. The maximum wedge force angle is irrelevant to initial microcrack size. This maximum wedge force angle increases with friction coefficient between crack interfaces.

Crack initiation stress and strength both decrease first and reach a minimum value, and then increase with increment of crack angle from 0° to 90°, which agrees well

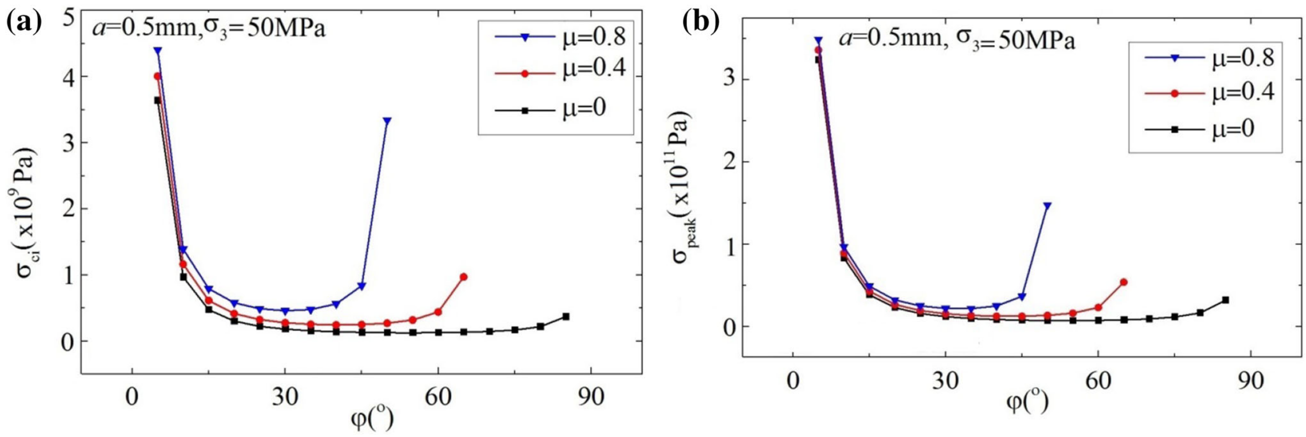


Fig. 14 Effects of friction coefficient on relationships between **a** crack initiation stress σci, **b** peak stress σpeak and crack angle φ

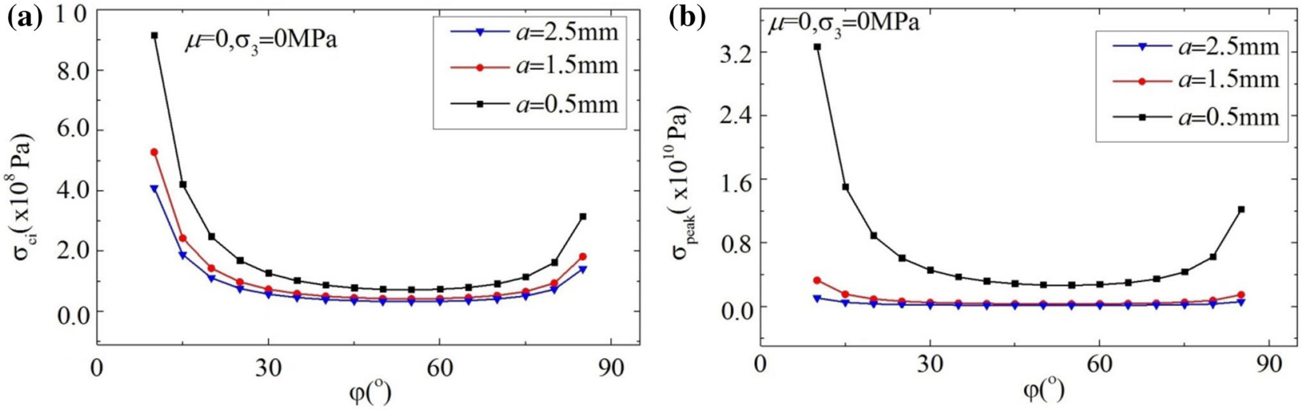


Fig. 15 Effects of initial crack size on relationships between **a** crack initiation stress, **b** peak stress and crack angle

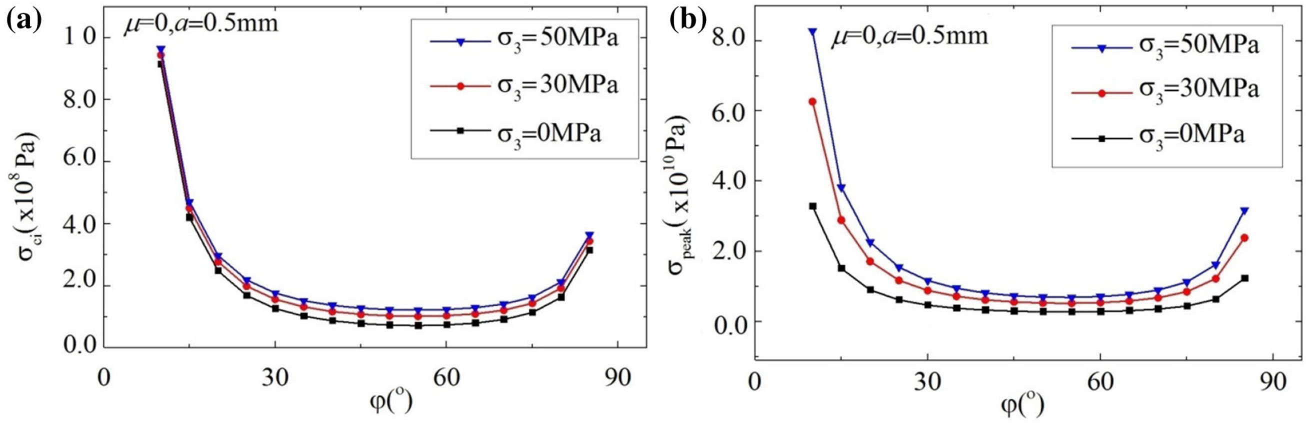


Fig. 16 Effects of confining pressure on relationships between **a** crack initiation stress, **b** peak stress and crack angle

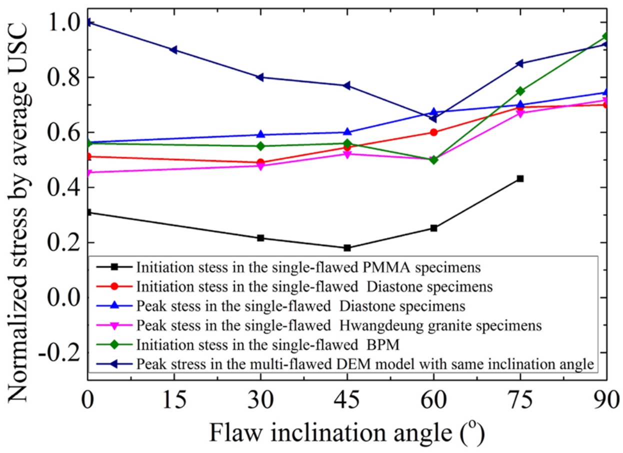


Fig. 17 Relationships between normalized stress by average UCS (uniaxial compressive strength) and flaw inclination angle from the published experimental data of PMMA, diastine and Hwangdeung granite specimens (Lee and Jeon 2011), numerical data of BPM (i.e., bonded-particle model) (Zhang and Wong 2012, 2013) and DEM (i.e., discrete element method) model (Vergara et al. 2016). *Note* this flaw inclination angle is the angle between inclination of flaw and horizontal direction

with the experiment result in macroscopic minor flawed specimens under uniaxial compressive stress. The range of crack angle which could result in crack initiation and rock failure of brittle rock decreases with increment of friction coefficient in compression. When the crack angle is beyond this range, wedge force would be compressive force and initial microcrack could not grow. When the crack angle is zero, wedge force would be zero and initial microcrack also could not grow.

Acknowledgments This work was supported by the National Natural Science Foundation of China (Grant nos. 51708016 and 51774018), the Postdoctoral Science Foundation of China (Grant no. 2017M610755), the National Basic Research Development Program (973 Program) of China (Grant no. 2015CB0578005), the scientific research fund for Beijing University of Civil Engineering and Architecture (Grant no. KYJJ2017006), and the Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT).

References

Ashby MF, Hallam SD (1986) The failure of brittle solids containing small cracks under compressive stress states. Acta Metall 34(3):497–510

Ashby MF, Sammis CG (1990) The damage mechanics of brittle solids in compression. Pure Appl Geophys 133(3):489–521

Baud P, Wong TF, Zhu W (2014) Effects of porosity and crack density on the compressive strength of rocks. Int J Rock Mech Min Sci 67(4):202–211

Bhat HS, Sammis CG, Rosakis AJ (2011) The micromechanics of Westerly granite at large compressive loads. Pure Appl Geophys 168(12):1–18

Bhat HS, Rosakis AJ, Sammis CG (2012) A micromechanics based constitutive model for brittle failure at high strain rates. J Appl Mech 79:031016

Bobet A, Einstein HH (1998) Numerical modeling of fracture coalescence in a model rock material. Int J Fract 92(3):221–252

Brace WF, Paulding B, Scholz C (1966) Dilatancy in the fracture of crystalline rocks. J Geophys Res 71:3939–3953

Brantut N, Baud P, Heap MJ, Meredith PG (2012) Micromechanics of brittle creep in rocks. J Geophys Res 117:B08412

Budiansky B, O'Connel RJ (1976) Elastic moduli of a cracked solid. Int J Solids Struct 12:81–97

Caia M, Kaisera PK, Tasakab Y, Maejima T, Morioka H, Minami M (2004) Generalized crack initiation and crack damage stress thresholds of brittle rock masses near underground excavations. Int J Rock Mech Min Sci 41:833–847

Cao RH, Cao P, Lin H, Pu CZ, Ou K (2016) Mechanical behavior of brittle rock-like specimens with pre-existing fissures under uniaxial loading: experimental studies and particle mechanics approach. Rock Mech Rock Eng 49(3):763–783

Chang SH, Lee CI (2004) Estimation of cracking and damage mechanisms in rock under triaxial compression by moment tensor analysis of acoustic emission. Int J Rock Mech Min Sci 41:1069–1086

Chen ZH, Tang CA, Huang RQ (1997) A double rock sample model for rockbursts. Int J Rock Mech Min Sci 34(6):991–1000

Clayton JD (2010) Deformation, fracture, and fragmentation in brittle geologic solids. Int J Fract 163:151–172

Cox SJD, Meredith PG (1993) Microcrack formation and material softening in rock measured by monitoring acoustic emissions. Int J Rock Mech Min Sci 30(1):11–24

Damjanac B, Fairhurst C (2010) Evidence for a long-term strength threshold in crystalline rock. Rock Mech Rock Eng 43(5):513–531

Diederichs MS, Kaiser PK, Eberhardt E (2004) Damage initiation and propagation in hard rock during tunneling and the influence of near-face stress rotation. Int J Rock Mech Min Sci 41:785–812

Fan LF, Wu ZJ, Wan Z, Gao JW (2017) Experimental investigation of thermal effects on dynamic behavior of granite. Appl Therm Eng 125:94–103

Germanovich LN, Salganik RL, Dyskin AV, Lee KK (1994) Mechanisms of brittle fracture of rock with pre-existing cracks in compression. Pure Appl Geophys 143:117–149

Hoek E, Bieniawski ZT (1984) Brittle fracture propagation in rock under compression. Int J Fract 26:276–294

Lajtai EZ (1971) A theoretical and experimental evaluation of the Griffith theory of brittle fracture. Tectonophysics 11(2):129–156

Lajtai EZ (1974) Brittle fracture in compression. Int J Fract 10(4):525–536

Lajtai EZ (1998) Microscopic fracture processes in a granite. Rock Mech Rock Eng 31(4):237–250

Lan HX, Martin CD, Hu B (2010) Effect of heterogeneity of brittle rock on micromechanical extensile behavior during compression loading. J Geophys Res 115:B01202

Lee H, Jeon S (2011) An experimental and numerical study of fracture coalescence in pre-cracked specimens under uniaxial compression. Int J Solids Struct 48:979–999

Li HQ, Wong LNY (2012) Influence of flaw inclination angle and loading condition on crack initiation and propagation. Int J Solids Struct 49:2482–2499

Martin CD (1997) Seventeenth Canadian Geotechnical Colloquium: the effect of cohesion loss and stress path on brittle rock strength. Can Geotech J 34:698–725

Martin CD, Chandler NA (1994) The progressive fracture of Lac du Bonnet granite. Int J Rock Mech Min Sci Geomech Abstr 31(6):643–659

Mavko GM, Nur A (1978) The effect of nonelliptical cracks on the compressibility of rocks. J Geophys Res 83(B9):4459–4468

Morgan S, Johnson CA, Einstein HH (2013) Cracking processes in Barre granite: fracture process zones and crack coalescence. Int J Fract 180:177–204

Park CH, Bobet A (2010) Crack initiation, propagation and coalescence from frictional flaws in uniaxial compression. Eng Fract Mech 77:2727–2748

Sagong M, Bobet A (2002) Coalescence of multiple flaws in a rock-model material in uniaxial compression. Int J Rock Mech Min Sci 39(2):229–241

Silva BGD, Einstein HH (2013) Modeling of crack initiation, propagation and coalescence in rocks. Int J Fract 182:167–186

Singh UK, Digby PJ (1989) A continuum damage model for simulation of the progressive failure of brittle rocks. Int J Solids Struct 25(6):647–663

Vergara MR, Jan MVS, Lorig L (2016) Numerical model for the study of the strength and failure modes of rock containing non-persistent joints. Rock Mech Rock Eng 49(4):1211–1226

Wan LH, Cao P, Huang YH, Wang YX (2010) Study of subcritical crack growth of rocks and threshold values in different environments. Chin J Rock Soil Mech 31(9):2737–2742

Wang B, Zhu JB, Wu AQ, Hu JM, Xiong ZM (2008) Experimental study on mechanical properties of jinping marble under loading and unloading stress paths. Chin J Rock Mech Eng 27(10):2138–2145

Weibull W (1951) A statistical distribution function of wide applicability. J Appl Mech 18:293–297

Wong LNY, Li HQ (2013) Numerical study on coalescence of two pre-existing coplanar flaws in rock. Int J Solids Struct 50:3685–3706

Wong TF, Wong RHC, Chau KT, Tang CA (2006) Microcrack statistics, Weibull distribution and micromechanical modeling of compressive failure in rock. Mech Mater 38:664–681

Wu Z, Fan L, Liu Q, Ma G (2017) Micro-mechanical modeling of the macro-mechanical response and fracture behavior of rock using the numerical manifold method. Eng Geol 225(20):49–60

Yang SQ, Jing HW (2011) Strength failure and crack coalescence behavior of brittle sandstone samples containing a single fissure under uniaxial compression. Int J Fract 168(2):227–250

Zhang XP, Wong LNY (2012) Cracking processes in rock-like material containing a single flaw under uniaxial compression: a numerical study based on parallel bonded-particle model approach. Rock Mech Rock Eng 45:711–737

Zhang XP, Wong LNY (2013) Crack initiation, propagation and coalescence in rock-like material containing two flaws: a numerical study based on bonded-particle model approach. Rock Mech Rock Eng 46(5):1001–1021

Zhou XP, Zhang YX, Ha QL, Zhu KS (2008) Micromechanical modelling of the complete stress- strain relationship for crack weakened rock subjected to compressive loading. Rock Mech Rock Eng 41(5):747–769

Zhou X, Fan L, Wu Z (2017) Effects of microfracture on wave propagation through rock mass. Int J Geomech 17(9):04017072

Acta Geophysica (2018) 66:153–165
https://doi.org/10.1007/s11600-018-0117-2

RESEARCH ARTICLE - SOLID EARTH SCIENCES

Characterization of site conditions for selected seismic stations in eastern part of Romania

B. Grecu[1] · **B. Zaharia**[1] · **M. Diaconescu**[1] · **A. Bala**[1] · **E. Nastase**[1] · **E. Constantinescu**[1] · **D. Tataru**[1]

Received: 6 October 2017 / Accepted: 3 February 2018 / Published online: 15 February 2018

Abstract

Strong motion data are essential for seismic hazard assessment. To correctly understand and use this kind of data is necessary to have a good knowledge of local site conditions. Romania has one of the largest strong motion networks in Europe with 134 real-time stations. In this work, we aim to do a comprehensive site characterization for eight of these stations located in the eastern part of Romania. We make use of a various seismological dataset and we perform ambient noise and earthquake-based investigations to estimate the background noise level, the resonance frequencies and amplification of each site. We also derive the Vs_{30} parameter from the surface shear-wave velocity profiles obtained through the inversion of the Rayleigh waves recorded in active seismic measurements. Our analyses indicate similar results for seven stations: high noise levels for frequencies larger than 1 Hz, well defined fundamental resonance at low frequencies (0.15–0.29 Hz), moderate amplification levels (up to 4 units) for frequencies between 0.15 and 5–7 Hz and same soil class (type C) according to the estimated Vs_{30} and Eurocode 8. In contrast, the eighth station for which the soil class is evaluated of type B exhibits a very good noise level for a wide range of frequencies (0.01–20 Hz), a broader fundamental resonance at high frequencies ($\sim$ 8 Hz) and a flat amplification curve between 0.1 and 3–4 Hz.

Keywords Ambient seismic noise · Resonance frequency · Soil amplification · Rayleigh waves · Vs_{30}

Introduction

Romania is considered one of the most vulnerable countries in Europe at strong earthquakes. Its seismicity is dominated by the Vrancea intermediate-depth earthquakes. These earthquakes are generated within a narrow focal volume located beneath the region of the maximum curvature of the Eastern Carpathians, at depths between 60 and approximately 200 km. The strong earthquakes occurred in this area produce significant damage over an extended area elongated in the NE–SW direction, predominantly in the extra-Carpathian area. The crustal seismicity is not as substantial as the subcrustal one, and it is generated in several seismogenic zones across the country (Radulian et al. 2000).

In the last four decades, 28 moderate-to-strong earthquakes with magnitudes (M_w) between 5.0 and 7.4 occurred on the Romanian territory, of which only 5 in the crustal domain (from Romanian catalog ROMPLUS—www.infp.ro). To monitor and adequately record the strong ground motion during such kind of events, the National Institute for Earth Physics (NIEP) has developed the Romanian Strong-Motion Network (RSMN). At present, RSMN consists of 134 stations installed in different environments (free-field, buildings, vaults) all over the country. The RSMN stations record the ground motion continuously and using a real-time communication system the data is sent to the Romanian National Data Center (RONDC) in Magurele, Romania.

Strong motion data are very important for seismic hazard studies as they are used as input for the derivation of the ground-motion prediction equations (GMPEs). The quality and the proper understanding of the recorded strong ground motion depend on the performance of the seismic equipment as well as on a complete characterization of the site. This characterization includes, besides the geological and technical information about the location and equipment, a good knowledge of the local site conditions as they

✉ B. Grecu
bgrecu@infp.ro

[1] National Institute for Earth Physics, Măgurele, Romania

could change the frequency content of the ground motion, amplify it and extend its duration.

In this paper, we investigate the site conditions at 8 RSMN stations located in the eastern part of Romania (Fig. 1). The motivation of the selection of the stations is twofold. Firstly, they cover now an area which was poorly monitored before 2013 when an unusual seismic swarm occurred NW to the city of Galati. Secondly, the distribution of the accelerations recorded by RSMN stations shows, after each significant earthquake with magnitude $M_{\mathrm{L}} \geq 4.0$, higher values for most of the stations used in the study and located in the vicinity of the Galati city than for almost all other stations. This behavior is seen in Fig. 2 which portrays the distribution of the maximum accelerations recorded by RSMN stations during the 2nd of August 2017 Vrancea earthquake ($M_{\mathrm{L}} = 4.9$, $H = 133$ km) and the 16th of August 2017 Galati earthquake ($M_{\mathrm{L}} = 4.0$, $H = 11$ km).

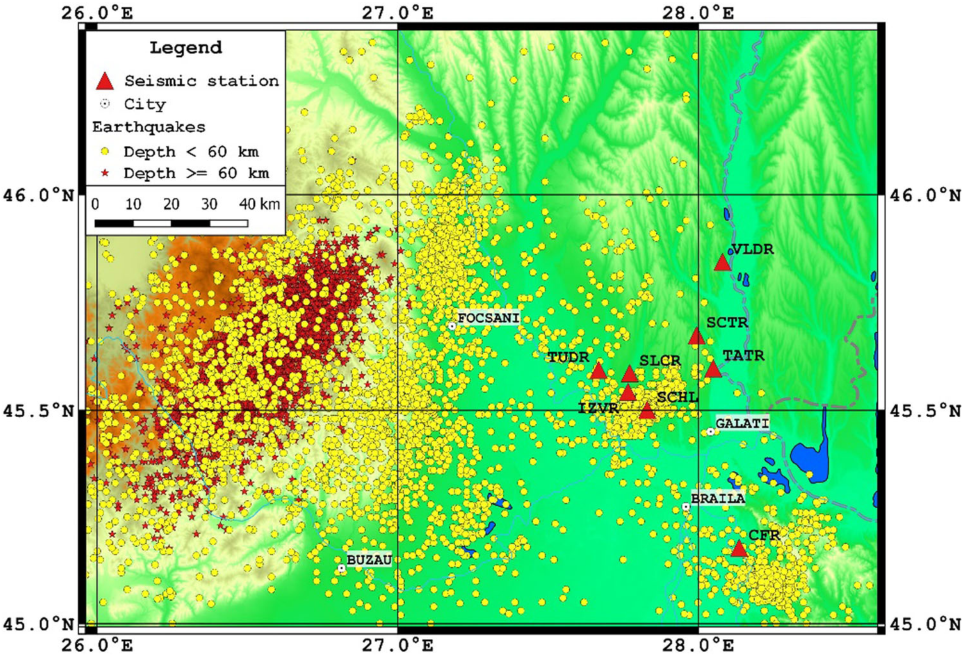
The investigations we performed in this study include characterization of the noise level at the stations, estimation of the fundamental characteristics of the ground motion using noise and earthquake data and determination of the shallow velocity structure and the Vs_{30} parameter.

Seismotectonic setting

The Eastern part of Romania has a particular tectonic environment which consists of several major tectonic units (East-European Platform, Scythian Platform, Moesian Platform and North Dobrogea tectonic unit) and major faults which bound these tectonic units: Peceneaga–Camena fault—the tectonic limit between Moesian Platform and North Dobrogea, Sf. Gheorghe fault—the boundary between North Dobrogea and Scythian Platform and Trotuş fault—the limit between Scythian Platform and East European Platform. Three of the investigated sites (Scanteiesti (SCTR), Tatarca (TATR) and Vladesti (VLDR)) are located on Scythian Platform while all other stations are on North Dobrogean Promontory (Fig. 3). The internal structure of the Scythian platform is less known than that of the East-European Platform (Matenco et al. 2003), due to thicker Tertiary sediments in the Bârlad Depression and to underthrusting below the Eastern Carpathians nappe pile. Carcaliu (CFR) site is located on the Macin nappe area of the North Dobrogean tectonic unit. The North Dobrogea tectonic unit is situated between the Scythian and Moesian Platforms, and it is composed of a complex deformed Hercynian basement and a Triassic–Cretaceous sedimentary cover, unequally developed (Ionesi 1989). West of the Danube, the basement and Mesozoic sediments are covered by a succession of Tertiary deposits, forming the North Dobrogea Promontory.

Before 2013, the seismic activity in the area was characterized by the occurrence of dispersed earthquakes of small-to-moderate size, the largest event with a moment magnitude of 4.2 being recorded on the 11th of September 1980 (ROMPLUS catalog Oncescu et al. 1999). In August 2013, an uncommon seismic swarm started in the Galati area. The activity during the swarm period was characterized by a large number of events (940) with local magnitudes from 0.1 to 4.0 occurred in the time interval from

Fig. 1 Map with the location of the investigated sites plotted together with the seismicity (red stars—intermediate-depth earthquakes, yellow circles—crustal events)



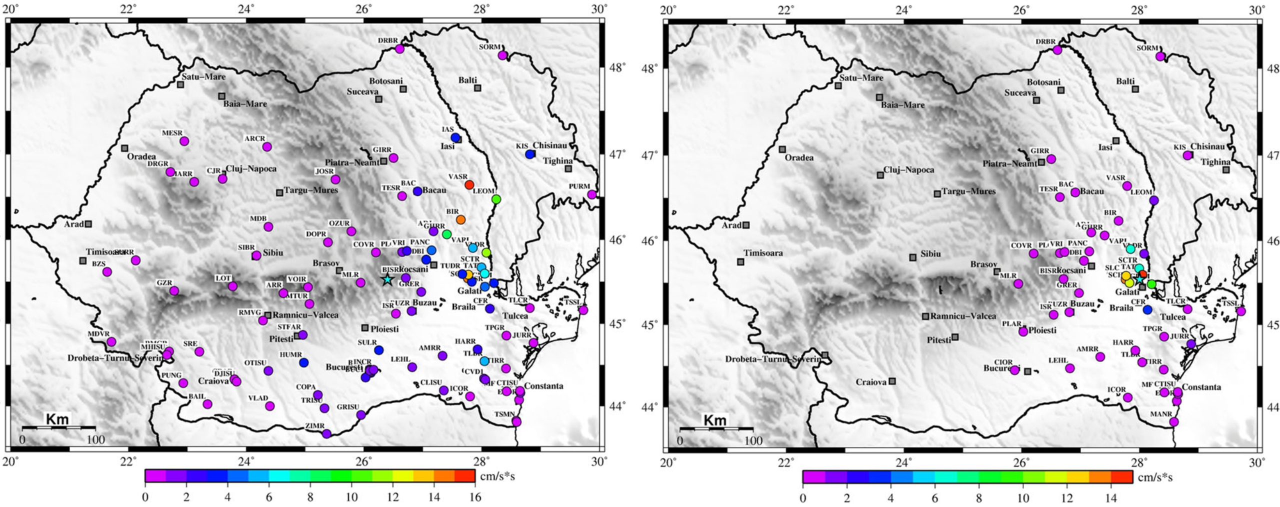


Fig. 2 The distribution of the maximum acceleration recorded during two local events: (left) the 2nd of August 2017 Vrancea intermediate-depth earthquake ($M_L = 4.9$) and (right) the 16th of August 2017 crustal earthquake ($M_L = 4.0$). Blue star—earthquakes epicenter

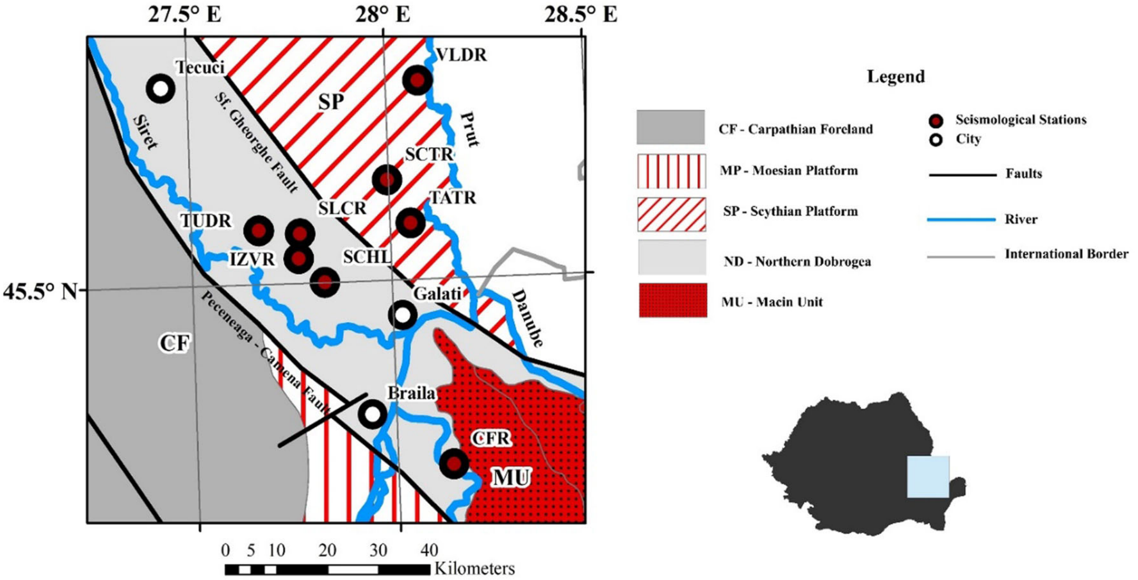


Fig. 3 Tectonic map of the investigated area (after Sandulescu 1984). Seismic stations are also depicted with red dots

15th August to 5th November (Popa et al. 2016). The earthquakes epicenters were located between the Sf. Gheorghe fault and Peceneaga–Camena crustal fault.

The region located to the SE of Galati city is characterized by the seismic activity clustered in the Predobrogean seismogenic zone (Radulian et al. 2000) and also by numerous artificial events generated in many quarries present in the area (Ghica et al. 2016).

Dataset description and investigations performed

The RSMN operated by NIEP in the study area was sparse before 2013, with only one station (CFR) existing to the SE of Galati city. The situation changed starting with spring 2013, when three more stations were deployed, one before and two during the Galati swarm. The network was upgraded with a new station in 2014 and two more in the middle of 2015 (Fig. 4). All sites are equipped with both broadband velocity sensors and accelerometers. The data used in this study are of three types: seismic noise, earthquake and active seismic data.

Seismic noise data

We used the seismic noise data with two purposes: first, to characterize the noise level at the stations and second, to estimate the fundamental resonance frequency of the sites. In the first case, we assessed the level of seismic noise by computing the Probability Density Functions (PDFs) for a large number of Power Spectral Densities (PSDs) (McNamara and Buland 2004). Many studies took advantage of the PDF versatility and used it to evaluate seismic stations performance and compute the noise level at the seismic sites (McNamara and Buland 2004; Diaz et al. 2010; Evangelidis and Melis 2012; Grecu et al. 2012). For this analysis, we used the entire dataset available at each station location.

In the late 80's and the 90's, numerous studies focused on local site investigations and used the ratio between the horizontal and vertical spectral components of seismic noise (hereinafter HVNR—"Horizontal to Vertical Noise Ratio") to characterize the seismic response of the subsoil. The peak of the HVNR curve is often associated with the fundamental resonance frequency of the site, and its amplitude and sharpness are related to the shear-wave impedance contrast in the subsoil (Nakamura 1989; Bard 1999). The HVNR analysis was performed at each station using 10 days of continuous data, recorded from 1st to 10th of January 2017 (the period was randomly selected) using the velocity sensors. The long duration of the recordings allowed us to use windows of 100 s length and, therefore, to go in our investigations to lower frequencies (up to 0.1 Hz) (SESAME project-2005). We used an automatic procedure based on an anti-triggering algorithm to avoid transient noise and to select stationary time windows. Figure 4 shows the number of windows used in HVRN noise analysis for each station. The data selection and processing were done using Geopsy software (www.geopsy.org).

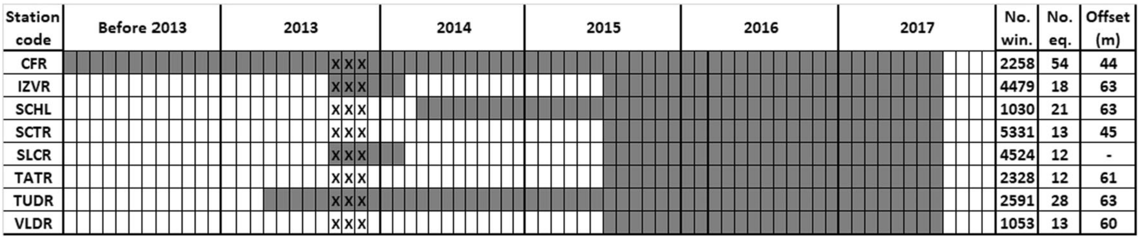
Earthquake data

In this study, we used earthquake data to obtain information on the resonance frequencies (Lermo and Chávez-García 1993; Field and Jacobs 1995) as well as on local amplification of the investigated sites. The used method (hereinafter HVSR—"Horizontal to Spectral Vertical Ratio") is very similar to the HVNR method used for noise data. Figure 5 portrays the methodology we followed. The HVSR analyses were performed on 60 s time windows, starting from the P-waves onset. This window length allowed us to include the S-wave train which usually contains the most energetic part of the record as well as the coda wave of which spectral shape depends only on the local heterogeneities in the crust (Philips and Aki 1986). Grecu et al. (2011) used both S-wave and coda wave to investigate the site effects at several stations installed in eastern part of Romania during a temporary seismic experiment. They found no significant differences regarding the resonant frequencies of the HVSRs computed for the two types of waves, while the level of the amplification for S-wave is slightly higher than for coda waves.

The HVSR analysis was performed using earthquake data recorded by the strong motion sensor located at each station. We used 54 local earthquakes with $M_w \geq 4.0$ (from Romanian ROMPLUS catalog) recorded between 2010 and 2017. Since the stations have been installed in various time periods and consequently have different recording periods, the number of earthquakes recorded is different from one station to another. In Fig. 4 are shown the number of earthquakes used in the HVSR analysis for each site.

Active seismic data

Active seismic measurements have been performed at all station sites, except for station Slobozia Conachi (SLCR) for which no seismic survey was possible. These measures aimed to record surface waves generated by an active source (sledgehammer) and invert their dispersive properties for the determination of the V_S vertical profile and, the Vs_{30} parameter. The dataset was acquired using a 3-C



Station code	Before 2013	2013	2014	2015	2016	2017	No. win.	No. eq.	Offset (m)
CFR		X X X					2258	54	44
IZVR		X X X					4479	18	63
SCHL		X X X					1030	21	63
SCTR		X X X					5331	13	45
SLCR		X X X					4524	12	-
TATR		X X X					2328	12	61
TUDR		X X X					2591	28	63
VLDR		X X X					1053	13	60

Fig. 4 The data availability, the number of windows (No. win.) and earthquakes (No. eq.) used for HVNR and HVSR analyses and the profile lengths (offset) for each station. With X is marked the period of the Galati swarm

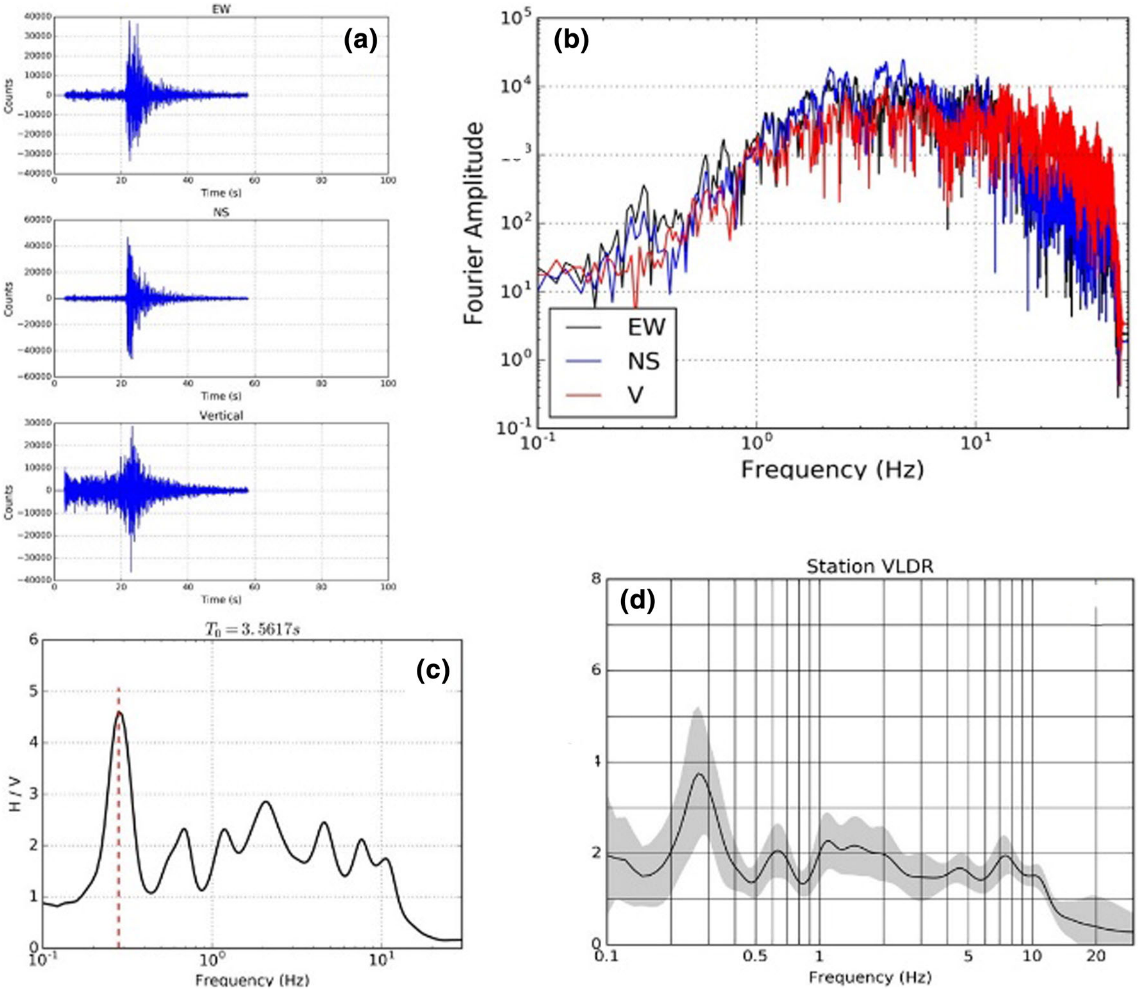


Fig. 5 The methodology used for estimating the local amplification at a given station: **a** selection of earthquake recordings and cut off of 60 s window length (the waveforms are from a subcrustal earthquake with $M_w = 4.6$ occurred in Vrancea area on 2nd of August 2017) **b** computation of the Fourier spectra for Z, NS and EW components for the selected earthquake **c** computation of the spectral ratio for the chosen earthquake by dividing the smoothed average spectrum of the horizontal components to the vertical one; the spectra are smoothed using the Konno and Ohmachi (1998) recording window ($b = 20$) and the horizontal average spectrum is computed as geometric mean of the two horizontal components (**d**) the HVSR curve obtained by averaging the results obtained for each event

geophone with a natural frequency of 2 Hz. The use of 3-C geophone allowed obtaining multichannel data consisting of vertical and radial components of Rayleigh waves (the geophone was oriented such that its horizontal NS axis was along the profile). No Love waves were used in the analysis since their recording requires a different setup for generating them (wooden beam and horizontal force) which was not available during the measurements. Because of logistic limitations, the maximum profile length (offset—the distance between the active source and the recording sensor) was 63 m. In Fig. 4 are shown the offsets of the surveys at each station. The profile length limits the maximum penetration depth to about two-third of the adopted offset if some conditions are met (Dal Moro 2014).

Each survey consisted of two sets of measurements. For each group, we generated four shots, and we stacked the

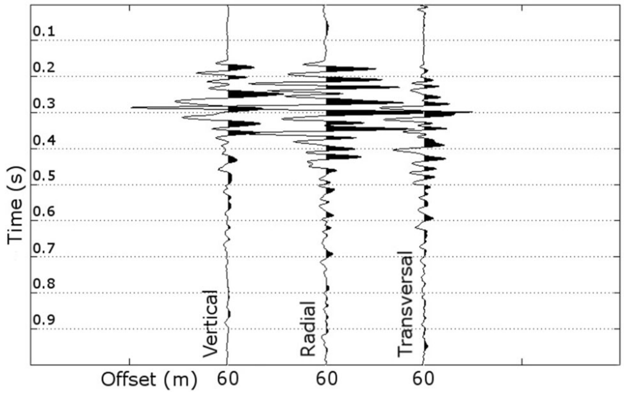


Fig. 6 Stacked traces obtained at station VLDR

resulting traces to attenuate the incoherent noise and obtain the final record. Figure 6 shows the stacked field data collected at station VLDR. In our analysis, we used the best measurements of the two sets.

The analysis of the surface waves (Rayleigh type) follows the approach described by Dal Moro (2014). In this method, the signals recorded in the field are transformed into the frequency–velocity domain to yield the velocity spectrum (VS) which is further used in an inversion procedure to retrieve the shallow velocity structure. The advantage of this method compared to classical one's lies in the manner the VS is used for the inversion. Thus, in this approach, the whole VS is used while in the traditional method the modal dispersion curves are picked subjectively by the user and used as input for the inversion.

Results

Background noise level

To estimate the background noise level at each station, we used the statistical mode of the PDF computed for the vertical component, as this curve corresponds to the highest probability noise level of a given site (McNamara and Buland, 2004). Figure 7a shows an example of the PDF computed for station CFR. The high-probability region corresponds to the power values associated with the background seismic noise. This domain is relatively close to the Peterson's (1993) new low noise model (NLNM) for periods between 0.1 s and 179 s, indicating thus a good background noise level for the station. An increase of almost 30 dB of the noise level is observed for periods smaller than 0.1 s. The noise level at these periods is strongly affected by anthropogenic noise sources, which in this case, are related to the human activities carried out in a rural household located very close (about 15–20 m) to the station. Figure 7b portrays the stations PDF mode noise levels plotted together with three reference noise models: the Peterson's New High Noise Model (NHNM), New Low Noise Model (NLNM) and the Romanian Noise Model (RONM) (Grecu et al. 2017). All stations show high noise levels, very close to or even exceeding the NHNM, for periods lower than 1 s and larger than 20 s, except for CFR site. The variation of the noise levels is much smaller in the microseismic band (2–20 s) where noise levels are related mainly to the energy released by oceanic waves (Stutzmann et al. 2009). The higher noise levels observed for IZVR, SCHL, SCTR, SLCR, TATR and TUDR sites at small periods (< 1 s) are related to the location of these stations. They are installed on soft and thick sediments within villages or within a monastery (TUDR station) where human activities are more significant than for CFR station. At periods larger than 20 s, the difference between the noise level at CFR and the others is given mainly by the instrumentation and the proper thermal isolation of the sensor of station CFR.

Site condition analysis

Detailed analysis of the HVNR results obtained at seismic stations indicates two types of HVNR curves: one dominated by a clear peak at low frequencies corresponding to the fundamental resonance peak and one with a broad peak at high frequencies. The former is characteristic for seven stations (IZVR, SCHL, SCTR, SLCR, TATR, TUDR, VLDR) while the latter is observed at just one site (CFR) (Fig. 8). In the first case, the frequency of the fundamental resonance peak increases from 0.15 Hz at TUDR station towards 0.29 Hz at TATR station while its amplitude varies from 3.9 observed at station SCHL to 7 at station VLDR. A secondary peak, much broader, is also seen at

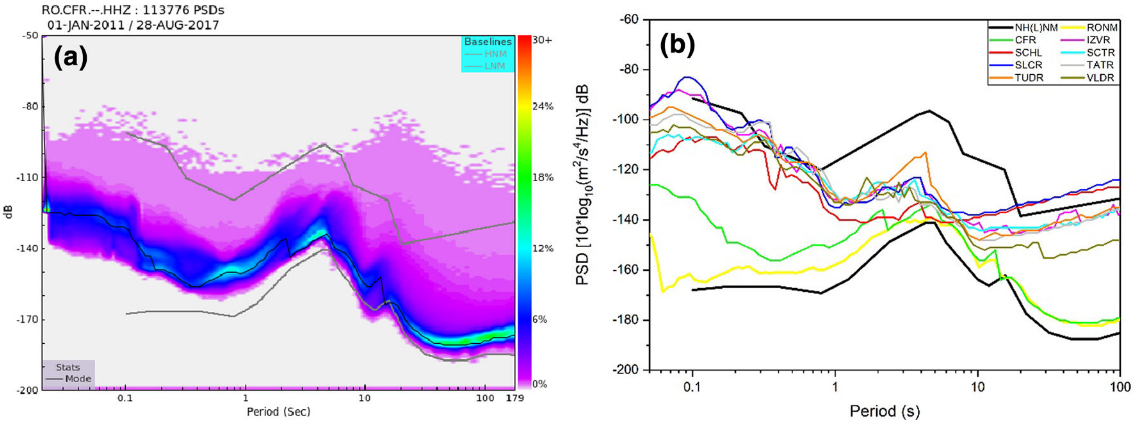


Fig. 7 **a** PDF for station CFR, the black line being the mode of the PDF and the gray lines correspond to the NLNM and NHNM curves **b** stations mode plotted together with three reference models: NHNM, NLNM and RONM

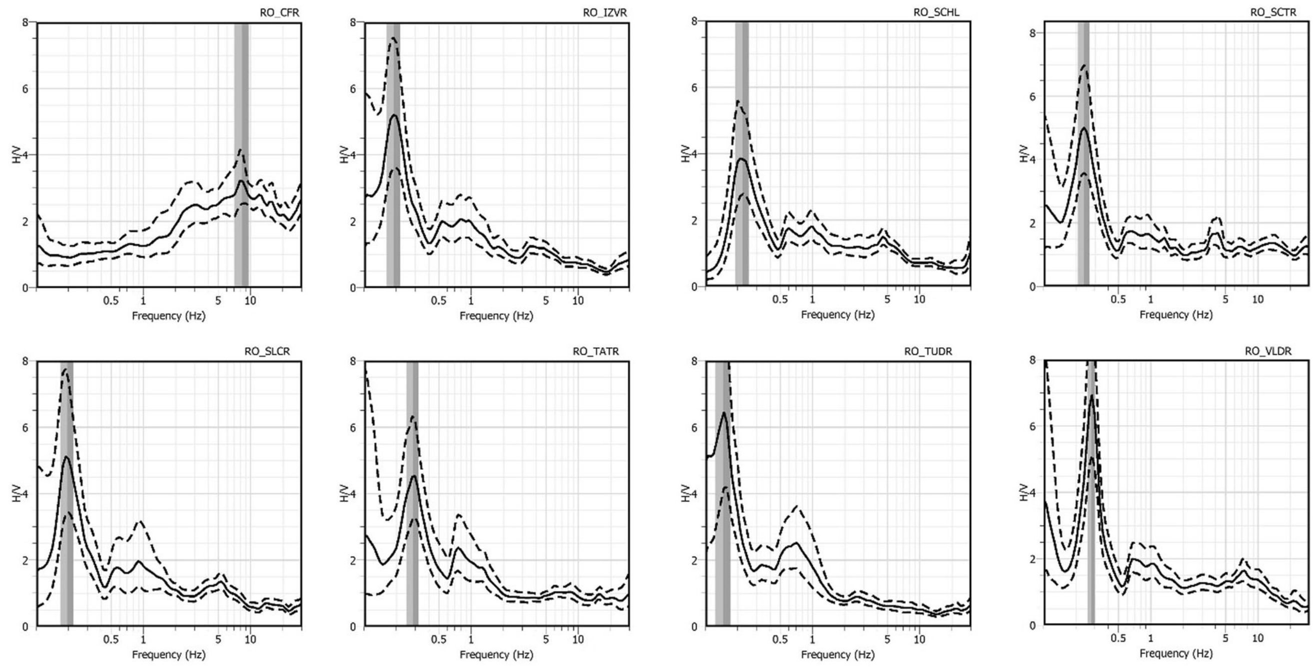


Fig. 8 The HVNR results for all stations

frequencies between 0.5 and 1 Hz. This peak is well defined and its amplitude exceeds 2 only for stations TATR and TUDR. For station CFR, the HVNR curve shows nearly flat amplitude below 2 for frequencies smaller than 2 Hz. For frequencies larger than 2 Hz, an increase in the amplitude of the HVNR curve is observed and its maximum of 3.2 units is reached at about 8 Hz. The resonance frequency of each site depends on the nature and thickness of the layers' underneath. All the stations located to the N and NE of Galati city are installed on relatively thick sediments. For example, TUDR station is located on the external margin of Carpathian foredeep, on 3 km depth of sedimentary rocks; the first 1.3 km are soft rocks of Neogene formations. Underneath lies the basement of Pre-Dobrogean depression with V_S velocities of about 3300 m/s (Raileanu, Annual report 2006).

This fact is reflected in the low-frequency fundamental resonance peak of the HVNR curve. The difference in frequencies of the fundamental resonance peak observed at these stations is related to the variations in the sediment thickness (Ibs-von Seht and Wohlenberg 1999) beneath these stations.

For station CFR, which is located on North Dobrogean Orogen, on old rocks of Paleozoic origin, the Paleozoic sediments overlie a Hercynian basement consisting of crystalline schists of Proterozoic age (Balan et al. 2014). The increase at high frequencies of the HVNR amplitude at CFR station can be related to the presence of a thin weathered soil of just a few meters, overlying the hard rock.

The HVNR amplitude azimuthal variations were also investigated (Fig. 9). For all the stations, except CFR station, the amplitude corresponding to the low fundamental frequency varies very little with the azimuth. On the contrary, the HVNR amplitude obtained for station CFR shows a high dependency on the azimuth at frequencies between 8 and 9 Hz. Large azimuthal variations could indicate either an interaction with a nearby structure or complex site effects (topographical irregularities, rock degradation), the latter being the most likely for station CFR.

Figure 10 shows the HVSR curves obtained using earthquake data. On the one hand, the results confirm to some extent the findings obtained from noise data, and on the other hand, they show some features that are not present in HVNR curves.

For all stations, the fundamental resonance frequency highlighted by the HVNR technique is also present in the HVSR curves, but only for stations CFR, IZVR, SCHL, SCTR and VLDR its amplitude is larger or at least similar with the other resonances. The resonance peak observed in the frequency domain 0.5–1 Hz has comparable amplitude with the one identified in HVNR curves, except for station TATR for which a value of 3 is reached. Another similarity between HVNR and HVSR curves is well recognized for station CFR. Both spectral ratios are relatively flat for a wide frequency band (from 0.1 to 2 Hz) and show amplitudes up to 3, for frequencies between 5 and 10 Hz. The main difference between the HVNR and HVSR curves is observed in the frequency band 1–7 Hz. In this range, all

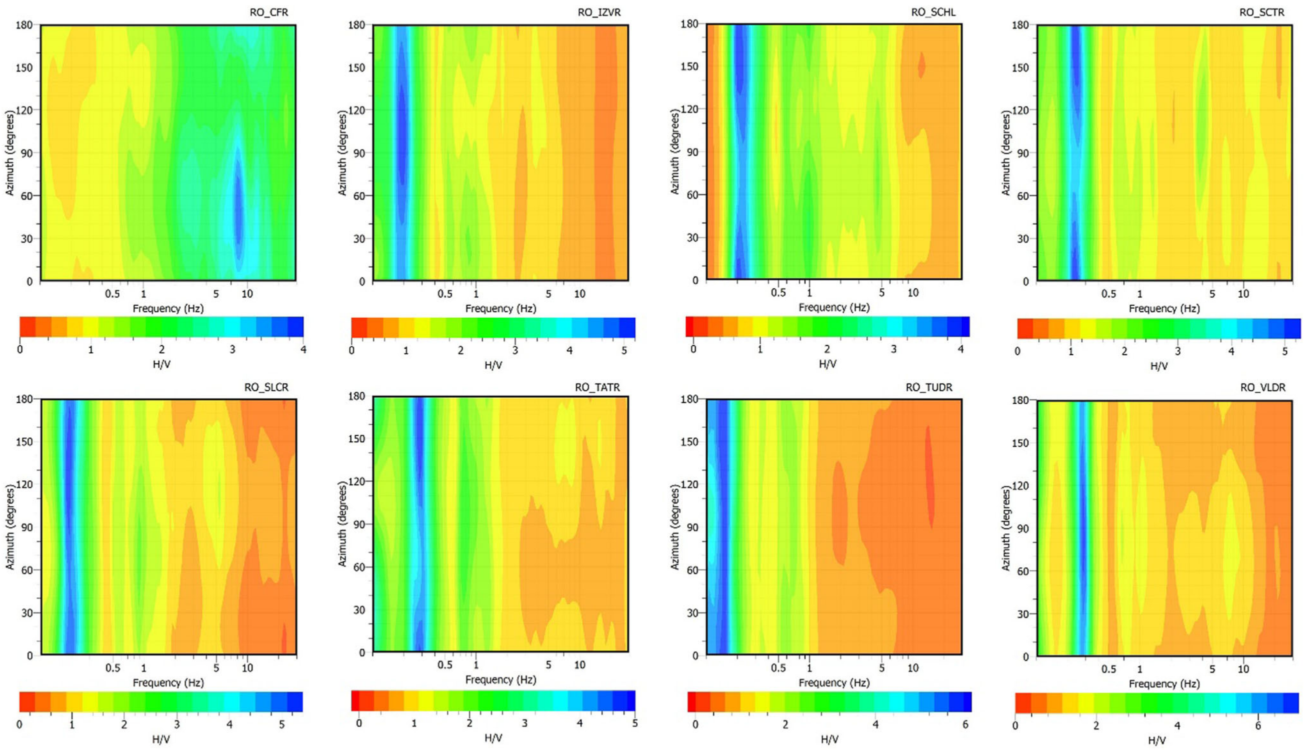


Fig. 9 The variation of the HVNR amplitude with azimuth

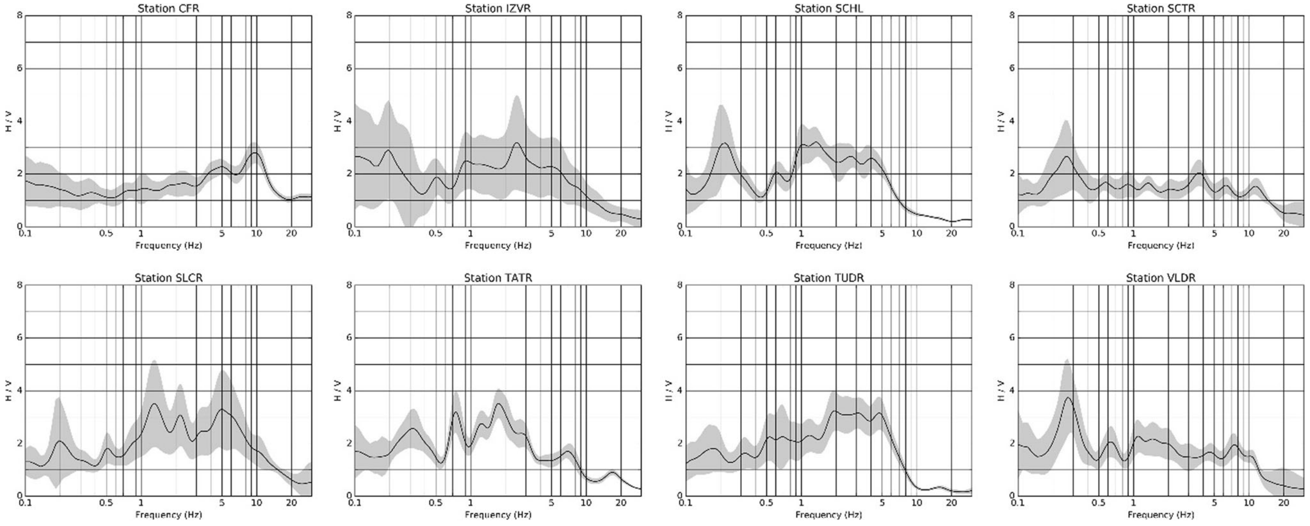


Fig. 10 HVSR results for all stations

the stations located in the Galati area exhibit amplifications around 3.

Shallow velocity structure, Vs_{30}

The derivation of the shallow velocity structure underneath each site and estimation of Vs_{30} for soil class evaluation was done using the tools of the HoliSurface software (www.winmasw.com) and consisted mainly of three steps. In the first step, we transformed the stacked trace from the offset-time domain into the frequency-velocity domain, and we obtained the VS (Fig. 11a). Next, we used different test velocity models to compute several synthetic VSs. The software uses the modal summation technique to calculate the synthetics. In this step, we modified the parameters of the test velocity models (shear-wave velocity and depth)

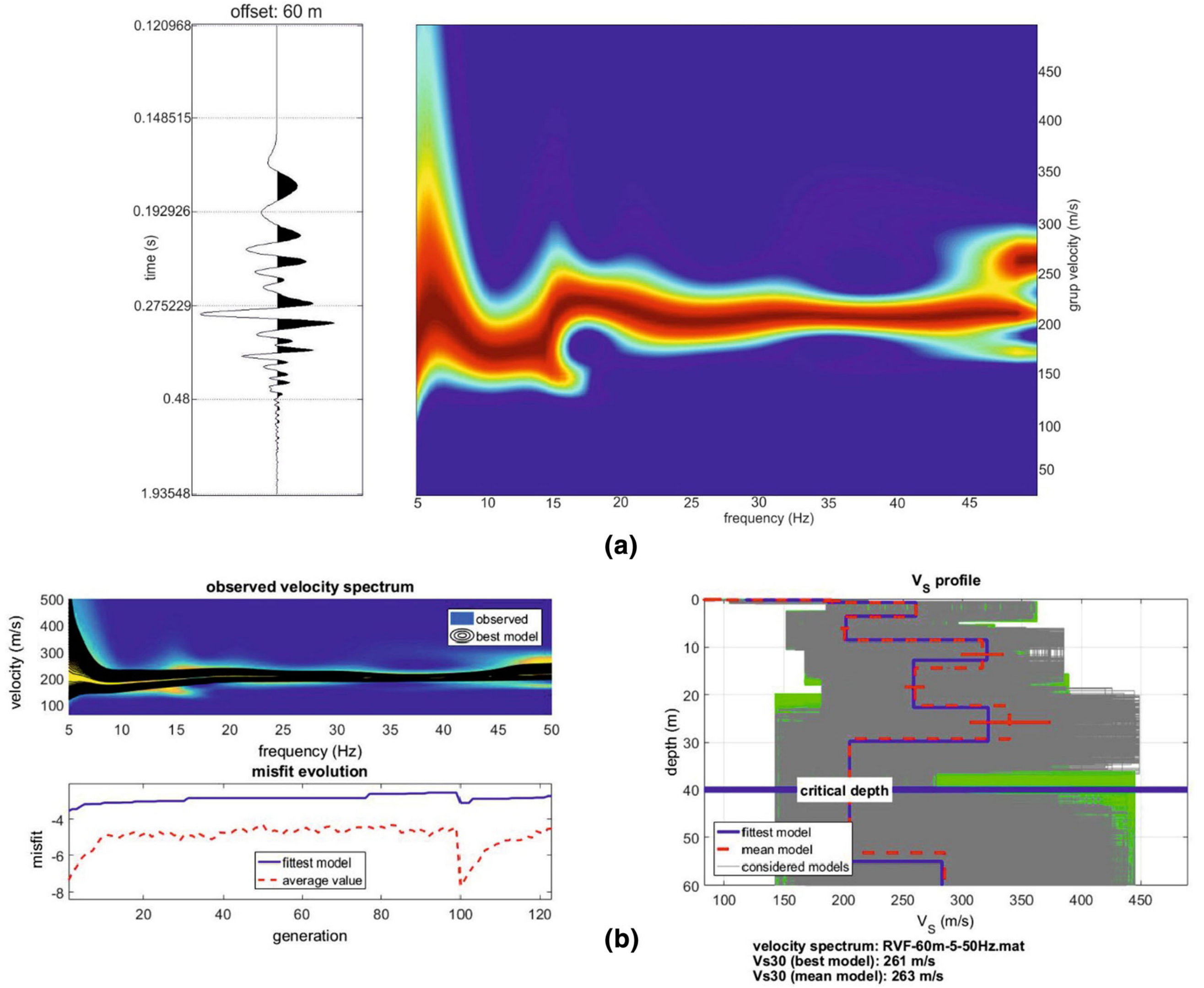


Fig. 11 **a** The radial component obtained in the field at station VLDR after the summation procedure and its observed velocity spectrum (**b**) inversion results (upper left—the fit between the VS of the best model and the observed FVS, right—velocity profiles for the best and mean models with blue line and red dotted line, respectively, and all models with gray lines) and the misfit evolution during the inversion procedure (lower left)

until we reached an acceptable fit between the observed VS and the computed VS. Then, we chose the final test model as starting model in the inversion performed in the last step. The inversion scheme is based on a genetic algorithm (Dal Moro et al. 2007) and, in summary, tries to minimize the misfit between the observed VS and the computed VS by changing the starting model iteratively. Finally, the Vs_{30} is automatically calculated from the derived shear-wave profile, and the soil class is ascertained. This final step is exemplified in Fig. 11b for station VLDR. The figure portrays the fit between the best FVS and observed FVS (upper left panel), the misfit evolution (lower left panel) and the retrieved Vs profiles for the best and mean model (right panel). In the case of station VLDR, the Vs_{30} is 261 and 263 m/s for the best model and mean model, respectively, while the soil class according to EC8 is C.

Since in our study we considered only vertical-impact source we could perform the inversion only for the vertical and radial components of the Rayleigh wave. In Fig. 12 are compared the shear-wave velocity profiles derived from the inversion of the vertical and radial traces recorded at station TATR. Some differences can be observed between the two profiles. These differences are also passed to the Vs_{30} values estimated from the two datasets. The Vs_{30} derived from the vertical component of the Rayleigh wave has a value of 267 m/s while for the radial component is 239 m/s. However, the differences are not large and the soil classification is C in both cases.

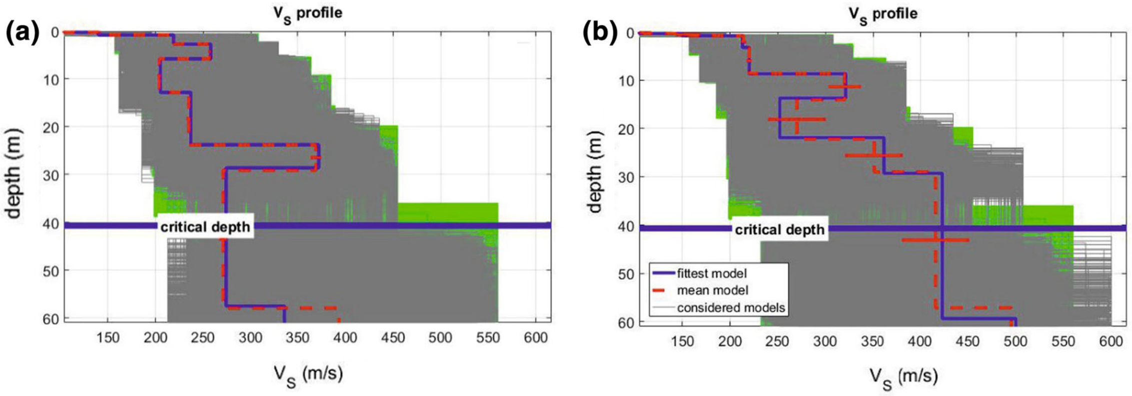


Fig. 12 V_s profiles obtained from the inversion of the vertical (**a**) and radial (**b**) component of Rayleigh wave (blue and red lines represent the best and mean models, respectively; gray lines—all models)

Considering both components (vertical and radial) of the Rayleigh wave recorded at each station, we have also performed a joint analysis of the two datasets. This common inversion has the advantage of better constraining the solution by excluding the models that cannot explain both datasets. The joint inversion has a different way of computing and treating the misfits. In this case, the misfits of each dataset are kept separately, and a ranking procedure evaluates their values based on the Pareto dominance criterion (Dal Moro and Ferigo 2011). Finally, the Pareto bi-objective space allows to choose the best model, and the symmetry of the Pareto front models constitutes an index of the general consistency of the whole inversion process (Dal Moro (2008); Dal Moro and Ferigo (2011)). The results using the joint inversion, for the TATR site, can be seen in Fig. 13. In this case, the computed value for Vs_{30} is 263 m/s. If we compare this value with those obtained from the inversion of just one dataset we notice that the differences are relatively small and the soil class for TATR site is still C. Table 1 gives a summary of the Vs_{30} values obtained from the inversion of the Rayleigh wave.

Discussion and concluding remarks

In this study, we applied different techniques and used various datasets (noise, earthquake and seismic active data) with the aim of investigating the site conditions at eight permanent stations of the Romanian Strong Motion Network.

The noise data were used with two purposes: first, to determine the noise level at the stations and second, to estimate the fundamental resonance frequency of each site. The investigation of the noise level at individual stations, which was performed using the PDFs, outlined only one station (CFR—located on North Dobrogean orogen) with a good noise level over a wide period domain (0.05–100 s). The other seven stations (IZVR, SCHL, SCTR, SLCR, TATR, TUDR, VLDR) show noise levels much higher, in particular for periods smaller than 1 s and periods larger than 10 s. In the first case, the noise is generated mainly by anthropic activities, and the noise level depends on the sources and the distance to the sources. All seven stations are installed within villages, and the different human activities in each village are reflected in different noise levels. It has been shown (Grecu et al. 2012, 2016) that the noise level in the microseismic domain (2–10 s) depends, among other things, on the thickness of the sediments underneath the station. All seven stations (IZVR, SCHL, SCTR, SLCR, TATR, TUDR, and VLDR) are located on much thicker sediments of Neogene origin, than station CFR which is located on Paleozoic sediments. The latter is more compact and with high seismic velocity up until surface; this is well reflected in the lower noise level observed in the microseismic domain. At larger periods (> 10 s), the higher noise levels are a consequence of the instrumentation and installation of the sensors.

The noise single station analysis, performed using the HVNR methodology, shows a distinct fundamental resonance peak at low frequencies for stations IZVR, SCHL, SCTR, SLCR, TATR, TUDR, VLDR. Its large amplitude and the absence of any azimuthal dependency are a strong indicator of the existence of a high-velocity contrast between the thick stack of sediments of about 3000 m and the bedrock at all these sites. This contrast is documented at 2600/3650 m/s near VLDR, and at 2800/3300 m/s at TUDR station (Raileanu, Annual report 2006). From Figs. 4 and 8 it can be observed that the resonance frequency of these stations increases from 0.15 to 0.28–0.29 Hz from the most south-western station (TUDR) to the most north-eastern stations (TATR and VLDR). This increase is well correlated with the decrease of Neogene sediments from W to E from about 1300 m (near TUDR) to about 1000 m (near VLDR) (Raileanu, Annual report 2006).

For station CFR, the fundamental resonance frequency is observed at high frequency (∼ 8 Hz) with an amplitude

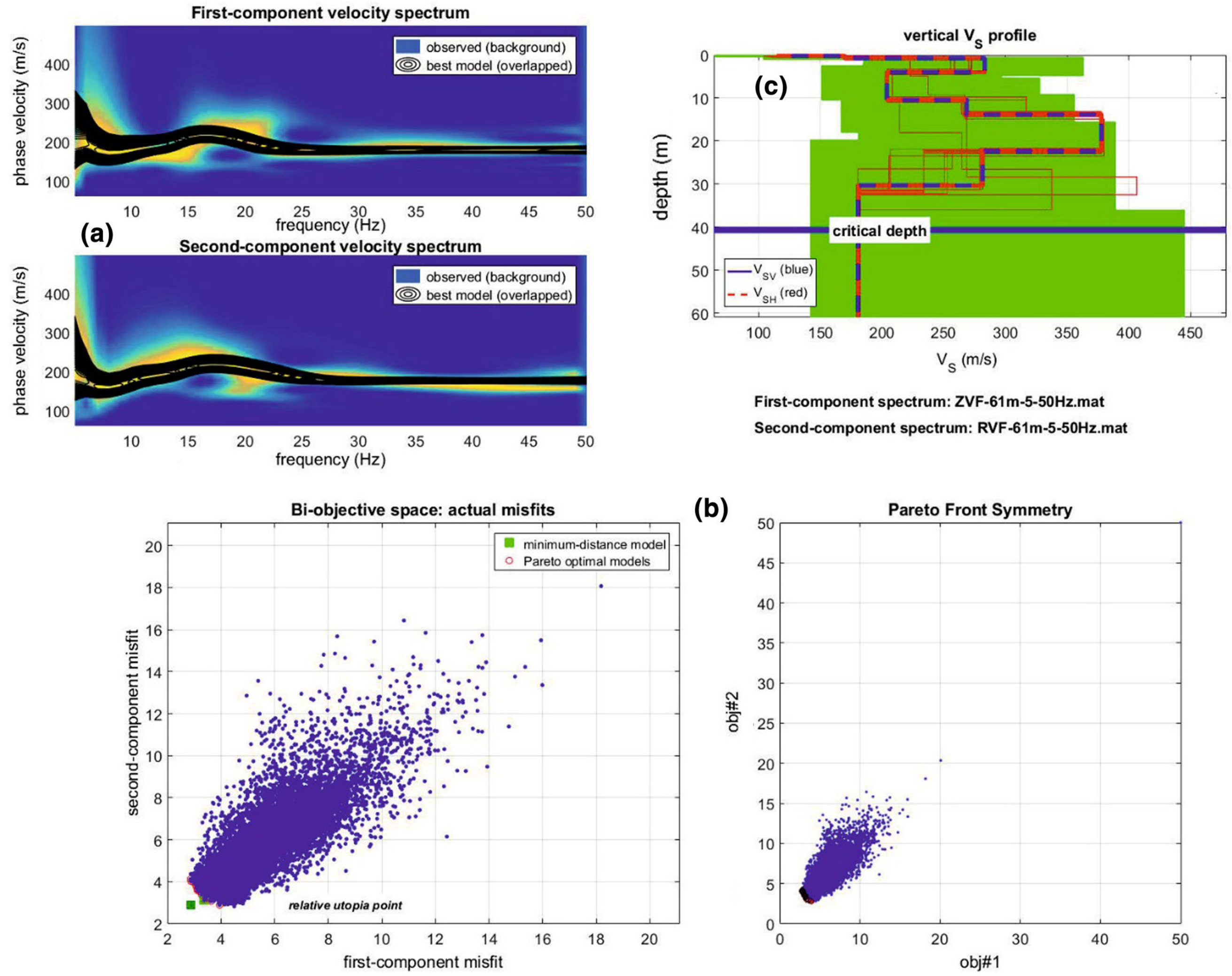


Fig. 13 **a** Joint analysis of the vertical and radial components (group velocities) of Rayleigh waves (colors in the background represent the field data while overlaying black contour lines represent the velocity spectra of the retrieved model) at station TATR. **b** Pareto front models of the joint inversion of vertical and radial components of Rayleigh wave. Distribution of the evaluated models in the bi-objective space (obj#1: vertical Rayleigh-wave misfit; obj#2: radial Rayleigh-wave misfit). **c** The retrieved vs profile

Table 1 Summary of the results obtained from HVNR analysis and inversion of the Rayleigh wave

Station	Computed Vs_{30} (m/s)			Soil classification (EC 8)	Vs_{30} (m/s) (EC 8)	Fundamental resonance	
	Vertical	Radial	Joint inversion			Frequency (Hz)	Amplitude
CFR	458	516	461	B	360–800	8.04	3.2
IZVR	233	240	252	C	180–360	0.19	5.2
SCHL	239	297	301	C	180–360	0.22	3.9
SCTR	207	190	227	C	180–360	0.24	5.0
SLCR	–	–	–	–	–	0.19	5.1
TATR	267	239	263	C	180–360	0.29	4.5
TUDR	239	276	214	C	180–360	0.15	6.4
VLDR	282	261	274	C	180–360	0.28	5.1

not larger than 3. At this station, the basement is placed at 1200 m depth, but the velocity contrast is very low, from 3400 to 3500 m/s.

Also, for this station, a strong dependency on the azimuth of the resonance amplitude is noticed. This suggests a complex site effect due most probably to the surface rock degradation and topographical irregularities observed at the site.

In the absence of a reference site, we applied a similar technique with HVNR method to earthquake data to investigate the site amplification. The HVSR technique provided consistent results in all sites. For the stations IZVR, SCHL, SCTR, SLCR, TATR, TUDR, VLDR amplifications were identified for a wide frequency band, from 0.15 Hz to 5–7 Hz. The fundamental resonance frequency obtained from noise data is present also in all HVSR curves, but only for station VLDR dominates the spectral ratio. If we look at the HVNR results for this station (Fig. 9), we can observe that the fundamental resonance peak has the largest amplitude and is the narrowest of all HVNR curves. This suggests a strong velocity contrast between sediments and bedrock which is also transmitted to the earthquake ground motion. For station CFR, the HVSR and HVNR curves are in good agreement, with no amplification till 2–4 Hz and an increase of the amplification up to 10 Hz. In case of HVSR, the maximum amplitude reaches three between 8 and 10 Hz.

Finally, using active seismic data (Rayleigh waves) generated by vertical force, it was possible to retrieve surface shear-wave velocity profiles, and then to calculate Vs_{30} values and propose the soil class for each site according to EC8 soil classes. For each station, we used three types of datasets for inversion: Rayleigh wave data recorded on the vertical and radial components and the combined dataset of the first two. The inversions of these data provided slightly different velocity profiles and Vs_{30} values (Table 1), but the soil class deduced no matter what dataset we used in the inversion was the same at each station. However, we strongly recommend the joint inversion of multiple datasets (Rayleigh and Love waves, Radial-to-Vertical Spectral Ratio—RVSR) to better constrain the final solution.

Acknowledgements This work was partly supported by a grant from the Romanian National Authority for Scientific Research and Innovation (ANCSI)-UEFISCDI, project number PN-II-RU-TE-2014-4-0701 and partly by a project carried out within Nucleu Program, supported by ANCSI, project number PN 16 35 01 01. The authors express their thanks to Viorel Pirvu who helped with the active seismic measurements performed at seismic stations.

Compliance with ethical standards

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

References

Balan SF, Ioane D, Cioflan C, Panea I, Apostol B, Malita Z, Chitea F, Anghelache MA (2014) Scenarios for local seismic effects of Tulcea (Romania) crustal earthquakes—preliminary approach of the seismic risk characterization for Tulcea city. In: Bostenaru D, Armas M, Goretti A (eds) Earthquake hazard impact and urban planning. Springer, Dordrecht, pp 85–103

Bard PY (1999) Microtremor measurements: a tool for site effects estimation. In: Irikura K, Kudo K, Okada H, Sasatani T (eds) The effects of surface geology on seismic motion. Balkema, Rotterdam, pp 1251–1279

Dal Moro G (2008) VS and VP vertical profiling and poisson ratio estimation via joint inversion of Rayleigh waves and refraction travel times by means of bi-objective evolutionary algorithm. J Appl Geophys 66:15–24

Dal Moro G (2014) Surface wave analysis for near surface applications. Elsevier, Amsterdam, p 244

Dal Moro G, Ferigo F (2011) Joint analysis of Rayleigh and love wave dispersion for near-surface studies: issues, criteria and improvements. J Appl Geophys 75:573–589

Dal Moro G, Pipan M, Gabrielli P (2007) Rayleigh wave dispersion curve inversion via genetic algorithms and marginal posterior probability density estimation. J Appl Geophys 61(1):39–55

Diaz J, Villasenor A, Morales J, Pazos A, Cordoba D, Pulgar J, Garcia-Lobon JL, Harnafi M, Carbonell R, Gallart J, TopoIberia Seismic Working Group (2010) Background noise characteristics at the IberArray Broadband Seismic Network. Bull Seismol Soc Am 100(2):618–628

Evangelidis CP, Melis NS (2012) Ambient noise levels in Greece as recorded at the Hellenic Unified Seismic Network. Bull Seismol Soc Am 102(6):2507–2517

Field EH, Jacob KH (1995) A comparison and test of various site-response estimation techniques, including three that are not reference-site dependent. Bull Seismol Soc Am 85(4):1127–1143

Ghica DV, Grecu B, Popa M, Radulian M (2016) Identification of blasting sources in the Dobrogea seismogenic region, Romania using seismo-acoustic signals. Phys Chem Earth Parts A/B/C 95:125–134

Grecu B, Raileanu V, Bala A, Tataru D (2011) Estimation of site effects in the eastern part of Romania on the basis of H/V ratios of S and coda waves generated by Vrancea intermediate-depth earthquakes. Rom J Phys 56:563–577

Grecu B, Neagoe C, Tataru D (2012) Seismic noise characteristics at the Romanian broadband seismic network. J Earthq Eng 16(5):644–661

Grecu B, Negoe C, Tataru D, Borleanu F, Zaharia B (2016) Analysis of seismic noise in the Romanian-Bulgarian cross-border region (submitted to Journal of Seismology)

Grecu B, Neagoe C, Partheniu R, Nastase E, Zaharia B (2017) New seismic noise model for Romania, science and technologies in geology, exploration and mining. In: Proceedings of 17th International Multidisciplinary Scientific Geoconference, pp 285–292, Eds. STEF92 Technology Ltd (**ISBN 976-619-7408-00-3**)

Ibs-von Seht M, Wohlenberg J (1999) Microtremor measurements used to map thickness of soft sediments. Bull Seis Soc Am 89(1):250–259

Ionesi L (1989) Geologia Romaniei: unitati de platforma si orogenul Nord Dobrogean (translated title: The geology of Romania: platform units and the North-Dobrogean orogen). Thesis, Univ. Al. I. Cuza, Iasi, Romania. p 253 (**in Romanian**)

Konno K, Ohmachi T (1998) Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components. Bull Seis Soc Am. 88(1):228–241

Lermo J, Chávez-García FJ (1993) Site effect evaluation using spectral ratios with only one station. Bull Seismol Soc Am 83(5):1574–1594

Matenco L, Bertotti G, Cloetingh S, Dinu C (2003) Subsidence analysis and tectonic evolution of the external Carpathian-Moesian Platform region during Neogene times. Sed Geol 156:71–94

McNamara DE, Buland RP (2004) Ambient noise levels in the continental United States. Bull Seis Soc Am 94:1517–1527

Nakamura Y (1989) A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. QR Railw Tech Res Inst 30:25–33

Oncescu MC, Marza V, Rizescu M, Popa M (1999) The Romanian earthquakes catalogue between 984 and 1997. In: Wenzel F, Lungu D (eds) Vrancea earthquakes: tectonics, hazard and risk mitigation. Kluwer Academic Publishers, Berlin, pp 43–47

Peterson J (1993) Observation and modeling of seismic background noise, U.S. Geol. Surv. Tech. Rept. 93–322, pp 1–95

Philips WS, Aki K (1986) Site amplification of coda waves from local earthquakes in central California. Bull Seism Soc Am 76:627–648

Popa M, Oros E, Dinu C, Radulian M, Borleanu F, Rogozea M, Munteanu I, Neagoe C (2016) The 2013 earthquake swarm in the galati area: first results for a seismotectonic interpretation. In: Vacareanu R, Ionescu C (eds) The 1940 Vrancea Earthquake. issues, insights and lessons learnt. Springer International Publishing, Switzerland. https://doi.org/10.1007/978-3-319-29844-3_17

Radulian M, Mandrescu N, Panza GF, Popescu E, Utale A (2000) Characterization of seismogenic zones of Romania. In: Panza G, Radulian M, Trifu C (eds) Seismic hazard of the Circum-Pannonian Region. Birkhäuser, Basel, pp 57–77

Raileanu V (2006) Annual report for the Contract no: 31 N/ 23.01.2006, Project—Advanced research of the disaster management of the strong Romanian earthquakes, Director of the project dr. Raileanu V., NIEP

Sandulescu M (ed) (1984) Geotectonica Romaniei (translated title: Geotectonics of Romania). Tehnica, Bucharest, p 335 **(in Romanian)**

Site Effects Assessment using Ambient Excitations (SESAME) European project (2005) Deliverable D23.12—Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations: measurements, processing and interpretation. http://www.sesame-fp5.obs.ujf-grenoble.fr

Stutzmann E, Schimmel M, Patau G, Maggi A (2009) Global climate imprint on seismic noise. Geochem Geophys Geosyst. https://doi.org/10.1029/2009GC002619

Acta Geophysica (2018) 66:167–177
https://doi.org/10.1007/s11600-018-0115-4

RESEARCH ARTICLE - SOLID EARTH SCIENCES

Pre-seismic geomagnetic and ionosphere signatures related to the Mw5.7 earthquake occurred in Vrancea zone on September 24, 2016

Dragos Armand Stanica[1] · **Dumitru Stanica**[1] · **Jan Błęcki**[2] · **Tomasz Ernst**[3] · **Waldemar Jóźwiak**[3] · **Jan Słomiński**[2]

Received: 25 October 2017 / Accepted: 25 January 2018 / Published online: 2 February 2018

Abstract

To emphasize the relationship between the pre-seismic geomagnetic signals and Vrancea seismicity, in this work it is hypothesized that before an earthquake initiation, the high stress reached into seismogenic volume generates dehydration of the rocks and fracturing processes followed by release of electric charges along the faulting systems, which lead to resistivity changes. These changes were explored on September 2016 by the normalized function Bzn obtained from the geomagnetic data recorded in ULF range (0.001–0.0083 Hz). A statistical analysis was also performed to discriminate on the new Bzn* time series a pre-seismic signature related to the Mw5.7 earthquake. Significant anomalous behavior of Bzn* was identified on September 21, with 3 days prior to the onset of the seismic event. Similar information is provided by registrations of the magnetic and electron concentration variations in the ionosphere over the Vrancea zone, by Swarm satellites, 4 days and 1 day before the earthquake.

Keywords Pre-seismic geomagnetic signature · Mw5.7 earthquake · Seismic active Vrancea zone · Swarm satellite · Ionospheric effects

Introduction

In the last few decades on the ground-based observations, more data regarding the existence of the electromagnetic precursors related to the earthquakes have been accumulated (Biagi et al. 2011; Fraser-Smith et al. 1990; Fenoglio et al. 1995; Han et al. 2015; Hattori et al. 2013; Hayakawa and Fujinawa 1994; Huang 2011; Johnston 1997; Kopytenko et al. 1994; Nagao et al. 2002; Stanica and Stanica 2010, 2011, 2012; Stanica et al. 2015; Uyeda et al. 2011; Uyeda 2013, 2015; Varotsos 2005). The satellite observations over the seismic areas indicate changes of the ionosphere state before the earthquakes (Błęcki et al. 2010, 2011; Hayakawa et al. 2011; Parrot 1995; Parrot et al. 2015). Even if all the above works have presented a posteriori observations related to the different pre-seismic signatures, their very optimistic results, together with the theoretical models of fracture induced electromagnetic phenomena (Rabinovitch et al. 2007; Hunt et al. 2007), laboratory experiments on the rock samples (Freund 2000; Freund et al. 2006), ionosphere state observation by the vertical magnetic component (Ernst et al. 2010), the correlation of resistivity variations and geodynamic processes (Bataleva et al. 2013) and very long time electromagnetic data acquisition, processing and analysis realized for a unique and very active seismic zone in Romania, offer a great advantage in emphasizing useful information regarding the possible interrelations between the Vrancea's earthquakes generation and geomagnetic signature. Thus, in this work, we focused on the geomagnetic methodology

---

✉ Waldemar Jóźwiak
jozwiak@igf.edu.pl

Dragos Armand Stanica
armand@geodin.ro

Jan Błęcki
jblecki@cbk.waw.pl

[1] Institute of Geodynamics of the Romanian Academy, Bucharest, Romania

[2] Space Research Centre, Warsaw, Poland

[3] Institute of Geophysics, Warsaw, Poland

able to identify in the ultra-low frequency (ULF) range a pre-seismic anomalous signal related to the Mw5.7 earthquake triggered on September 24, 2016, and to bring strong arguments in favor of the need to asses such information, daily, on the institute website (http://www.geodin.ro), as it is shown in Fig. 1, with the aim to promote an international scientific co-operation.

Methodology, data processing and analyzing

An earthquake of Mw5.7 struck the Vrancea zone, Romania, on September 24, 2016 at 02:11:20 local time (LT). The epicenter was located at the geographic coordinates 45.710N and 26.620E, with a focal depth at 92 km and, about 150 km NE of Bucharest, as it was determined by the Euro-Mediterranean Seismic Centre (http://www.emsc-csem.org). Although this earthquake was moderate, it was felt in Bucharest and generated panic to a certain extent through its inhabitants, being intensively mediated on the TV “Breaking News”, as seismologists systematically claimed a possible imminence of a large earthquake, based on the recurrence rate of about 40 years, on the analogy of the previous two earthquakes incidence of Mw7.7 on November 10, 1940 and Mw7.5 on March 4, 1977, respectively.

Basic concept of the geomagnetic precursor

It was demonstrated by Word et al. (1970) that for a 2D geoelectric structure, the vertical component (Bz) of the geomagnetic field is produced mainly by the horizontal magnetic component perpendicular to strike (B⊥) and, as a consequence, the normalized function (Bzn) having the form:

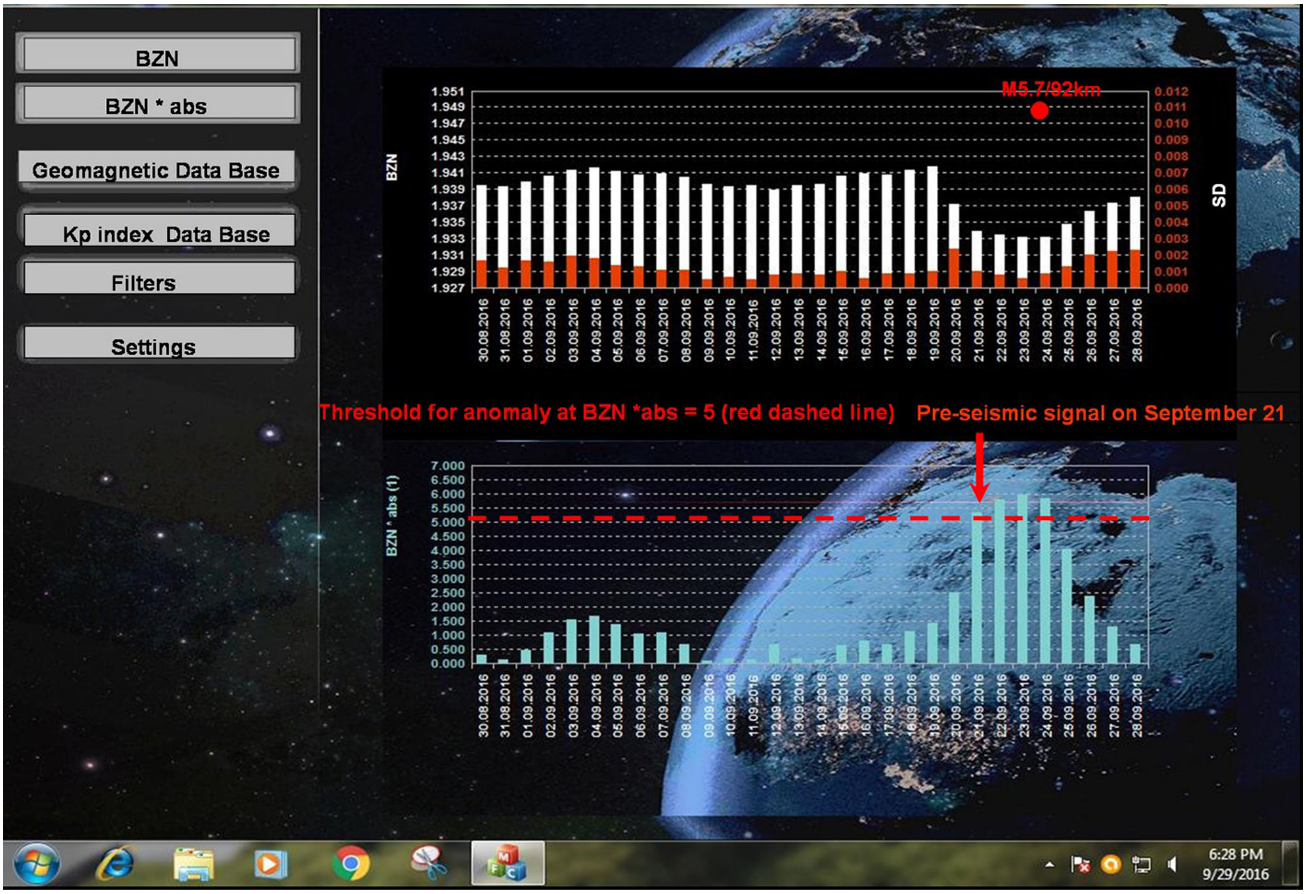


Fig. 1 Daily mean distribution of the geomagnetic parameters Bzn and Bzn* presented on the Institute of Geodynamics of the Romanian Academy website (photo image captured on September 28, 2016). White vertical bar is Bzn; red vertical bar is standard deviation (SD); red full circle is earthquake; the ratio 5.7/92 km is earthquake magnitude/hypocentre depth; blue vertical bar is Bzn*abs (abs means absolute value); red dashed line is threshold for anomaly; red vertical arrow indicates the pre-seismic signal occurrence; the explanation of the parameter Bzn and Bzn* will be done below

$$\mathrm{Bzn}(f)=\frac{\mathrm{Bz}(f)}{\mathrm{B}\perp(f)} \tag{1}$$

should be time invariant and it becomes unstable due to the geodynamic processes related to the intermediate-depth seismicity, being associated with the resistivity changes along the good conducting pass in the lithosphere (Stanica and Stanica 2011), according to the relation (2):

$$|\mathrm{Bzn}(f)|=\sqrt{\frac{\rho_{\parallel}(f)}{\rho z(f)}}, \tag{2}$$

where $\rho_{\parallel}$ is resistivity parallel [Ωm] to the geoelectric strike, ρz is vertical resistivity [Ωm] and f is frequency [Hz].

Consequently, the existence of a 2-D structure gives rise to normalized function Bzn that has the magnitude proportional to the intensity of the geomagnetic component perpendicular to the geoelectric strike, which is in turn determined by the electric field concentrations released in pre-seismic conditions and propagated along the strike, due to the torsion process of the Vrancea's seismogenic volume (Stanica et al. 2004; Stanica and Stanica 2012).

To fulfill the conditions imposed by relation (1), the Geodynamic Observatory Provita de Sus (GOPS) is placed on the Carpathian electrical conductivity anomaly (CECA), at about 100 km westwards of the seismic active Vrancea zone (Fig. 2). This conductivity anomaly is generated by a 2D geoelectric structure (Pinna et al. 1993; Stanica et al. 1999) and it was identified in the ULF range (0.001–0.0083 Hz), on the basis of the magnetotelluric dimensionality parameters skewness and strike (Stanica and Stanica 2010).

According to relation (3), the ULF range (0.001–0.0083 Hz) may be associated with the intermediate-depth earthquakes interval (60–180 km) where possible geomagnetic signals are generated. This supposition is based on the electromagnetic skin depth relation, which in this study is equated with penetration depth of the electromagnetic field into the Earth and can be approximated as:

$$p(f) \approx 500\frac{1}{\sqrt{\sigma f}}, \tag{3}$$

where p is the penetration depth (m), σ is conductivity (S/m) and f is frequency (Hz).

If in relation (3) it is assumed that the Earth's lithosphere has an average conductivity of about 10^{-2} S/m and $f = 0.001$ Hz (minimum value in ULF range), then the geomagnetic signal generated in Vrancea's seismogenic volume, at about 160 km depth, can reach to the GOPS.

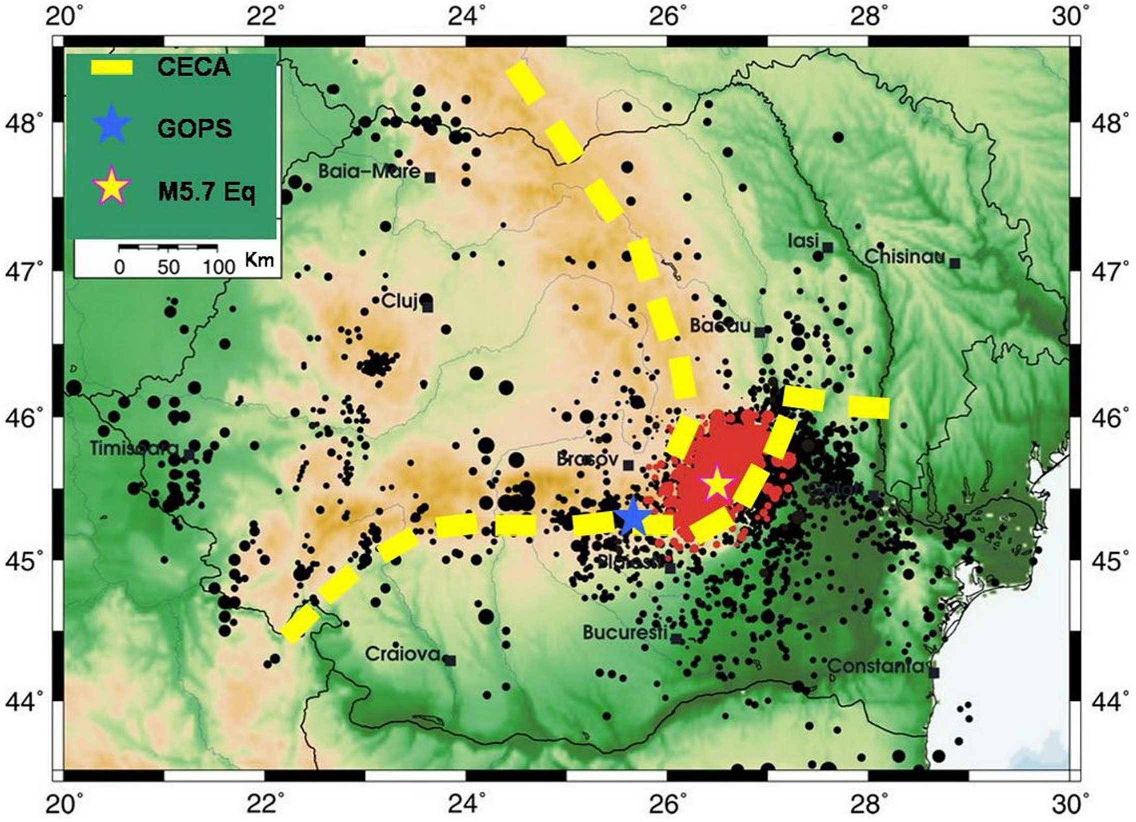


Fig. 2 Map with crustal (black dots) and intermediate (read dots) earthquakes in Vrancea zone. Carpathian Electrical Conductivity Anomaly (yellow dashed line); epicenter of M5.7 earthquake (yellow star); Geodynamic Observatory Provita de Sus (blue star)

Table 1 $B^{\perp}$, Bz, Bzn and Bzn (FFT-BPF) time series obtained for 20 min on September 22, 2016, Bzn mean and Bzn (FFT-BPF) mean are daily averaged values

Date	Time (hh:mm:ss)	$B^{\perp}$ (μT)	Bz (μT)	Bzn	Bzn mean	SD	Bzn (FFT-BPF)	Bzn (FFT-BPF) mean	SD
09-22-2016	00:00:29	22.188	42.883	1.932711	1.93338	0.00076	1.93372	1.93343	0.00053
09-22-2016	00:01:29	22.188	42.883	1.932711			1.93372		
09-22-2016	00:02:29	22.188	42.883	1.932711			1.93371		
09-22-2016	00:03:29	22.189	42.883	1.932624			1.93371		
09-22-2016	00:04:29	22.189	42.883	1.932624			1.93371		
09-22-2016	00:05:29	22.188	42.883	1.932711			1.93371		
09-22-2016	00:06:29	22.187	42.883	1.932798			1.93371		
09-22-2016	00:07:29	22.187	42.883	1.932798			1.93371		
09-22-2016	00:08:29	22.188	42.883	1.932711			1.93371		
09-22-2016	00:09:29	22.188	42.883	1.932711			1.93371		
09-22-2016	00:10:29	22.188	42.883	1.932711			1.93371		
09-22-2016	00:11:29	22.188	42.883	1.932711			1.93371		
09-22-2016	00:12:29	22.188	42.883	1.932711			1.93371		
09-22-2016	00:13:29	22.187	42.883	1.932798			1.9337		
09-22-2016	00:14:29	22.188	42.883	1.932711			1.9337		
09-22-2016	00:15:29	22.187	42.883	1.932798			1.9337		
09-22-2016	00:16:29	22.186	42.883	1.932886			1.9337		
09-22-2016	00:17:29	22.185	42.882	1.932928			1.9337		
09-22-2016	00:18:29	22.185	42.883	1.932973			1.9337		
09-22-2016	00:19:29	22.185	42.882	1.932928			1.9337		

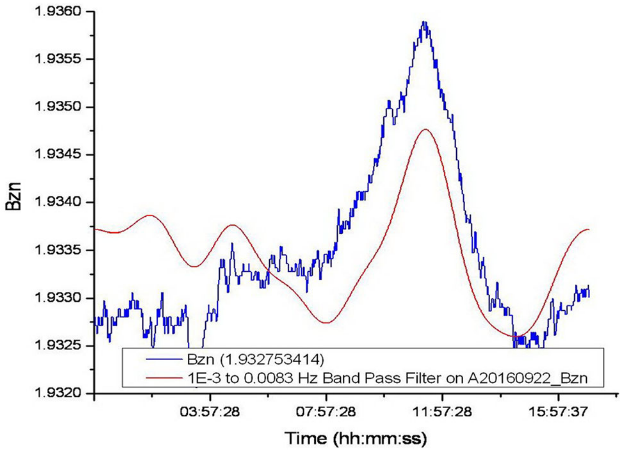


Fig. 3 FFT Band-pass filtering (red line) applied on Bzn time series (blue line) for a time windows of 1024 samples recorded on September 22, 2016

To identify the distance for pre-seismic signal detection, depending on the earthquake magnitude, we used the Morgunov and Malzev (2007) relation:

$$R^{*}(\mathrm{km}) = 10^{0.5M-0.27}, \tag{4}$$

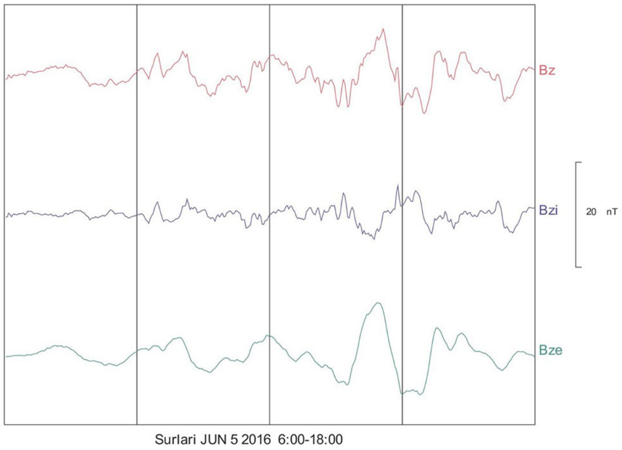


Fig. 4 Twelve-hour time series from Magnetic Observatory Surlari. Bz is total vertical component; Bzi is internal (induced) part of vertical component; Bze is external part of vertical component

where R^* is epicentral distance and M is earthquake magnitude.

In particular, for the Mw5.7 earthquake that occurred on September 24, the epicentral distance $R^* \approx 380$ km. As the distance between GOPS and the earthquake epicenter is about 100 km (Fig. 2), the condition for pre-seismic signal detection imposed by relation (4) is fulfilled.

Satellite observations

The registrations performed by the Swarm satellites over Vrancea zone, before the M5.7 earthquake, are used as an additional information about variations in the ionosphere. Swarm is a constellation of three satellites to measure the Earth's magnetic field and identify the sources of its variations originated from core, ionosphere, magnetosphere, mantle, crust and as well as the oceans. This mission consists of the three identical Swarm satellites (A, B, and C), which were launched on 22 November 2013 into a near-polar orbit. Swarm A and C form the lower pair of satellites flying side by side (1.4° separation in longitude) at an altitude of about 470 km (inclination angle is equal to 87.30°), whereas Swarm B is cruising at higher orbit of about 520 km (inclination angle is equal to 87.75°). They are equipped with a set of six identical instruments—absolute scalar magnetometer, vector field magnetometer, star tracker, electric field instrument, GPS receiver, and accelerometer. Further, we use data from vector field magnetometer to study magnetic field variations and Langmuir probe (being a part of electric field instrument) to provide the representation of the main filed.

The magnetometers installed on-board of Swarm satellites measure main magnetic field with sampling rate 50 Hz (3 components and absolute value). On this way, we are looking for the ionosphere effects associated with seismic activity, that are mainly seen in the variations of the electromagnetic fields in very broad frequency range from fraction of Hz up to several Hz (Olsen et al. 2013).

Geomagnetic and satellite data processing and analysis

Geomagnetic data

The ground-based monitoring system used for this study is installed at GOPS (Fig. 2) and consists of: (1) data logger (MAG-03 DAM) and a three-axes fluxgate magnetic sensor (MAG-03MS), both used for the geomagnetic time series ($B^{\perp}$ and Bz) collection with a sampling rate of 60 s; (2) Computer with programs dedicated to the data acquisition, storage and daily transfer, via internet cable, to Bucharest at the Institute of Geodynamics of the Romanian Academy (IG-RA).

The automatic system installed at IG-RA is composed by: (1) Computer (server) used to receive and storage daily the geomagnetic data records; (2) work-station with specific programs for data processing, analysis and display on the institute webpage.

In this study, the daily mean distributions of the Bzn and Bzn* with their standard deviation (SD) are analyzed on the span of time 01–30 September 2016, by the following procedures:

- FFT Band-pass filtering (BPF) analysis in the ULF range was applied to Bzn time series, for two successive time windows of 1024 samples, with about 30% overlapping on 1440 data acquired each day (Table 1; Fig. 3);
- Statistical analysis based on the standardized random variable has been used to discriminate the pre-seismic anomalous intervals as follows:

$$\mathrm{Bzn}^{*} = \frac{X - \overline{X}}{\overline{Y}}, \tag{5}$$

where X is daily mean value of Bzn, $\overline{X}$ is mean value of Bzn obtained for 30 consecutive days before X, $\overline{Y}$ is mean value of SD obtained for 30 consecutive days before X, Bzn* is threshold for anomaly using SD.

On this way, it is expected to eliminate seasonal variation of the normalized function Bzn and to discriminate with high accuracy the pre-seismic geomagnetic signature related to the seismic event.

A very important element of the initial stage of interpretation is to ascertain that the geomagnetic component Bz is not of external origin, i.e., it is not an over-ground effect of ionosphere currents. We verified it on the basis of records from two magnetic observatories belonging to the INTERMAGNET network: Surlari (Romania) and Tihany (Hungary).

We divided the Bz component variations into the internal (induction) and external parts (Ernst and

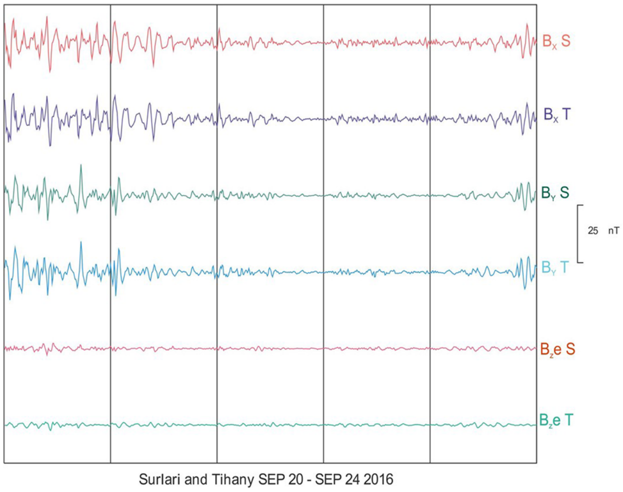


Fig. 5 Five-day simultaneous time series from Magnetic Observatories Surlari and Tihany. B_x S and B_y S are horizontal components for at Surlari; B_x T and B_y T are horizontal components from Tihany; Bze S is external part of vertical component from Surlari; Bze T is external part of vertical component from Tihany

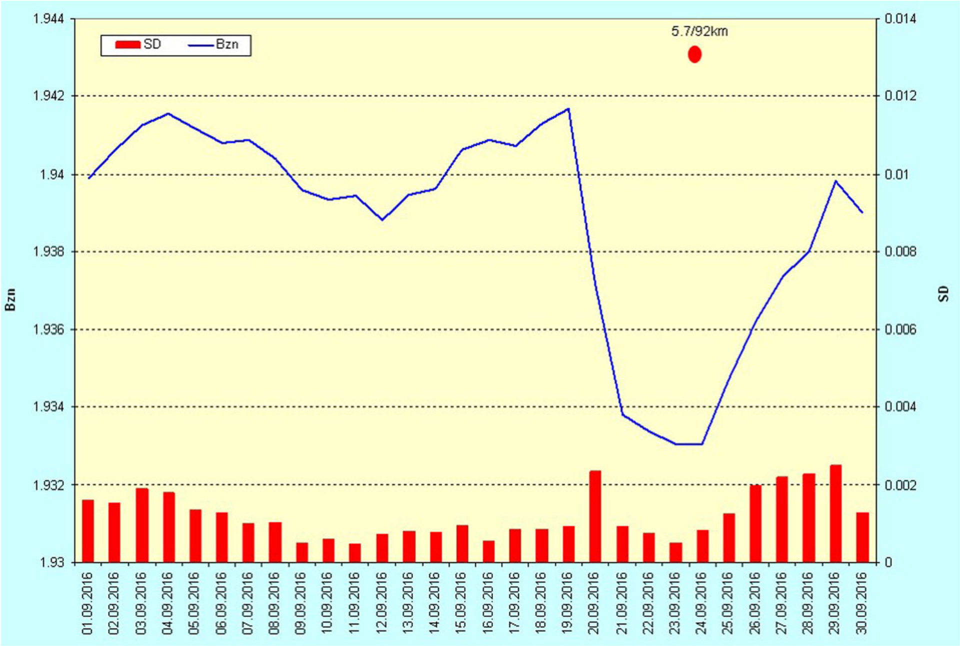


Fig. 6 Daily mean distribution of Bzn and SD on September 2016. Blue line is Bzn; red vertical bar is SD; red star is earthquake; the ratio 5.7/92 km is earthquake magnitude/hypocentre depth, in km

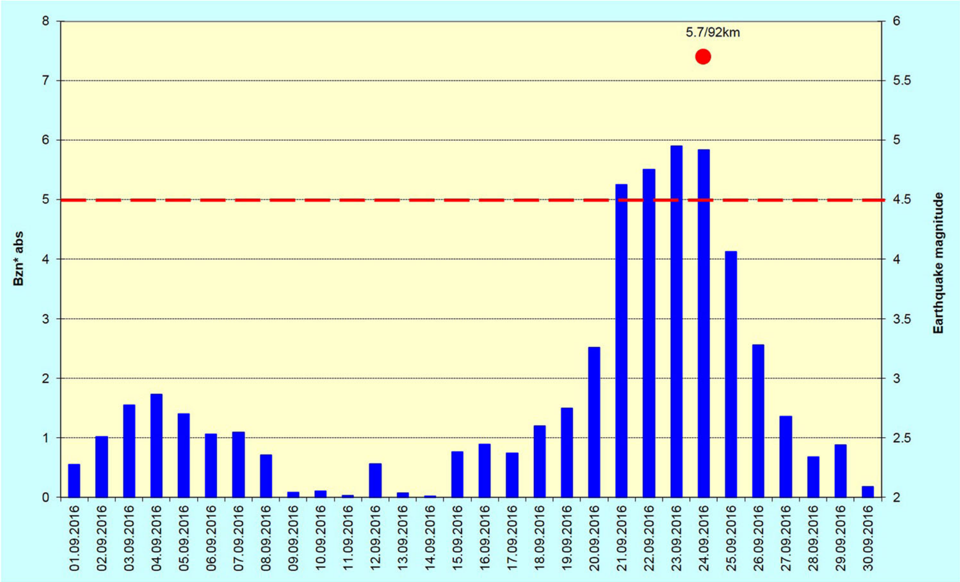


Fig. 7 Daily mean distribution of the Bzn* and earthquakes magnitude obtained on September 2016. Blue bar is Bzn*abs; red full circle is earthquake; the ratio 5.7/92 km is earthquake magnitude/hypocentre depth in km; red dashed line is threshold for anomaly using SD

Jankowski 2005; Ernst et al. 2010). It is worth noting that the external part of Bz variations is often much bigger than the induction (internal) one.

Such a situation is shown in Fig. 4, from which it is also evident how important it is to take into account and analyze the external part Bze in the study of pre-seismic processes.

In Fig. 5, we present 5-day simultaneous time series of the horizontal geomagnetic components B_x and B_y and the external part of Bz (denoted Bze) recorded at the observatories Surlari and Tihany, where it is clearly emphasized that changes of Bze in the both observatories are negligible (< 3 nT), which makes it sure that Bz used in relation (1) is not of external origin, and the anomalous behavior of the Bzn may be interpreted as the earthquake precursor.

Satellite data

The registrations performed by the Swarm satellites limits our studies only to very low frequencies called ULF (ultra-low frequencies up to single Hz) and ELF (extra low frequencies). The magnetic vector is measured with a frequency of 50 Hz, what gives theoretically the possibility to study variations up to 25 Hz, but the filter inside the instrument limits this range to 15 Hz.

Next problem of our studies is related to the extremely small values of these variations (typically fraction of nT) in comparison with the main geomagnetic field measured in tens thousands of nT. Their observation requires sensitive magnetometers and carefully selected methods of signal analysis. The use of the vector field magnetometer on board Swarm, intended for observation of the main geomagnetic field, requires application of special filtering techniques.

Steps distinguished in the data processing chain are as follows:

1. δBi residuals retrieval from the measured signal for three B components;
2. FFT transformation applied to the δBi residuals;
3. Generation of time–frequency spectrograms for δBi along the orbit in the frequency range up to 25 Hz;
4. Integration of lightning's database with derived Swarm δBi spectra.

In the δBi residuals retrieval procedure, a second-order polynomial approximation is applied to set of 1024 samples of the VFM waveform measurements. Also, non-linear least squares method and the Levenberg–Marquardt algorithm are used to provide the representation of the main filed. Thus, δBi is obtained as a difference between measured signal and computed best fit. In the next step FFT is applied, giving power spectrum for each component and intensity of the B field residuals, in the frequency range up to 25 Hz. The FFT window is 1024 samples wide, advanced by 256 samples for overlapping.

Results and discussion

The continuous monitoring of the Bzn and Bzn* time series realized in the last 10 years at GOPS has shown small variability related to their normal trend observed in non-geodynamic conditions and has emphasized consistent anomalies some days before the onset of the seismic event.

To have a comprehensive view on the applied methodology, in this work the daily mean distributions of the Bzn and Bzn* obtained in September 2016, in correlation with Mw5.7 earthquake, are presented in Figs. 6 and 7.

In Fig. 6, there is a significant anomalous domain of minimum, extended on the interval September 19–September 26, with values ranging from 1.933 to 1.942, easily identifiable on the Bzn distribution. This abnormal change of Bzn may be associated with the increasing magnitude of $B^{\perp}$ (see relation 1) which is basically controlled by the electric charges released and propagated along the CECA, prior the onset of the seismic event, as the result of high stress reached in seismogenic volume (Stanica and Stanica 2012).

The Bzn* time series, displayed in absolute value (Bzn*abs) on the same time interval as Bzn, is shown in Fig. 7. On the Bzn* anomaly, extended on the interval September 21–25, a pre-seismic geomagnetic signature corresponding to a magnitude greater than 5·SD (red dashed line) is emphasized starting with 3 days before the onset of the Mw5.7 earthquake on September 24.

The analysis of the ULF/ELF emissions and electron density in the ionosphere in the vicinity of the earthquake epicenter has been performed for 4 days, on the time interval September 20–23 and 1 day before the Vrancea earthquake. A very weak effect of increasing of the intensity and variations of the electron density has been registered close to the epicenter. For these days, the disturbances registered by Swarm are presented below in details.

Figure 8 shows the part of orbit of the Swarm B satellite during observations on September 20. The color of the dots corresponds to the intensity of the magnetic field variations in the determined point. The enhancement of the intensity is seen in the closest vicinity part to the epicenter.

In Fig. 9 are shown the spectra of the magnetic field variations and changes of the electron concentration in the vicinity (around 600 km) of the Vrancea earthquake epicenter 4 days before this event. The variations of the electron density are seen in time interval between 08:17 and 08:19UT. The satellite flown by, in this time, a distance of about 900 km. The variations of the magnetic field were present during this entire interval. The effect shown in this case is very weak in comparison to the effect reported from DEMETER registration (Parrot et al. 2006; Błęcki et al. 2010), but clearly seen. The spectra have a maximum in the lowest part of the frequency range and correspond to the range below the oxygen ions O^+ gyrofrequency which is of the order of 34 Hz. It can be associated with Alfven waves.

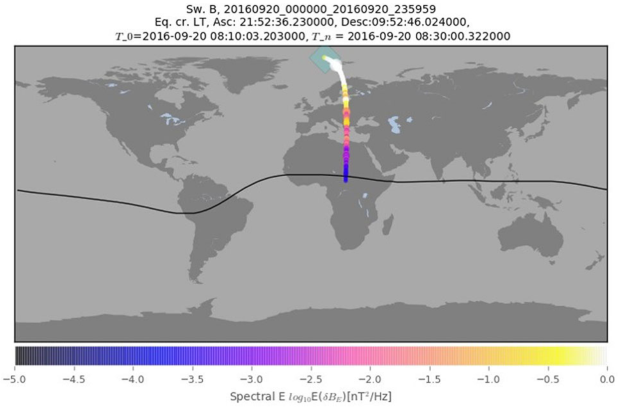


Fig. 8 The part of Swarm B orbit related to the flight over Vrancea zone on September 20. The color of the dots on the orbit line corresponds to the intensity of the magnetic field variations

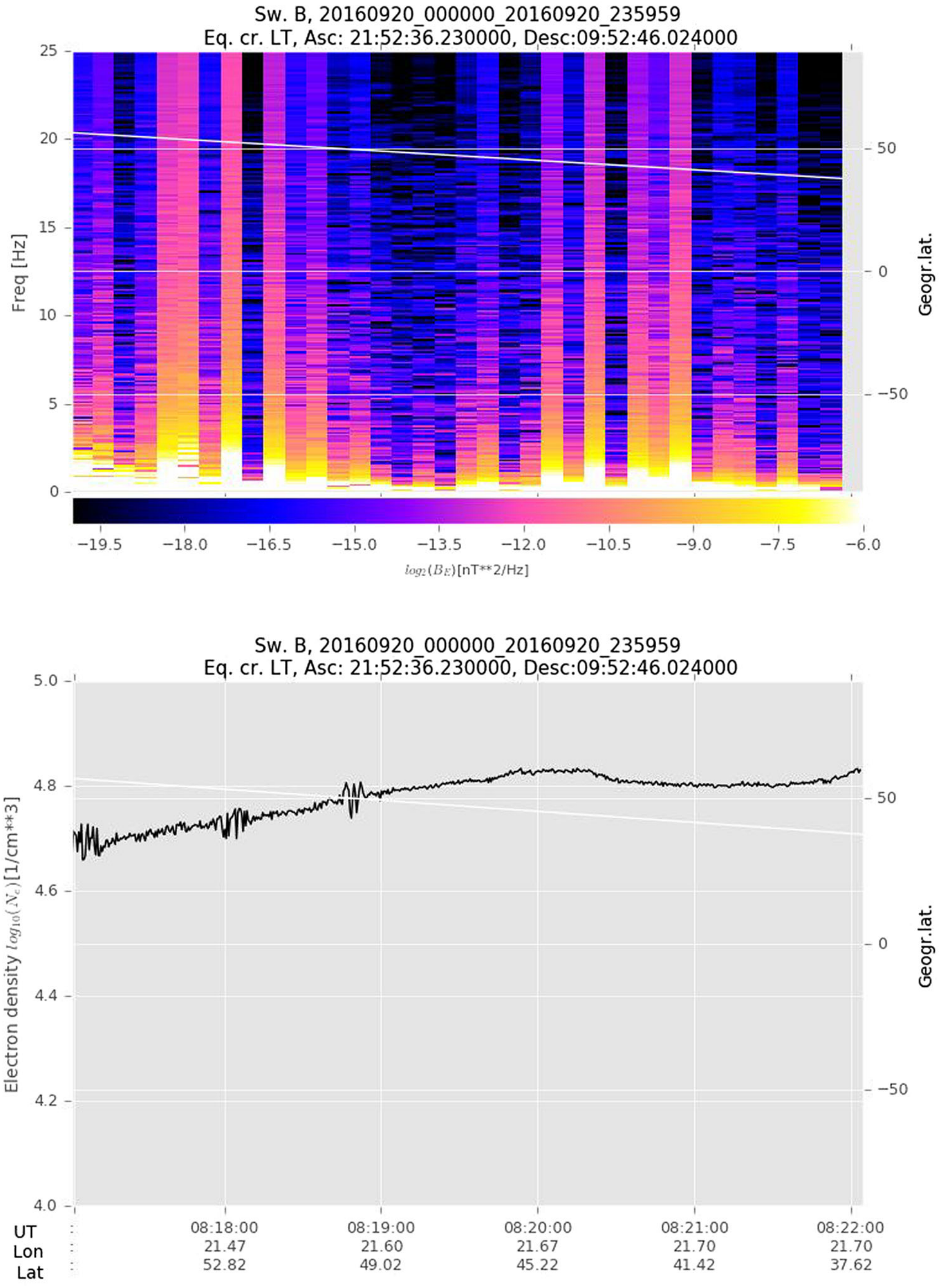


Fig. 9 The spectra of the magnetic field variations (upper panel) and electron concentration registered by Swarm B satellite in the vicinity of the Vrancea zone on September 20

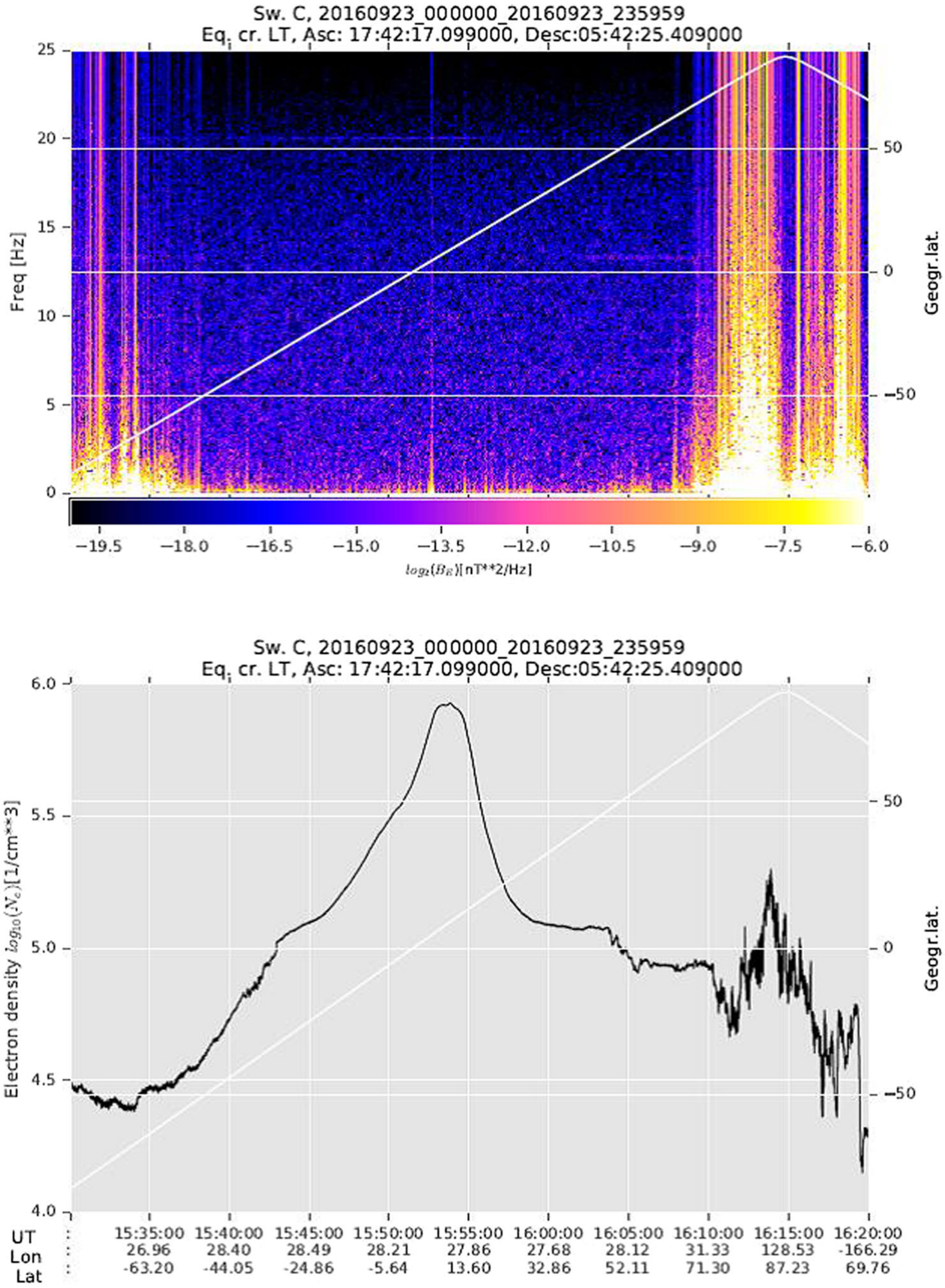
Figure 10 presents spectra of the magnetic field variations and electron concentration taken on September 23 (1 day before the discussed earthquake) from half orbit crossing the Vrancea zone. One can see strong effects associated with high latitude regions (auroral oval, ionospheric through) in comparison with the weak variations in the vicinity (distance about 500 km) of the epicenter. The closest vicinity was at 16:04UT. The disturbances are seen around 16:04 to 16:05UT.

The weakness of the variations seen by Swarm satellites in comparison to reported after DEMETER mission can be related to the smaller effects in the magnetic field than in electric measured by DEMETER, the earthquake was also weaker and may be the local conditions not amplified the electromagnetic effect.

Conclusions

In this paper, we have investigated the ULF geomagnetic data recorded on September 2016 to find a possible pre-seismic signal associated with Mw5.7 earthquake occurred

Fig. 10 The same as in Fig. 9, but for the September 23 and taken for entire half orbit crossing area in vicinity of the Vrancea zone. White line represents the position of the satellite



on September 24. On this interval of analysis, a very clear and unusual anomaly of minimum, extended from September 19 to September 26, has been detected on the Bzn distribution. The new Bzn* time series, obtained after applying a statistical analysis based on Eq. 5, indicates that the variability of Bzn is not at random one, this being a significant and reliable anomalous pre-seismic effect associated with Mw5.7 earthquake. The both results suggest that the pre-seismic anomalous behavior of the Bzn and Bzn*, emphasized by Figs. 6 and 7, was triggered with 3 days before the onset of the seismic event. Complimentarily, the analysis of external part of the vertical geomagnetic field, recorded in the observatories Surlari and Tihany, demonstrated that vertical component (Bz) recorded at GOPS is of internal origin and, consequently, the pre-seismic anomalous signatures observed in the Bzn time series may be interpreted as the earthquake precursor. The measurements originated from Swarm satellites also indicate some disturbances of the magnetic field and electron concentration over Vrancea zone with 3 days before the Mw5.7 earthquake, but the effect was rather small because the earthquake was weaker.

Acknowledgements This work has been supported by the Joint Research Project between the Romanian Academy and Polish Academy of Sciences. The authors thank to the Institute of Geodynamics of the Romanian Academy for making geomagnetic data available to the user community and ESA for access to data from Swarm satellites. Some results presented in this paper are based on data collected at Magnetic Observatories. We thank the national institutes that support them and the INTERMAGNET for promoting high standards of magnetic observatory practice. This work was partially supported within statutory activities No 3841/E-41/S/2017 and Grant NCN 2014/15/B/ST10/00789 of the Ministry of Science and Higher Education of Poland.

References

Bataleva EA, Batalev VYu, Rybin AK (2013) On the correlation of crustal conductivity variations and geodynamic processes. Izvestya. Phys Solid Earth 3:105–113

Biagi PF, Maggipinto T, Righetti F, Loiacono D, Schiavulli L, Ligonzo T, Ermini A, Moldovan IA, Moldovan AS, Buyuksarac A, Silva HG, Bezzeghoud M, Contadakis ME (2011) The European VLF/LF radio network to search for earthquake precursors: setting up and natural/man-made disturbances. Nat Hazards Earth Syst Sci 11:333–344. https://doi.org/10.5194/nhess-11-333-2011

Błęcki J, Parrot M, Wronowski R (2010) Studies of the electromagnetic field variations in ELF frequency range registered by DEMETER over the Sichuan region prior to the 12 May 2008 Earthquake. Int J Remote Sens 31:3615–3629. https://doi.org/10.1080/01431161003727754

Błęcki J, Parrot M, Wronowski R (2011) Plasma turbulence in the ionosphere prior to earthquakes, some remarks on the DEMETER registrations. JAES 41:450–458. https://doi.org/10.1016/j.jseaes.2010.05.016

Ernst T, Jankowski J (2005) On the plane wave approximation of the external geomagnetic field in the regional induction study. Izv Phys Solid Earth 41(5):363–370

Ernst T, Jankowski J, Nowozynski K (2010) A new magnetic index based on the external part of vertical geomagnetic variation. Acta Geophys 58(6):963–972. https://doi.org/10.2478/s11600-010-0014-9

Fenoglio MA, Johnston MJS, Bierlee JD (1995) Magnetic and electric fields associated with changes in high pore pressure in fault zones: application to the Loma Prieta ULF emissions. J Geophys Res 100:12951–12958

Fraser-Smith AC, Bernardi A, Mc Gill PR, Ladd ME, Halliwell RA, Villard OG Jr (1990) Low frequency magnetic field measurements near the epicenter of the M 7.1 Loma Prieta earthquake. Geophys Res Lett 17:1465–1468

Freund F (2000) Time-resolved study of charge generation and propagation in igneous rocks. J Geophys Res B 105:11001–11019

Freund FT, Takeuchi A, Lau RWS (2006) Electric currents streaming out of stressed igneous rocks—a step towards understanding pre-earthquake low frequency EM emissions. Phys Chem Earth 31:389–396

Han P, Hattori K, Xu G, Ashida R, Chen CH, Febriani F, Yamaguchi H (2015) Further investigation of geomagnetic diurnal variation associated with the 2011 off the Pacific coast of Tohoku earthquake (Mw 9.0). J Asian Earth Sci 114:431–434. https://doi.org/10.1016/j.jseas.2015.02.022

Hattori K, Han P, Huang Q (2013) Global variation of ULF geomagnetic fields and detection of anomalous changes at a certain observatory using reference data. Electr Eng Jpn 182:9–18. https://doi.org/10.1002/eej.22299

Hayakawa M, Fujinawa AY (1994) Electromagnetic phenomena related to the earthquake prediction. Terra Scientific Pub Comp, Tokyo

Hayakawa M, Hobara Y, Ohta K, Hattori K (2011) The ultra-low-frequency magnetic disturbances associated with earthquakes. Earthq Sci 24:523–534. https://doi.org/10.1007/s11589-011-0814-2

Huang Q (2011) Retrospective investigation of geophysical data possibly associated with the Ms8.0 Wenchuan earthquake in Sichuan, China. J Asian Earth Sci 41:421–427. https://doi.org/10.1016/j.jseaes.2010.04.014

Hunt A, Gershenzon N, Bambakidis G (2007) Pre-seismic electromagnetic phenomena in the framework of percolation and fractal theories. Tectonophysics 431:23–32. https://doi.org/10.1016/j.tecto.2006.05.02

Johnston MJS (1997) Review of electric and magnetic fields accompanying seismic and volcanic activity. Surv Geophys 18:441–475. https://doi.org/10.2183/pjab.86.257

Kopytenko YA, Matiashvili TG, Voronov PM, Kopytenko EA (1994) Observation of electromagnetic ultra-low frequency lithospheric emission in the Caucasian seismically active zone and their connection with earthquakes. In: Hayakawa M, Fujinawa AY (eds) Electromagnetic phenomena related to earthquake prediction. Terra Scientific Pub Comp, Tokyo, pp 175–180

Morgunov VA, Malzev SA (2007) A multiple fracture model of pre-seismic electromagnetic phenomena. Tectonophysics 431:61–72. https://doi.org/10.1016/j.tecto.2006.05.030

Nagao T, Enomoto Y, Fujinawa Y, Hata M, Hayakawa M, Huang Q, Izutsu J, Kushida Y, Maeda K, Oike K, Uyeda S, Yoshino T (2002) Electromagnetic anomalies associated with 1995 Kobe earthquake. J Geodyn 33:401–411

Olsen N, Friis-Christensen E, Floberghagen R, Alken P, Beggan CD, Chulliat A, Doornbos E, Teixeira da Encarnacao J, Hamilton B, Hulot G, van den Ijssel J, Kuvshinov A, Lesur V, Lühr H, Macmillan S, Maus S, Noja M, Olsen PEH, Park J, Plank G, Püthe C, Rauberg J, Ritter P, Rother M, Sabaka TJ, Schachtschneider R, Sirol O, Stolle C, Thebault E, Thomson AWP, Tøffner-Clausen L, Velimsky J, Vigneron P, Visser PN (2013) The Swarm satellite constellation application and research facility (SCARF) and Swarm data products. Earth Planets Space 65:1189–1200. https://doi.org/10.5047/eps.2013.07.001

Parrot M (1995) Use of satellites to detect seismo-electromagnetic effects. Adv Space Res 15:27–35

Parrot M, Berthelier JJ, Lebreton JP, Sauvaud JA, Santolik O, Blecki J (2006) Examples of unusual ionospheric observations made by the DEMETER satellite over seismic regions. Phys Chem Earth 31:486–495

Parrot M, Berthelier JJ, Blecki J, Brochot JY, Hobara Y, Lagoutte D, Lebreton JP, Němec F, Onishi T, Pinçon JL, Píša D, Santolík O, Sauvaud JA, Slominska E (2015) Unexpected events recorded by the ionospheric satellite DEMETER. Surviv Geophys 36:483–511. https://doi.org/10.1007/s10712-015-9315-5

Pinna E, Soare A, Stanica D, Stanica M (1993) Carpathian conductivity anomaly and its relation to deep substratum structure. Acta Geod Geophys Montan 27(1):35–45

Rabinovitch A, Friend D (2007) A possible source of fracture induced electromagnetic radiation. Tectonophysics 431:15–21. https://doi.org/10.1016/j.tecto.2006.05.027

Stanica D, Stanica DA (2010) Constraints on correlation between the anomalous behaviour of electromagnetic normalized functions (ENF) and the intermediate depth seismic events occurred in Vrancea zone (Romania). Terr Atmos Ocean Sci 21:675–683. https://doi.org/10.3319/TAO.2009.09.09.01(T)

Stanica D, Stanica DA (2011) Anomalous pre-seismic behaviour of the electromagnetic normalized functions related to the intermediate depth earthquakes occurred in Vrancea zone, Romania. Nat Hazards Earth Syst Sci 11:3151–3156. https://doi.org/10.5194/nhess-11-3151

Stanica D, Stanica DA (2012) Earthquakes precursors. In: D'Amico S (ed) Earthquake research and analysis, statistical studies, observations and planning. InTech Open Access Publisher. ISBN 978-953-51-0134-5, pp 79–100. https://doi.org/10.5772/2461

Stanica M, Stanica D, Marin-Furnica C (1999) The placement of the Trans-European Suture Zone on the Romanian Territory. Earth Planets Space 51:1073–1078

Stanica D, Stanica M, Piccardi L, Tondi E, Cello G (2004) Evidence of geodynamic torsion in Vrancea zone (Eastern Carpathians). Rev Roum Gephys 48:15–19

Stanica DA, Stanica D, Vladimirescu N (2015) Long-range anomalous electromagnetic effect related to M9 Great Tohoku earthquake. Earth Sci 4(1):31–38. https://doi.org/10.11648/j.earth.20150401.13

Uyeda S (2013) On earthquake prediction in Japan. Proc Jpn Acad Ser B 89:391–400

Uyeda S (2015) Current affairs in earthquake prediction in Japan. J Asian Earth Sci 114:431–434. https://doi.org/10.1016/j.jseas.2015.07.006

Uyeda S, Nagao T, Kakogawa M (2011) Earthquake prediction and precursor. Encycl Solid Earth Geophys 5:168–178. http://dx.doi.org/10.107/978-90-481-8702-7_4

Varotsos P (2005) The physics of seismic electric signals. TERRAPUB, Tokyo

Word RD, Smith HW, Bostick FX Jr (1970) An investigation of the magnetotelluric tensor impedance method. Electronics Research Center, University of Texas, Austin, p 78712

Acta Geophysica (2018) 66:179–189
https://doi.org/10.1007/s11600-018-0131-4

RESEARCH ARTICLE - APPLIED GEOPHYSICS

Near-surface structure of the Carpathian Foredeep marginal zone in the Roztocze Hills area

M. Majdański[1] · **J. Grzyb**[1] · **B. Owoc**[1] · **T. Krogulec**[2] · **A. Wysocka**[2]

Received: 27 October 2017 / Accepted: 22 March 2018 / Published online: 29 March 2018

Abstract

Shallow seismic survey was made along 1280 m profile in the marginal zone of the Carpathian Foredeep. Measurements performed with standalone wireless stations and especially designed accelerated weight drop system resulted in high fold (up to 60), long offset seismic data. The acquisition has been designed to gather both high-resolution reflection and wide-angle refraction data at long offsets. Seismic processing has been realised separately in two paths with focus on the shallow and deep structures. Data processing for the shallow part combines the travel time tomography and the wide angle reflection imaging. This difficult analysis shows that a careful manual front mute combined with correct statics leads to detailed recognition of structures between 30 and 200 m. For those depths, we recognised several SW dipping tectonic displacements and a main fault zone that probably is the main fault limiting the Roztocze Hills area, and at the same time constitutes the border of the Carpathian Forebulge. The deep interpretation clearly shows a NE dipping evaporate layer at a depth of about 500–700 m. We also show limitations of our survey that leads to unclear recognition of the first 30 m, concluding with the need of joint interpretation with other geophysical methods.

Keywords Carpathian Foredeep · Seismic imaging · Traveltime tomography · Near-surface · Joint seismic interpretation

Introduction

The reflection seismic methods are well known and successfully used in industrial applications over the decades. They are also more often used in smaller scales to recognize near-surface structures. The potential of high-resolution seismic reflection methods has been proven by number of groups studding unconsolidated or postglacial sediments (Buker et al. 1998; Bachrach and Nur 1998; Steeples and Miller 1998; Francese et al. 2007). Standard reflection methods can resolve structures starting from several meters down to several hundreds of meters (Pullan and Hunter 1990; Steeples and Miller 1990; Feroci et al. 2000; Steeples 2000; Dusar et al. 2001; Sugiyama et al. 2003; Bruno et al. 2010; Green et al. 2010).

To recognize the near-surface geological structures an optimal solution is to utilise wide-angle observation and combine them with high resolution reflection data. A successful applications of this approach (Bruno et al. 2010, 2013) shows, that seismic velocities recognized from observed refractions significantly enhance the reflection image of shallow structures.

Here, we report the result of joint interpretation of wide-angle seismic along 2D profile designed to recognize the structure of the marginal zone of the Carpathian Forefront. Recorded large offsets were helpful not only to estimate velocities, but also to image a deep bedrock reflection, that would not be possible using standard deployment. Moreover, designed survey with wireless stations and especially designed accelerated weight drop source was cost-effective and can be performed in a very short time giving a productive tool for the geological studies. Combination of reflection seismic imaging and travel time tomography, that is the tool to recover a high resolution images, has been used before in tectonic scale (Malinowski et al. 2013), the industrial scale (Majdański et al. 2016; Vesnaver et al.

✉ M. Majdański
mmajd@igf.edu.pl

[1] Institute of Geophysics, Polish Academy of Sciences, ks. Janusza 64, 01-452 Warsaw, Poland

[2] Faculty of Geology, University of Warsaw, Żwirki i Wigury 93, 02-089 Warsaw, Poland

1999), but also in near-surface studies (Bruno et al. 2010, 2013).

Geological setting

The present-day area of the Roztocze Hills constitutes as a narrow range of hills, stretching from north-west to south-east from Kraśnik in Poland to Lviv in Ukraine. The area divides the eastern part of the Sandomierz Basin from the Lublin Upland and the Pobuża Basin (Kondracki 1994). The area where the measurements were made is located on the Roztocze Hills along the Carpathian Foredeep marginal zone (Fig. 1). During the Neogene (Miocene) time, this area belonged to the north-eastern part of the outer ramp of the Carpathian Foreland Basin (Wysocka 2006) and it is recognized as their forebulge (Jankowski and Margielewski 2015; Wysocka et al. 2016).

This region is dominated on the surface by sedimentary rocks, mainly Cretaceous marls or sandstones and Neogene (Miocene) sandstones, limestones and marls. These rocks were formed in shallow sea with high hydrodynamic energy.

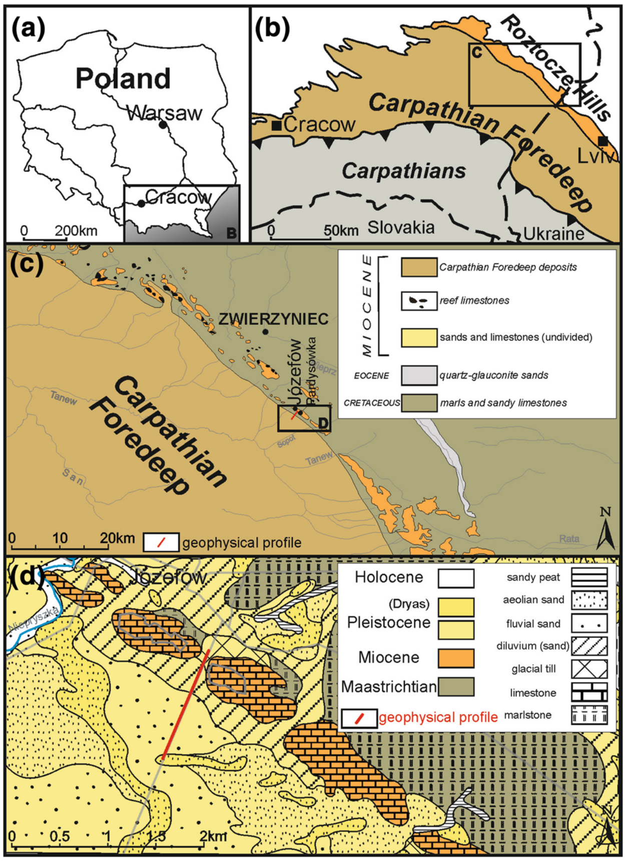


Fig. 1 Schematic map showing: **a** general location of the study region in Poland, **b** tectonic structures in the area; **c** location of seismic profile in the marginal zone of the Carpathian Foredeep; **d** detailed geological map in the vicinity of the profile (after Kurkowski 1998)

Miocene rocks are preserved only in the form of erosional patches along south-western marginal zone (Wysocka and Jasionowski 2006). Carpathian Foredeep in the north eastern part consists mostly of Miocene flat bedded lightly lithified shales, of the same age as the rocks described above, covered by thin layer of Quaternary sediments. In the whole Miocene sequence, the only one characteristic layer, evaporates, consists mainly of anhydrites and gypsum. This layer has variable thickness from 0 up to 60 m and has great significance as the main correlation level and seismic marker on the scale of the entire Carpathian Foredeep (Myśliwiec 2004; Krzywiec 2001). These beds can be observed in the Kozaki 1 drill core. Drilling was done approximately 7.5 km to the south-west from the beginning of the profile. These evaporates are mainly gypsum with a thickness of 34 m and occurs at a depth of 609–643 m. This salina basin was developed during the Badenian salinity crisis in northern Central Paratethys (Bąbel 2004).

Due to the difficulty of stratigraphic recognition of sedimentary rocks on the area of the Roztocze Hills, evaporate layer is the only certain horizon which allowed to determine the relative age of the succession. The lack of evaporates layer on the Polish part of the Roztocze Hills area and different types of rocks with different thicknesses are the main problems in the deposits correlation along the north-eastern Carpathian Foredeep marginal zone.

Measurements were made between Józefów and Pardysówka quarries, which are located in the distance of about 950 m from each other, on the most south-western range of hills of the Roztocze Hills area. Both of them are characterized by Miocene complex, consisting mainly of conglomerates and limestones (Wysocka et al. 2006). The whole complex is located directly on the Cretaceous basement and is inclined slightly toward the south (Roniewicz and Wysocka 1999). In both quarries many normal and reverse faults were found, associated with post-sedimentary Miocene or younger seismic activity in the area. Their planes are parallel to the Carpathian Foredeep marginal zone (Jaroszewski 1977; Jankowski and Margielewski 2015). All observed faults are crossing all the layers at a low angle (Wysocka et al. 2006).

The seismic survey and the field works

The measurements were performed during 3 days in May 2017 along the forest road near Józefów in the Roztocze Hills area, in a direction that crosses the studied structures perpendicularly. For seismic recording we used 60 standalone DATA-CUBE stations with GPS timing connected to 4.5 Hz vertical geophones. As a seismic source we used modified PEG-40 accelerated weight drop with designed handcart equipped with electric power generator. This autonomous system can be operated by two man in rough terrain, and gives reliable and repeatable seismic signals in efficient way. Additionally, the source was equipped with time measurement system based on GPS time, that stored in laptop the time of each stroke based on closure of the electric circuit. The timing system has precision of 1 ms. The whole production was divided for four deployments (Fig. 2) resulting in the total length of 1280 m. Both sources and receiver spacing was 5 m, and for each deployment additional 10 shots to each side was performed resulting in the fold spanning from 22 to 60. The profile was crossing the hill, resulting in complicated elevation along it, especially changing for the last deployment. The total difference in elevation was 22 m. For each shot position, four strokes has been measured to maximise signal to noise ratio. Those records were further stacked using diversity stack technique. An example of raw shot gathers are presented in left panel of Figs. 3 and 4a. A maximum energy in recorded signals is visible around 30 Hz with lower frequencies strictly connected with surface waves. All survey parameters are summarised in Table 1.

Data processing

Reflection seismic imaging

The data were processed using a simple but effective operations that are summarised in Table 2.

The processing has been performed separately for shallow area and for deep structures. For both paths the continuously recorded data has to be cut according to shot times. The next step was to add geometry and sort to shot gathers with detailed quality check. This step is time consuming but important in case of wireless stations, as mistakes in geometry are easy to make and might have a severe impact on the final quality. Further on, a trace editing was performed to exclude noisy channels resulted from poor geophone coupling. Elevation statics was estimated based on the topography and the floating datum has been set just below the lowest elevation at 240 m. Refraction statics was calculated using standard routines implemented in the commercial seismic software with limited offset of 100 m. This allows to recognize the weathering layer without taking into account large velocity variations resulted from change of topography. A next step in common processing was careful front mute to remove guided waves and preserve wide angle shallow reflections. This procedure is crucial, as explained in many papers (Buker et al. 1998; Robertsson et al. 1996), and has been manually performed for each shot. The effect of application of front mute is presented in Fig. 3 in middle panels, where clearly refracted arrivals have been removed while shallow reflections are preserved. The first recognition of near-surface velocities based on analysis of shallow reflections flattening shows variable velocities along the profile that varies between 1400 and 1600 m/s. The effect of reflection flattening is present in Fig. 3 (right panels),

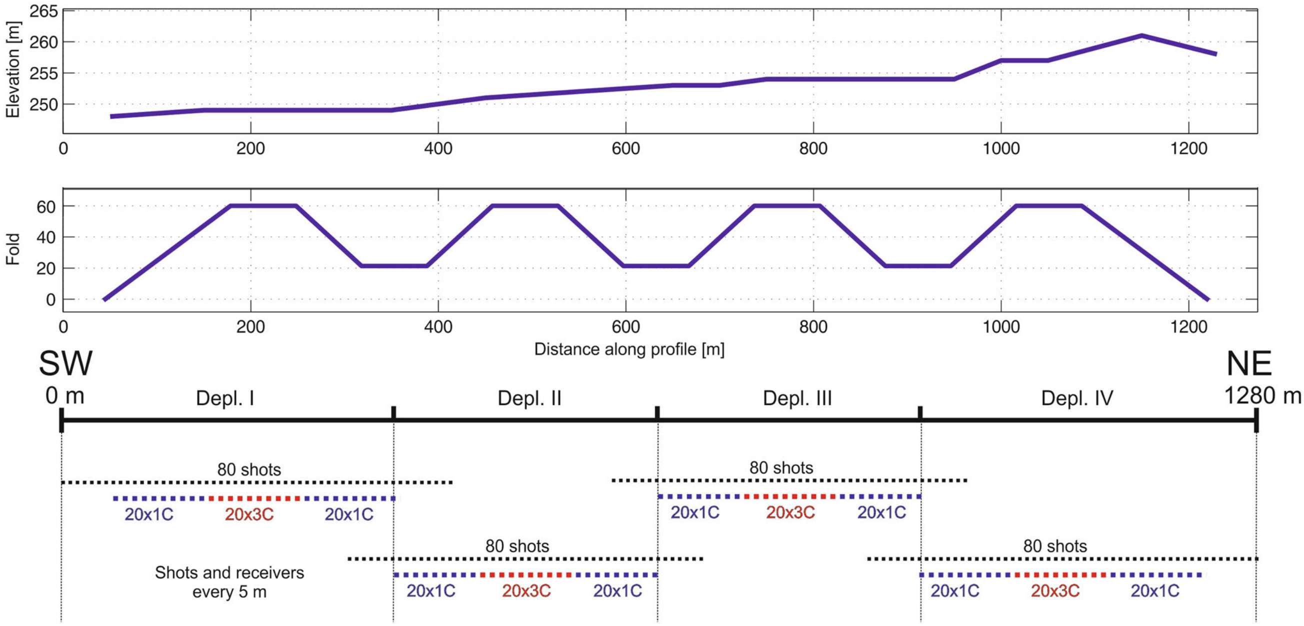


Fig. 2 Field measurement scheme showing four deployments (bottom) using 60 seismic stations and 80 shot positions in each. Thanks to overlaps of shots, the minimum fold was set as 22 (middle panel). The top panel shows the significant change of 22 m in the elevation along 1280 m long profile

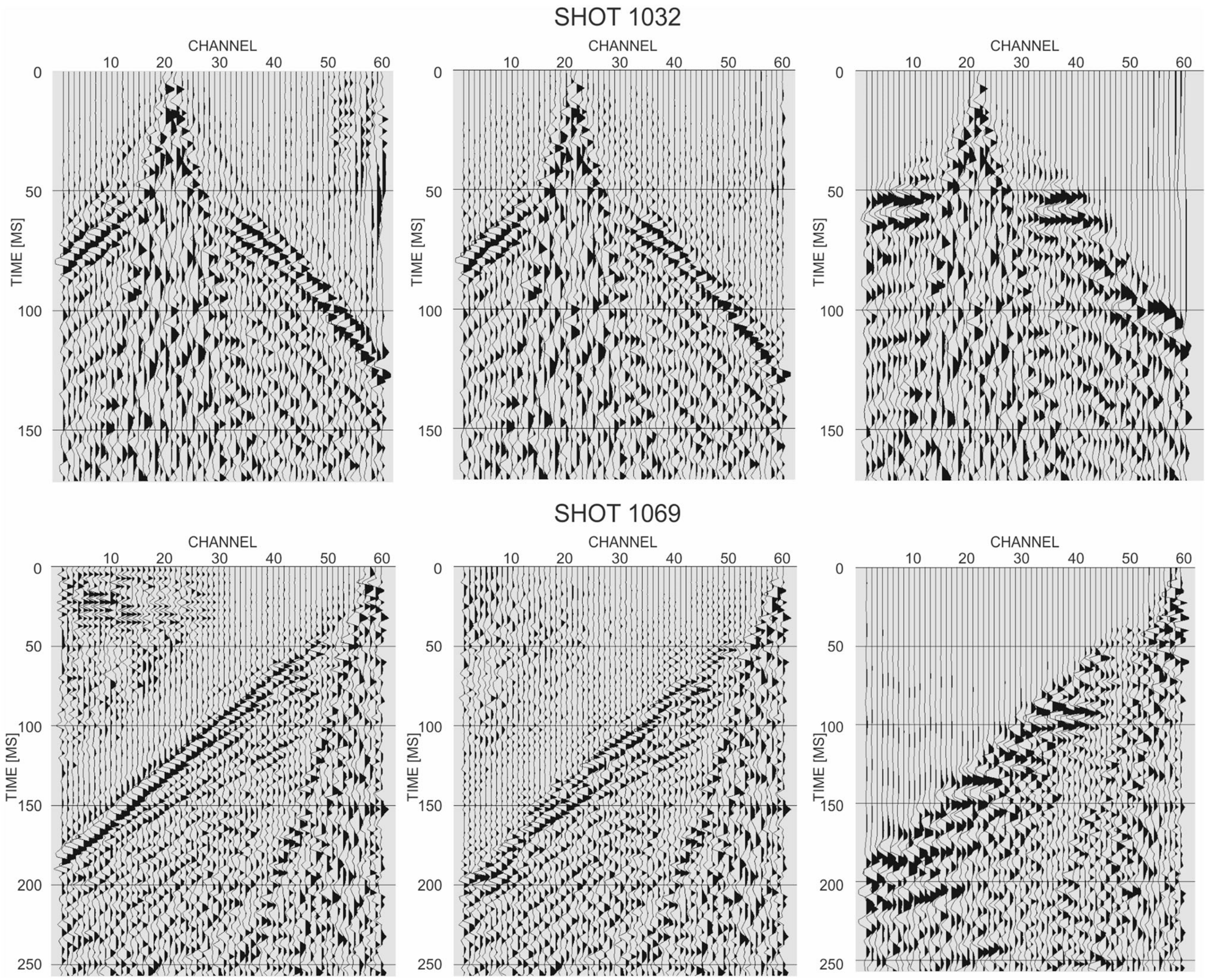


Fig. 3 Example of shot gathers for shots 1032 and 1069 (left panels) after diversity stack of four repeated weight drops. The middle panels show carefully selected front mute, that removes linear refraction arrivals. Right panels show the NMO effect with velocity of 1600 and 1400 m/s, respectively. Clear shallow reflections are visible at 55 ms for shot 1032, and 90 and 140 ms for shot 1069. All panels presented with AGC (100 ms) applied

showing also effect of optimal combination of 70% stretch mute and the front mute.

One of the problems in near-surface processing are strong amplitudes of surface waves that covers useful wide-angle reflections. In our case a simple high-pass filter was efficient enough to attenuate most of surface waves. The best results were obtained when filtering above 30 Hz as presented in Fig. 4c, d. Moreover, high-pass filter attenuates also long wavelength characteristics in shot gathers. Figure 4 shows also a spectral content for an example shot gather showing significant reduction of amplitudes with frequency. To recover sharp images we used spectral whitening technique that is strengthening high frequencies (see Fig. 4d). For this step several deconvolution techniques has been tested, including predictive deconvolutions and spectral whitening. Figure 5 shows an example results using surface consistent deconvolution, surface consistent spectral whitening, single trace deconvolution and spectral whitening. Choosing an optimal deconvolution we were looking for a sharp and coherent wide-angle reflections, like the one at 120 ms, and one at 270 ms. Both surface consistent approaches (Fig. 5a, b) result in poor enhance of high frequencies, thus smeared reflections. Although, the deconvolution performed for each trace separately results in the flattest spectrum (Fig. 5c) wide-angle reflections are not coherent. Finally, we decided to use spectral whitening in range of 10–180 Hz (Fig. 5d) because it gives the clearest image.

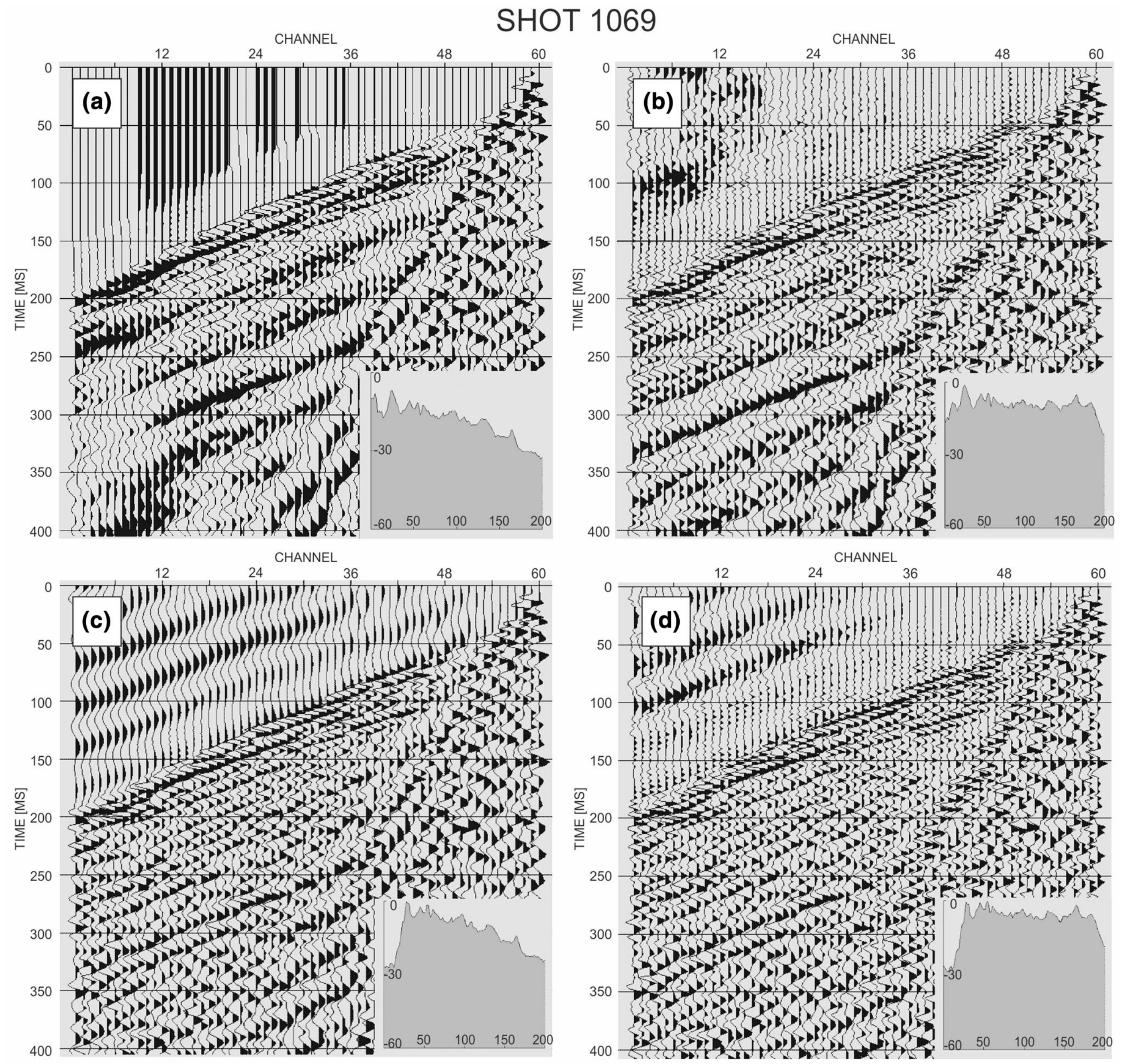


Fig. 4 Example of optimal processing for shot 1069. Raw gather after diversity stack of four repeated weight drops (**a**); gather after spectral whitening **b** shows flat spectrum up to the 180 Hz; **c**, **d** shows corresponding processing steps as in **a**, **b** with applied high pass filter above 30 Hz; inlets in each panel shows spectral content of the gather. All panels presented with AGC (100 ms) and front mute applied

The shallow structure processing

Having problems with standard semblance velocity analysis we wanted to recognize the velocity structure in the shallow part using other techniques. For that, we used well known travel time tomography. The first breaks were manually picked for every fifth shot resulting in 52 regularly spaced shots along the whole profile. All receivers were used in this procedure giving offsets from 5 to 360 m. The refraction arrivals were clearly visible at all offsets, but after careful elimination of noisy channels we prepared the set of 2600 travel times. This travel time set has been inverted using JIVE3D software (Hobro 1999) for a smooth velocity model with defined topography. The inversion procedure was performed according to standard scheme (Majdański et al. 2016; Zelt 1999) resulting in an interesting velocity distribution presented in Fig. 6. The velocity structure is rather smooth with strong increasing gradient with depth, but also with an increasing trend toward NE. Around CDP 480 a strong gradient marks

Table 1 Acquisition parameters

Feature	Measurement
Vertical stack	4
Sampling interval	2.5 ms
Record length	1 s
Receivers	4.5 Hz
Station interval	5 m
Shot interval	5 m
Active channels	60
Fold	22–60
Offset	0–345 m

sudden change of velocity that indicate a SW dipping fault. Three other smaller gradients indicate discontinuity of velocity with more horizontal SW dipping direction. To verify reliability of tomographic result, a middle panel of Fig. 6 shows ray paths from the last iteration of inversion. It proves that all mentioned observations are in well covered areas. This verified result for the first 60 m is alternative information that was further used to create correct velocity field for reflection imaging. With this additional information it was possible to perform semblance based velocity analysis for observed shallow reflections down to 200 ms. All shallow velocities has been further tested by NMO application on shot gathers along the profile. Flattening of wide-angle reflections (see example Fig. 3) confirming that those shallow velocities are correct. The final velocity model in depth domain is presented in the lower panel of Fig. 6. It was generated as a simple extrapolation of tomographic velocities in shallow areas for large depths.

The shallow processing includes both elevation and refraction static corrections and careful front mute. Spectral whitening has been applied to eliminate spectral differences between shots and effect of variable coupling of receivers. For NMO correction we used velocity model from Fig. 6. After the stacking procedure resulting section was FK filtered (with velocity of 50 m/s) and depth converted using the same velocity field. The final stack in depth domain is presented in Fig. 7 (top). Looking closely at the first 200 m we observe several flat reflections spanning from the beginning of the profile till CDP 450. Unfortunately there are no borehole information to verity and classify those reflections. A sudden change of reflectivity is observed at CDP 460 where the fault exist and is confirmed at the surface as a morphological edge. This is probably the main fault limiting the Roztocze Hills area from the south west, and at the same time constitutes the border of the Carpathian Forebulge. At CDP 460–500 there is an area of low reflectivity (yellow polygon in Fig. 7) that coincides with strong velocity gradient in tomography. In this area some discontinuities in small reflectivity is visible (marked with black dashed line). The flat reflections on a left side show small discontinuities that could be traced to form lines gently dipping in SW direction. They could be interpreted as small tectonic displacements that do not reach the surface. Three of those structures (solid yellow lines) are confirmed in tomographic result (lower panel in Fig. 7) by areas of dipping velocity gradients. Unreliable tomographic results are greyed out, so we can confirm internal faults at CDP 120, 210, 310 and 410. The other discontinuities are not that sharp in the seismic image and are marked with dashed lines.

The deep structure processing

In the deep structure processing path, two additional steps were needed to enhance energy at larger times. Firstly, we used spherical divergence procedure linearly dependent on

Table 2 Two paths of data processing

	Common processing	Deep focus (additional steps)
1	Data cutting	
2	Geometry and sorting	
3	Trace editing	
4	Elevation statics	
5	Refraction statics	
6	Front mute	
7		Spherical divergence with time
8	Spectral whitening (10–180 Hz)	
9		Tail mute
10	NMO	
11	Stack	
12	FK and high-pass filtering	
13	Depth conversion	

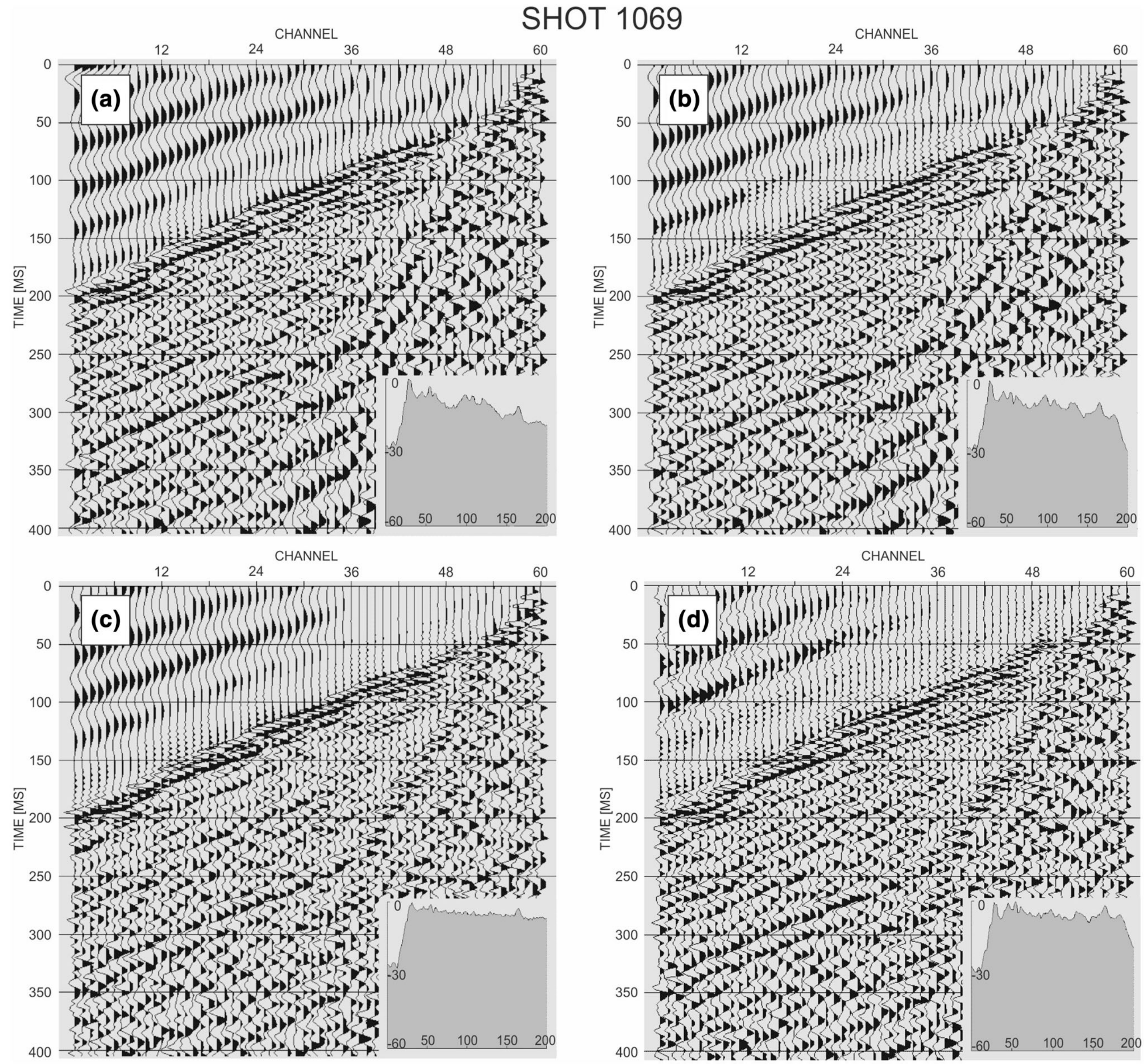


Fig. 5 Example shot 1069 deconvolution tests: surface consistent deconvolution (**a**), surface consistent spectral whitening (**b**), single trace deconvolution (**c**), spectral whitening (**d**). Inlets in each panel shows spectral content of the gather. Simple spectral whitening **d** results in the sharpest wide-angle reflections. All panels presented with AGC (100 ms), front mute and high-pass filter (> 30 Hz) applied

time. This procedure significantly boosts the amplitudes below 300 ms for both deep reflections and much stronger surface waves. Because our data set consist of large offsets it was possible to simply mute the surface waves with hand-picked tail mutes, and use only wide angle reflections for imaging. As presented in Fig. 8, a clear deep reflection at c.a. 550 ms is visible for times preceding the arrival of the surface waves. Moreover, this strong reflection was visible at a number of shots and for various frequencies. Spectral whitening in the range of 10–180 Hz was applied further on to minimise the effects of variable geophones and source plate coupling, and some source generated noise. Unfortunately, standard velocity analysis using semblance plots was not possible for deep structure because of lack of deep reflections between 200 and 600 ms. For stack we used extrapolated velocities as in Fig. 6. The effect of normal moveout and stack is presented in Fig. 9. The main features are the gaps at the deployment contact points. They are the effect of application of tail mutes presented in Fig. 8. This step significantly improves the quality of section, especially at the larger times. Thanks to large offsets for the profile we could simply mute all

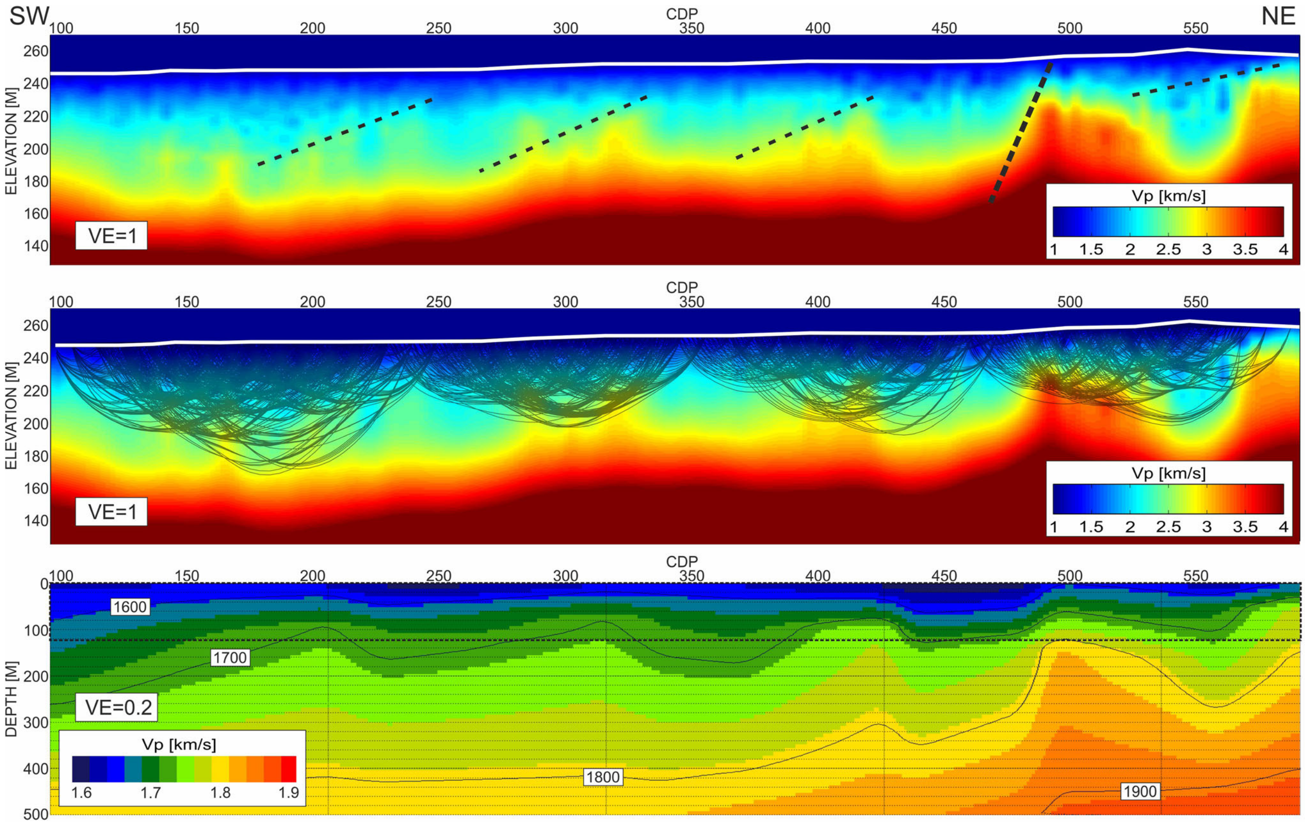


Fig. 6 Result of the first breaks travel time tomography showing the P wave velocity in the shallow structure (top). White line marks the elevation, black dashed lines mark strong dipping velocity gradients indicating the discontinuity of velocities. The middle panels shows the same result with seismic ray paths to mark the well recovered areas. The lower panel, with different vertical exaggeration (VE = 0.2) shows the velocity field based on tomographic result and the velocity analysis, that was further used to the depth conversion. Dashed rectangle marks the area with tomographic velocities

arrivals related to the surface waves and still recover wide-angle reflection. One important aspect in the processing was correct order of tail mutes, AGC and filters application. It is important to filter the data first then apply tail mute procedure. The other order result in strong filter artefacts below the tail mute times being enhanced by AGC procedure giving strong, artificial reflections at large times. The result of correct processing focused on the deep structures shows clear NE deepening reflection marked with a red line (Fig. 9) at depths from 450 to 700 m.

Conclusions

Shallow seismic investigations yielded detailed images of the Carpathian Foredeep marginal zone. A large offset survey combined with modified accelerated weight drop source was specially designed to allow both high resolution reflection image and refraction tomography. Four strokes of the source give enough energy to observe refractions at all offsets up to 350 m, but also to recognize a deep structure down to 700 m. Two processing paths were used to enhance both shallow and deep structures, resulting in detailed image of the near-surface faults, but also a sharp deep reflection. The 5-m spacing used for both shots and receivers was not dense enough to clearly recognize the first 30 m. Additional information like ERT experiment or different seismic processing, e.g., MASW (Park et al. 1999), and finally joint interpretation might be used to further verify this part of the structure. Still, a clear image of a main fault in the area was presented. This weight drop seismic profile was performed to recognize near-surface structures, that is why it was surprising to observe clear reflection at 700 m (Fig. 9). For surveys of this type with limited number of stations we suggest to use different spacing for shots and receivers. For example deploying stations with 8 m spacing and shooting with 2 m spacing will result in similar fold, but would limits number of deployments, limiting number of repeated shots. Thanks to dense shooting it would be possible to recover shallower structures, but wider deployments would give important long offset refracted arrivals. In the end acquisition time should be similar with higher resolution results. To make it optimal it would be beneficial to perform test shooting and

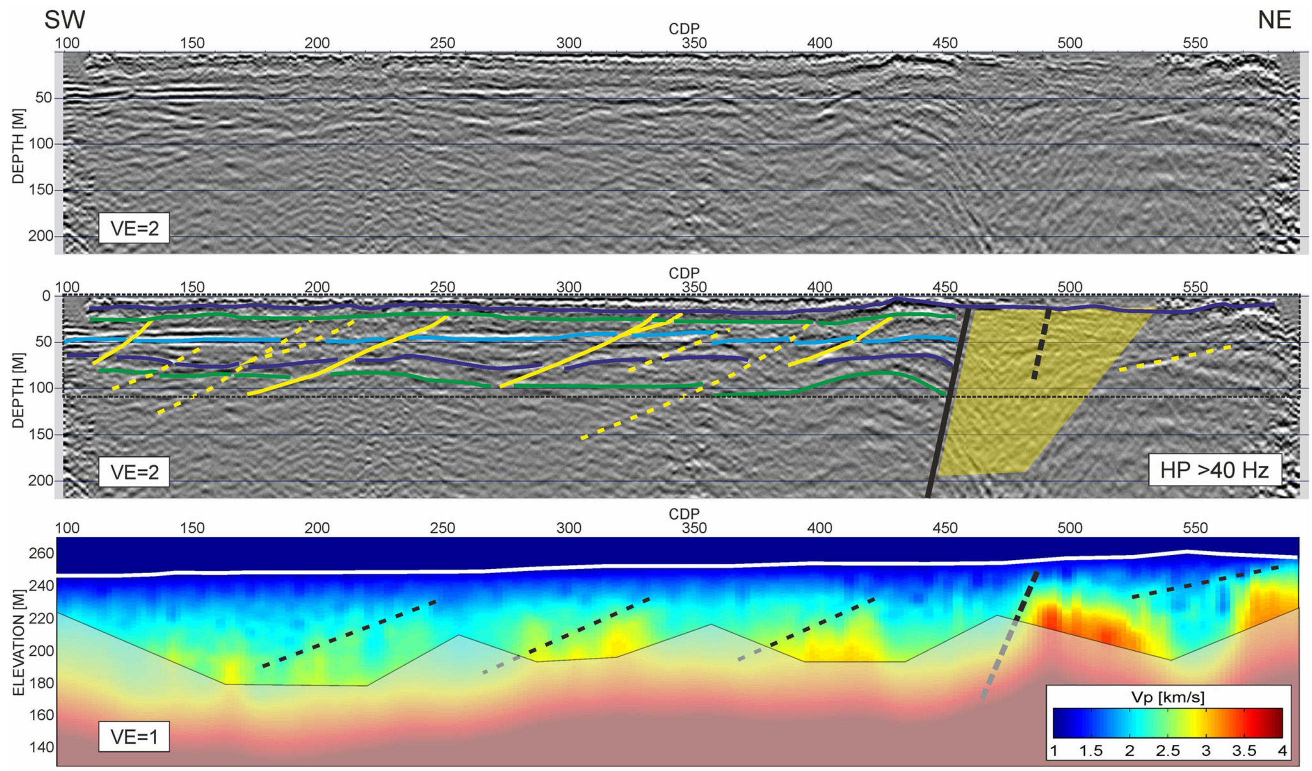


Fig. 7 Shallow structure in the depth domain obtained with shallow reflections enhanced processing (top). The middle panel shows several discontinuities in the flat reflections (marked with colour solid lines) that do not reach the surface. Those marked with solid yellow lines are confirmed with tomography. Yellow polygon marks the area of the low reflectivity that differs from surrounding structures, and corresponds to a zone of the strong velocity gradient in tomography. Dashed rectangular marks the area as in the lower panel. The lower panel, with different vertical exaggeration (VE = 1), shows tomographic result as in Fig. 6 with an unreliable greyed area without ray coverage

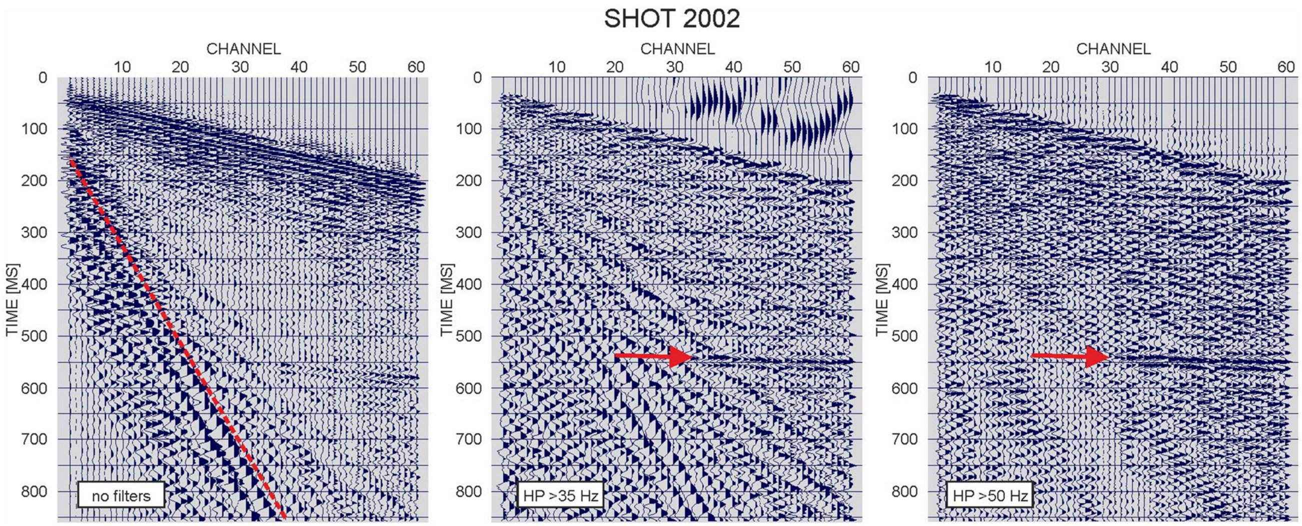


Fig. 8 Clear deep reflection (red arrow) at about 550 ms is visible at a number of shots. Example for shot 2002 shows raw data after diversity stack (left), application of front mute and high pass filter above 35 Hz (middle), and NMO with velocity of 1500 m/s and high pass filter above 50 Hz. Simple high pass filtration clearly removes surface waves and enhance reflections. Flattening of reflection for small NMO velocity shows deepening of the reflector toward NE. All panels presented with AGC (100 ms) applied. The red line marks the tail mute

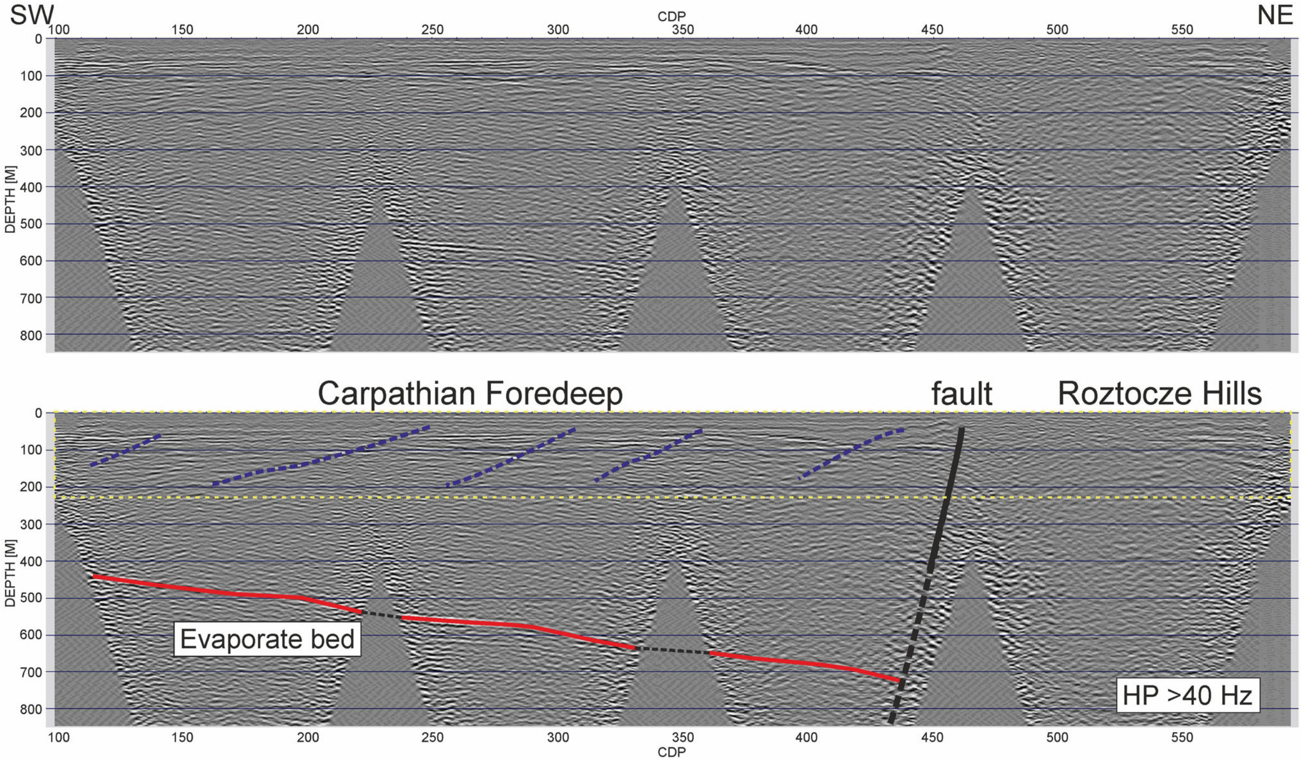


Fig. 9 Time domain stack with deep reflection enhanced processing. The data gaps at the deployment contact points are the effect of tail mute application to remove effect of surface waves. Clear deepening reflection at 450–700 ms marked with red line is visible along the profile. In shallow part between 50 and 200 ms several discontinuities of reflections are visible. Strong cut-of layering marked with black line is showing the Marginal Zone of the Carpathian Foredeep. Yellow rectangle marks the shallow area as in Fig. 7

recognize the maximum offset of clear observations, and design the survey for specific environment.

The horizon marked in red in Fig. 9 refers to the evaporate layers that occurs throughout the area of the Carpathian Foredeep and represents the best reference horizon for geophysical and geological research in this area. Such a evaporate layer is observed and well documented at much shallower depths in Ukraine.

Summarizing, we show that presented type of seismic survey with weight drop source and standalone stations can lead to a clear image of the structure from 30 down to 700 m, that could be performed effectively with low costs. The key to achieve detailed result is a careful data analysis using multiple techniques, that we hope to improve even further in the future.

Acknowledgements We would like to thank colleagues in the Imaging Department of IG PAS, as well as Editor Michał Malinowski for many suggestions that significantly improved the quality of this paper. This research was funded by National Science Centre, Poland (NCN) Grant UMO-2015/19/B/ST10/01833. Part of this work was supported within statutory activities no. 3841/E-41/S/2017 of the Ministry of Science and Higher Education of Poland.

References

Bąbel M (2004) Badenian evaporite basin of the northern Carpathian Foredeep as a drawdown salina basin. Acta Geol Pol 54(3):313–337

Bachrach R, Nur A (1998) High-resolution shallow seismic experiments in sand, part I: water table, fluid flow and saturation. Geophysics 63(1):1225–1233. https://doi.org/10.1190/1.1444423

Bruno PP, Improta L, Castiello A, Villani F, Montone P (2010) The Vallo di Diano fault system: new evidence for an active range-bounding fault in Southern Italy using shallow, high-resolution seismic profiling. Bull Seismol Soc Am 100(2):882–890

Bruno PP, Castiello A, Villani F, Improta L (2013) High-resolution densely spaced wide-aperture seismic profiling as a tool to aid seismic hazard assessment of fault-bounded intramontane basins: application to Vallo di Diano, Southern Italy. Bull Seismol Soc Am 103(3):1969–1980

Buker F, Green AG, Horstmeyer H (1998) Shallow seismic reflection study of a glaciated valley. Geophysics 63:1395–1407

Dusar M, Rijpens J, Sintubin M, Wouters L (2001) Plio-Pleistocene fault pattern of the Feldbiss fault system (southern border of the Roer Valley Graben, Belgium) based on high resolution reflection seismic data. Neth J Geosci 80:79–93

Feroci M, Orlando L, Balia E, Bosman C, Cardarelli E, Deidda G (2000) Some considerations on shallow seismic reflection surveys. J Appl Geophys 45:127–139

Francese RG, Hajnal Z, Schmitt D, Zaja A (2007) High resolution seismic reflection imaging of complex stratigraphic features in shallow aquifers. Memorie Descrittive della Carta Geologica d'Italia, LXXVI, pp 175–192

Green AG, Campbell FM, Kaiser AE, Dorn C, Carpentier S, Doetsch JA, Horstmeyer H, Nobes D, Campbell J, Finnemore M, Jongens R, Ghisetti F, Gorman AR, Langridge RM, McClymont AF (2010) International Conference on Environmental and Engineering Geophysics, p 15

Hobro JWD (1999) Three-dimensional tomographic inversion of combined reflection and refraction seismic travel-time data. Ph.D. Thesis, Department of Earth Sciences, University of Cambridge, Cambridge, UK

Jankowski L, Margielewski W (2015) Pozycja tektoniczna Roztocza w świetle historii rozwoju zapadliska przedkarpackiego. Biuletyn Państwowego Instytutu Geologicznego 462:7–28

Jaroszewski W (1977) Synsedymentacyjne przejawy mioceńskiej ruchliwości tektonicznej na Roztoczu Środkowym. Przegląd Geologiczny 25(8/9):418–427

Kondracki J (1994) Geografia Polski, mezoregiony fizyczno-geograficzne, PWN

Krzywiec P (2001) Contrasting tectonic and sedimentary history of the central and eastern parts of the Polish Carpathian Foredeep basin—results of seismic data interpretation. Mar Pet Geol 18:13–38

Kurkowski S (1998) Szczegółowa Mapa Geologiczna Polski 1:50 000, arkusz Józefów. Państwowy Instytut Geologiczny

Majdański M, Trzeciak M, Gaczyński E, Maksym A (2016) Seismic velocity estimation from post-critical wide-angle reflections in layered structures. Stud Geophys Geod 60:565–582

Malinowski M, Guterch A, Narkiewicz M, Probulski J, Maksym A, Majdański M, Środa P, Czuba W, Gaczyński E, Grad M, Janik T, Jankowski L, Adamczyk A (2013) Deep seismic reflection profile in Central Europe reveals complex pattern of Paleozoic and Alpine accretion at the East European Craton margin. Geophys Res Lett 40:1–6

Myśliwiec M (2004) Mioceńskie skały zbiornikowe zapadliska przedkarpackiego. Przegląd Geologiczny 52:581–592

Park CB, Miller RD, Xia J (1999) Multichannel analysis of surface waves. Geophysics 64:800–808

Pullan S, Hunter JA (1990) Delineation of buried bedrock valleys using the optimum offset shallow seismic reflection technique. In: Ward SH (ed) Geotechnical and environmental geophysics, vol. 3, Society of Exploration Geophysicists, no 5, pp 75–88

Robertsson JOA, Holliger K, Green AG, Pugin A, De Iaco R (1996) Effects of near-surface waveguides on shallow high resolution seismic refraction and reflection data. Geophys Res Lett 23:495–498

Roniewicz P, Wysocka A (1999) Charakterystyka sedymentologiczna utworów środkowomioceńskich północno-wschodniej, brzeżnej strefy zapadliska przedkarpackiego. Prace Państwowego Instytutu Geologicznego, CLXVIII, pp 83–97

Steeples DW (2000) A review of shallow seismic methods. Ann Geofis 43:1021–1030

Steeples DW, Miller RD (1990) Seismic reflection methods applied to engineering, environmental, and groundwater problems. In: Ward SW (ed) Geotechnical and environmental geophysics, vol 3, Society of Exploration Geophysicists, no 5, pp 1–30

Steeples DW, Miller RD (1998) Avoiding pitfalls in shallow seismic reflection surveys. Geophysics 63:1213–1224

Sugiyama Y, Mizuno K, Nanayama F, Sugai T, Yokota H, Hosoya T, Miura K, Takemura K, Kitada N (2003) Study of blind thrust faults underlying Tokyo and Osaka urban areas using a combination of high-resolution seismic reflection profiling and continuous coring. Ann Geophys 46:1071–1085

Vesnaver AL, Gohm G, Madrussani G, Petersen SA, Rossi G (1999) Tomographic imaging by reflected and refracted arrivals at the North Sea. Geophysics 64:1862–1952

Wysocka A (2006) Klastyczne utwory badeńskie Roztocza—przebieg sedymentacji w północnej marginalnej strefie basenu zapadliska przedkarpackiego. Przegląd Geologiczny 54:430–437

Wysocka A, Jasionowski M (2006) Polish part of Roztocze Hills. POKOS Materiały Konferencyjne 11–14

Wysocka A, Krzywiec P, Maksym A (2006) Kamieniołom Józefów. POKOS Materiały Konferencyjne 19–24

Wysocka A, Radwański A, Górka M, Bąbel M, Radwańska U, Złotnik M (2016) The middle miocene of the fore-Carpathian Basin (Poland, Ukraine and Moldova). Acta Geol Pol 66:351–401

Zelt CA (1999) Modelling strategies and model assessment for wide-angle seismic traveltime data. Geophys J Int 139:183–204

Acta Geophysica (2018) 66:191–201
https://doi.org/10.1007/s11600-018-0134-1

RESEARCH ARTICLE - APPLIED GEOPHYSICS

Generating porosity spectrum of carbonate reservoirs using ultrasonic imaging log

Jie Zhang[1,2] · Xin Nie[1,2,3] · Suyun Xiao[4] · Chong Zhang[1,2] · Chaomo Zhang[1,2] · Zhansong Zhang[1,2]

Received: 21 December 2017 / Accepted: 24 March 2018 / Published online: 31 March 2018

Abstract

Imaging logging tools can provide us the borehole wall image. The micro-resistivity imaging logging has been used to obtain borehole porosity spectrum. However, the resistivity imaging logging cannot cover the whole borehole wall. In this paper, we propose a method to calculate the porosity spectrum using ultrasonic imaging logging data. Based on the amplitude attenuation equation, we analyze the factors affecting the propagation of wave in drilling fluid and formation and based on the bulk-volume rock model, Wyllie equation and Raymer equation, we establish various conversion models between the reflection coefficient β and porosity ϕ. Then we use the ultrasonic imaging logging and conventional wireline logging data to calculate the near-borehole formation porosity distribution spectrum. The porosity spectrum result obtained from ultrasonic imaging data is compared with the one from the micro-resistivity imaging data, and they turn out to be similar, but with discrepancy, which is caused by the borehole coverage and data input difference. We separate the porosity types by performing threshold value segmentation and generate porosity–depth distribution curves by counting with equal depth spacing on the porosity image. The practice result is good and reveals the efficiency of our method.

Keywords Carbonate reservoir · Porosity spectrum · Well logging analysis · Ultrasonic imaging logging

List of symbols

A	Amplitude, L, m
f	Frequency, $1/t$, Hz
L	Propagation distance, L, m
$\bar{L}$	Average distance from the transducer to the borehole wall at the same depth, L, m
m	The ratio of number of pixels, n/n, 1
n	Unit volume particle number, n, 1
r	Radius of spherical particles, L, m
v	Wave propagation speed, L/t, m/s
v_0	Ultrasonic transmission speed in the drilling mud, L/t, m/s
V_{sh}	Shale volume content, L^3/L^3, %
Z	Wave impedance, m/L^2t, g/cm^3 m/s
α	Amplitude attenuation coefficient, L/L, 1
β	Reflection coefficient, L/L, 1
Δt	Near-borehole formation interval transit time, t, s
Δt_w	Water interval transit time, t, s
Δt_{ma}	Matrix interval transit time, t, s
Δt_{sh}	Shale interval transit time, t, s
λ	Wavelength of ultrasound, L, m
μ	Medium viscosity, mt/L^2, Pa s
μ_0	Kinematic viscosity of fluid, mt/L^2, Pa s
ρ	Density, m/L^3, g/m^3
ρ_0	Density of fluid, m/L^3, g/m^3
ρ_2	Near-bore formation density, m/L^3, g/m^3
ρ_{ma}	Density of matrix, m/L^3, g/m^3
ρ_s	Density of particles, m/L^3, g/m^3
ρ_w	Density of water/fluid, m/L^3, g/m^3
τ	Attenuation factor, L/L, 1
ϕ	Porosity, L^3/L^3, %
ϕ_0	Foreground average porosity, L^3/L^3, %
ϕ_1	Background average porosity, L^3/L^3, %

✉ Xin Nie
xin.nie@yangtzeu.edu.cn

[1] Key Laboratory of Exploration Technologies for Oil and Gas Resources (Yangtze University), Wuhan 430100, Hubei, China

[2] School of Geophysics and Oil Resource, Yangtze University, Wuhan 430100, Hubei, China

[3] Georgia Institute of Technology, Atlanta, GA 30332, USA

[4] Jianghan Oilfield Branch of Sinopec Group, Qianjiang 433124, Hubei, China

Introduction

There are two types of porosity in carbonate reservoirs including primary and secondary porosity, and the secondary porosity (includes vugs and fractures, etc.) whose structures and morphologies vary in different directions. Normally we use acoustic logging data to calculate the porosity. The usage of acoustic logging data is only limited to calculate the porosity through the wave velocity, not to reveal the pore structure property (Wang and Tao 2011; Wang et al. 2017). The conventional acoustic tools cannot provide the full azimuth information of the borehole. In practice, when the conditions of measurement and data acquired are limited, it is always a challenge to obtain the porosity distribution (Gu et al. 2017). Therefore, to accurately predict the production potential by conventional wireline logging was difficult. Accuracy has been greatly improved since micro-resistivity and ultrasonic borehole image logging was introduced to the industry. With the help of imaging logging, we can reveal the fracture density, orientation (e.g., dip and azimuth) and distribution of the fractures and caves visually in open hole. Based on the classic Archie saturation equation for the flushed zone, Newberry et al. (1996) proposed a method to convert micro-resistivity borehole images to porosity distribution of the formation. Using this method, we can obtain porosity spectrum which can be used in fine quantitative evaluation of the distribution of reservoir porosity. Predecessors have done much study on quantitatively evaluating the porosity distribution of carbonate reservoirs, and the application results were good (e.g., Hurley et al. 1998; Akbar et al. 2000; Tyagi and Bhaduri 2002, Xu et al. 2006; Ghafoori et al. 2009; Tetsushi et al. 2013). However, the micro-resistivity tool cannot cover the whole borehole wall, which causes white stripes between the pads. Therefore, is information is lost in the process, especially in the section where high-angle fractures and vugs develop. The ultrasonic image can cover 100% borehole wall, therefore, it can provide more complete information. However, the advanced usage of ultrasonic imaging logging data has been studied by very few. In this paper, we analyze the factors affecting the propagation of wave in drilling mud and formation, and propose a new method which is based on the amplitude attenuation equation. Using this method, we can calculate the near-borehole formation porosity distribution with ultrasonic borehole images and conventional wireline logs.

Analysis of influence factors on the ultrasonic imaging logging

Ultrasonic imaging tool uses a rotating transducer to launch ultrasonic pulse traveling through borehole fluid, receives the echo of the borehole wall, and records the echo amplitude and traveling time. Geological information, such as changes of lithology and physical properties, can be obtained using this tool. The measured echo amplitude is associated with the acoustic impedance and the shapes of the borehole wall. When the borehole diameter is “on gauge” (i.e. regular, without caves or key seats), the greater the acoustic impedance of the rock is, the larger the echo amplitude will be. Echo traveling time is associated with the geometric shape of borehole and the mud; as the mud is considered uniformly distributed in the borehole, it only reflects the status of borehole wall. The images reflect wave response from the surface of the borehole, thus can reveal the cracks and holes, similar to the micro-resistivity imaging.

The sound wave amplitude will decay exponentially with the changing of propagation distance in the medium. When passed through the medium interfaces, sound waves will result in refraction and reflection. Due to the difference of wave impedance, the reflection coefficient is different. According to the propagation distance and the differences of the two media, when the wave passes through the interface vertically, the reflection wave amplitude can be expressed as (Wang 2011):

$$A_1(f) = A_0(f) \times \beta \times \mathrm{e}^{-\alpha(f,\mu)\times L}, \tag{1}$$

where $A_1(f)$ is the reflection wave amplitude traveling from the interface of medium 1 and 2 for time t; $A_0(f)$ is the amplitude of the original wave launched by the acoustic transmitter in frequency f; β is the reflection coefficient; $\alpha(f, \mu)$ is the amplitude attenuation coefficient, which is a function of sound wave frequency f and medium viscosity μ; L is the propagation distance.

For an ultrasonic imaging logging tool, the acoustic transmitting frequency f is fixed; and at the same depth, borehole diameter is also a constant. Therefore, based on the analysis of Eq. (1), we identify two major factors that influence the ultrasonic echo amplitude: (1). attenuation coefficient α and (2) reflection coefficient β.

The influence of attenuation coefficient α

Ultrasonic attenuation process is complicated and difficult to analyze. One general equation of attenuation coefficient in fluid was presented by Urick (1948) and Urick and Ament (1949):

$$\alpha = \frac{2}{3}\pi r^3 n\left[\frac{1}{3}k^4 r^3 + k\left(\frac{\rho_\mathrm{s}}{\rho_\mathrm{f}} - 1\right)^2 \frac{S}{S^2 + \left(\frac{\rho_\mathrm{s}}{\rho_\mathrm{f}} - \tau\right)}\right], \tag{2}$$

where $k = \frac{2\pi}{\lambda}$, $S = \frac{9\sigma}{4r}\left(1 + \frac{\sigma}{r}\right)$, $\tau = \frac{1}{2} + \frac{9\sigma}{4r}$ and $\sigma = \frac{2\mu_0}{\rho_0 \omega}$.

In Eq. (2), r is the radius of spherical particles, m; ρ_s is the density of particles, g/m^3; ρ_0 is the density of fluid, g/m^3; μ_0 is the kinematic viscosity of fluid, Pa s; λ is the wavelength of ultrasound, m; n is the unit volume particle number; ω is the angular frequency, rad/s. According to Eq. (2), the influencing factors of attenuation coefficient α can be summed up into three classes: (a) micro-mud slurry: unit volume particle number n, radius of spherical particles r, density of particles ρ_s. (b) Macro-mud slurry: fluid density ρ_0, fluid kinematic viscosity μ_0. (c) Tool specification: tool transmitting frequency f, angular frequency ω, wavelength λ.

Therefore, it is obvious that the attenuation coefficient α just reflects the characteristics of the tool and drilling mud, but do not reflect the formation porosity.

The influence of reflection coefficient β

Z_1 and Z_2 are the wave impedance of drilling mud and formation, respectively; ρ and v are density and sound wave propagation speed. For the normal angle of incidence, their relationship with β can be shown as follows:

$$\beta = \frac{Z_2 - Z_1}{Z_2 + Z_1}. \tag{3a}$$

As Eq. (3a) shows, the main factors affecting reflection coefficient β are drilling mud and formation wave impedance Z_1 and Z_2. And the formation wave impedance Z_2 can be expressed as follows:

$$Z_2 = \rho_2 v_2 = \frac{\rho_2}{\Delta t_2}, \tag{3b}$$

where ρ_2 is the near-borehole formation density; Δt_2 is the near-borehole interval transit time. Both ρ_2 and Δt_2 can reflect formation porosity changes, so the reflection coefficient β can be used as a tool to obtain the formation porosity.

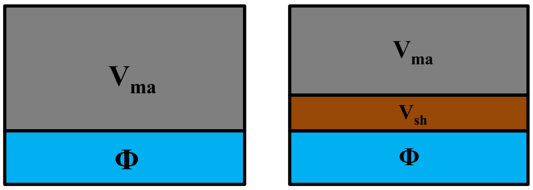
Porosity conversion models

Reflection coefficient around the well bore mainly reflects the information near the interface where the pore has been flushed by drilling mud filtrate. The pore fluid near the borehole can be considered to be pure brine. Therefore, two models of pure matrix (matrix built up only by pure carbonate minerals such as calcite or dolomite) and argillaceous matrix (matrix built up with both carbonate minerals and clay minerals) can be established (Fig. 1).

Wyllie time average equation

Pure matrix model (Fig. 1a) satisfies the Wyllie time average equation (Wyllie equation, Wyllie et al. 1956, 1958), which can be written as:

$$\begin{cases} \rho_2 = \rho_w \phi + (1 - \phi)\rho_{ma} \\ \Delta t_2 = \Delta t_w \phi + (1 - \phi)\Delta t_{ma} \end{cases}. \tag{4a}$$



(a) (b)

Fig. 1 Pure matrix model (**a**) and argillaceous matrix model (**b**)

Plug Eqs. (3b) and (4a) into Eq. (3a):

$$\beta = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{\rho_2 v_2 - Z_1}{\rho_2 v_2 + Z_1} = \frac{\rho_{ma} - Z_1\Delta t_{ma} + [(\rho_w - \rho_{ma}) - Z_1(\Delta t_w - \Delta t_{ma})]\phi}{\rho_{ma} + Z_1\Delta t_{ma} + [(\rho_w - \rho_{ma}) + Z_1(\Delta t_w - \Delta t_{ma})]\phi}. \tag{4b}$$

Let

$$a = (\rho_w - \rho_{ma}) - Z_1(\Delta t_w - \Delta t_{ma}); \quad b = \rho_{ma} - Z_1\Delta t_{ma};$$
$$c = (\rho_w - \rho_{ma}) + Z_1(\Delta t_w - \Delta t_{ma}); \quad d = \rho_{ma} + Z_1\Delta t_{ma};$$

Then

$$\beta = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{\rho_2 v_2 - Z_1}{\rho_2 v_2 + Z_1} = \frac{a}{c} + \frac{b - \frac{a}{c}d}{c\phi + d} \tag{4c}$$

Let

$$A = \frac{a}{c} \quad \text{and} \quad B = b - \frac{a}{c}d$$

Then

$$\beta = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{\rho_2 v_2 - Z_1}{\rho_2 v_2 + Z_1} = A + \frac{B}{c\phi + d} \tag{4d}$$

Based on Eq. (4d), we can reveal the relationship between reflection coefficient and porosity in both pure calcite and dolomite matrix (Fig. 2). As the figure shows: (a) when the drilling mud wave impedance is constant, the greater the porosity is, the smaller the reflection coefficient is; (b) when the porosity is greater than 0.15 in limestone with pure calcite matrix and 0.25 in dolostone with pure dolomite matrix, the calculated reflection coefficients are larger than the experimental value (Chelini et al. 1998).

Therefore, when the drilling mud wave impedance is known, the porosity can be calculated with the reflection coefficient based on this model.

Raymer equation

As Fig. 2 shows, when porosity is larger than a certain value, the reflection coefficients converted by the above model do not conform to the actual situation. Consequently, we combine rock volume model and Raymer conversion equation (Raymer equation, Raymer et al. 1980) to get the equations:

$$\begin{cases} \rho_2 = \rho_{\mathrm{w}}\phi + (1-\phi)\rho_{\mathrm{ma}} \\ v_2 = v_{\mathrm{w}}\phi + (1-\phi)^m v_{\mathrm{ma}} \\ \beta = \dfrac{Z_2 - Z_1}{Z_2 + Z_1} = \dfrac{\rho_2 v_2 - Z_1}{\rho_2 v_2 + Z_1} \end{cases} \tag{5}$$

Based on Eq. (5), the relationship between reflection coefficient and porosity can be calculated after a loop iteration (Fig. 3).

Argillaceous matrix model

Similar to the pure matrix model, argillaceous matrix model (Fig. 1b) satisfies the following equations:

$$\begin{cases} \rho_2 = \rho_{\mathrm{w}}\phi + V_{\mathrm{sh}}\rho_{\mathrm{sh}} + (1-\phi-V_{\mathrm{sh}})\rho_{\mathrm{ma}} \\ \Delta t_2 = \Delta t_{\mathrm{w}}\phi + V_{\mathrm{sh}}\Delta t_{\mathrm{sh}} + (1-\phi-\mathrm{V}_{\mathrm{sh}})\Delta t_{\mathrm{ma}} \end{cases}, \tag{6a}$$

where V_{sh} is the shale content. Plug Eq. (6a) into Eq. (3a):

$$\beta = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{\rho_2 v_2 - Z_1}{\rho_2 v_2 + Z_1} = \frac{\rho_{\mathrm{ma}} - Z_1\Delta t_{\mathrm{ma}} + (\rho_{\mathrm{sh}} - \rho_{\mathrm{ma}} - Z_1\Delta t_{\mathrm{sh}} + Z_1\Delta t_{\mathrm{ma}})V_{\mathrm{sh}} + [(\rho_w - \rho_{\mathrm{ma}}) - Z_1(\Delta t_{\mathrm{w}} - \Delta t_{\mathrm{ma}})]\phi}{\rho_{\mathrm{ma}} + Z_1\Delta t_{\mathrm{ma}} + (\rho_{\mathrm{sh}} - \rho_{\mathrm{ma}} + Z_1\Delta t_{\mathrm{sh}} - Z_1\Delta t_{\mathrm{ma}})V_{\mathrm{sh}} + [(\rho_w - \rho_{\mathrm{ma}}) + Z_1(\Delta t_{\mathrm{w}} - \Delta t_{\mathrm{ma}})]\phi}. \tag{6b}$$

After simplification

$$\beta = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{\rho_2 v_2 - Z_1}{\rho_2 v_2 + Z_1} = A + \frac{B'}{c\phi + d'}, \tag{6c}$$

where $A = \frac{a}{c}$ and $B' = b' - \frac{a}{c}d'$.

$$a = (\rho_w - \rho_{ma}) - Z_1(\Delta t_{\mathrm{w}} - \Delta t_{\mathrm{ma}}); \quad b' = \rho_{\mathrm{ma}} - Z_1\Delta t_{\mathrm{ma}} + (\rho_{\mathrm{sh}} - \rho_{\mathrm{ma}} - Z_1\Delta t_{\mathrm{sh}} + Z_1\Delta t_{\mathrm{ma}})V_{\mathrm{sh}};$$
$$c = (\rho_{\mathrm{w}} - \rho_{\mathrm{ma}}) + Z_1(\Delta t_{\mathrm{w}} - \Delta t_{\mathrm{ma}}); \quad d' = \rho_{\mathrm{ma}} + Z_1\Delta t_{\mathrm{ma}} + (\rho_{\mathrm{sh}} - \rho_{\mathrm{ma}} + Z_1\Delta t_{\mathrm{sh}} - Z_1\Delta t_{\mathrm{ma}})V_{\mathrm{sh}}.$$

Workflow of porosity spectrum calculation and porosity segmentation

The whole workflow of porosity spectrum calculation and porosity segmentation is shown in Fig. 4.

Calculation workflow of porosity spectrum

Using ultrasonic imaging data to build porosity spectrum can be divided into five steps:

Step 1: Based on Eq. (4b) or (5), use the mud wave impedance Z_1 and conventional density and acoustic wireline logs (e.g. DEN and AC) data to calculate formation reflection coefficient β'.

$$\beta' = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{\frac{\rho_2}{\Delta t_2} - Z_1}{\frac{\rho_2}{\Delta t_2} + Z_1}. \tag{7a}$$

Step 2: Introduce the echo response Eq. (1) and input transmitting amplitude A_0, received echo amplitude AMP_i and echo traveling time TIM_i at the ith point in the depth of total n points to calculate attenuation factor τ, and

$$\overline{\mathrm{LnAMP}} = \sum_{i=0}^{n} \ln(\mathrm{AMP}_i), \quad \overline{\mathrm{TIM}} = \sum_{i=0}^{n} \mathrm{TIM}_i \quad \text{and}$$
$$\overline{L} = \overline{\mathrm{TIM}} \times v_0.$$

Then, Eq. (1) can be turned into:

$$\frac{\overline{\mathrm{LnAMP}}}{\mathrm{LnA}_0} = \mathrm{Ln}\beta' - \alpha \times v_0 \times \overline{\mathrm{TIM}}, \tag{7b}$$

where $\overline{L}$ is the average distance from the transducer to the wall at the same depth; v_0 is the ultrasonic velocity in the mud.

Step 3: The τ can be set to be attenuation factor:

$$\tau = \alpha \times v_0. \tag{7c}$$

The attenuation factor τ at the depth can be obtained:

$$\tau = \frac{\mathrm{Ln}\beta' - \frac{\overline{\mathrm{LnAMP}}}{\mathrm{LnA}_0}}{\overline{\mathrm{TIM}}}. \tag{7d}$$

Step 4: Substitute τ into Eq. (1), and use AMP_i and TIM_i to calculate reflection coefficient β_i of the pixel.

$$\beta_i = \frac{\mathrm{AMP}_i}{A_0 \times \mathrm{e}^{-\tau \times \mathrm{TIM}_i}}. \tag{7e}$$

Step 5: Substitute β_i to Eqs. (4d), (5) or (6c), and then we can calculate the porosity value of each pixel which should be added up in equal depth spacing to draw porosity spectrum.

In addition, due to the porosity spectra transformation from ultrasonic images is based on rock volume model, the

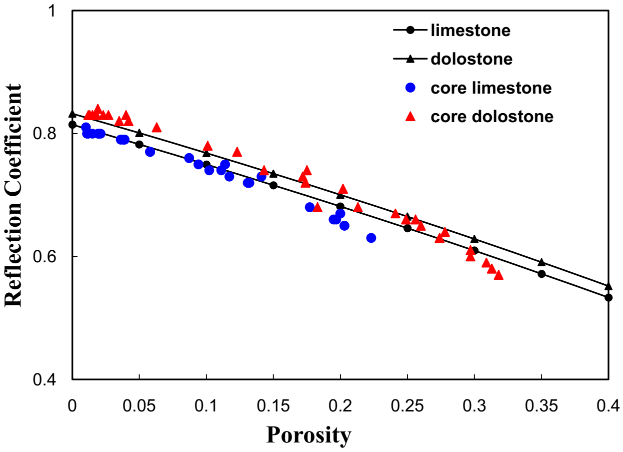


Fig. 2 The relationship between reflection coefficient and porosity in both pure calcite and dolomite matrix models deduced using Wyllie equation. The blue dots and red triangles are from experimental data, and the black curves with dots and triangles are from the conversion of Wyllie equation. Drilling mud wave impedance is 1850 g/cm^3 m/s

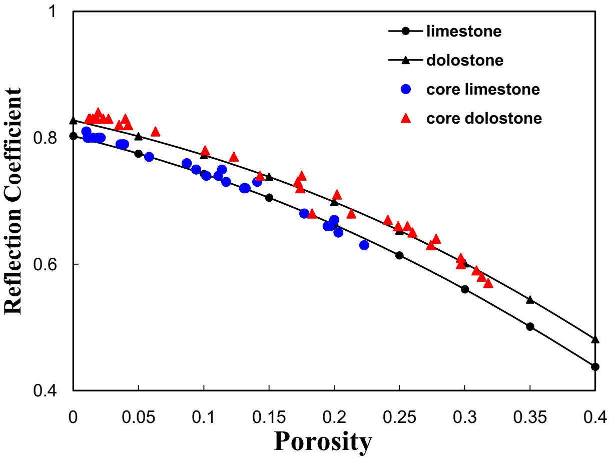


Fig. 3 The relationship between reflection coefficient and porosity in both pure calcite and dolomite matrix models deduced using Raymer equation. The blue dots and red triangles are from experimental data, and the black curves with dots and triangles are from the conversion of Raymer equation ($m = 2.5$). Drilling mud wave impedance is 1850 g/cm^3 m/s

porosity spectrum transformation model is only applicable under the conditions that the lithology is simple, and porosity distribution is near-homogeneous and porosity variety is small.

The segmentation of primary and secondary porosity

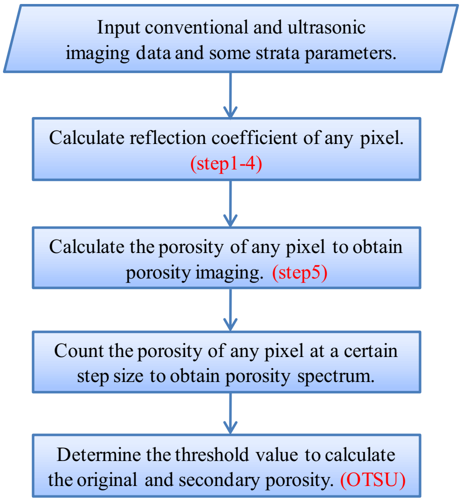


Fig. 4 The calculation workflow chart of obtain primary and secondary porosity from ultrasonic images

Porosity spectrum usually presents two or three separate peaks. Small apertures correspond with high cumulative frequency, large range of variation, while the large ones correspond with low cumulative frequency, small range of variation (Wang 2011). This phenomenon exactly reflects the state of primary pore and secondary pore in the underground rock. Thus, the frequency diagram can be used to distinguish original and secondary pore to obtain the two porosities, respectively. The methods to determine the threshold value between primary and secondary porosities mainly include artificial threshold division method (Wang 2011) and OTSU method (Otsu 1979).

OTSU method is an algorithm determining the image binarization threshold segmentation, which can execute image binarization segmentation to make the interclass variance between foreground and background image maximum.

Assuming the size of porosity image I is $M \times N$ pixels; secondary porosity (foreground) and primary porosity (background) segmentation threshold is t; the ratio of number of pixels in the foreground to the whole image is m_0; average porosity is ϕ_0; the ratio of number of pixels in the background and the whole image is m_1; average porosity is ϕ_1. The total average porosity of porosity image is ϕ_a; Variance between classes is g. Consequently, the next equation can be obtained:

$$g = m_0 m_1 (\phi_0 - \phi_1)^2. \tag{8}$$

Ergodic methods should be taken to make the interclass variance g largest to obtain the threshold t between the original and secondary porosity, average original porosity ϕ_0 and average secondary porosity ϕ_1.

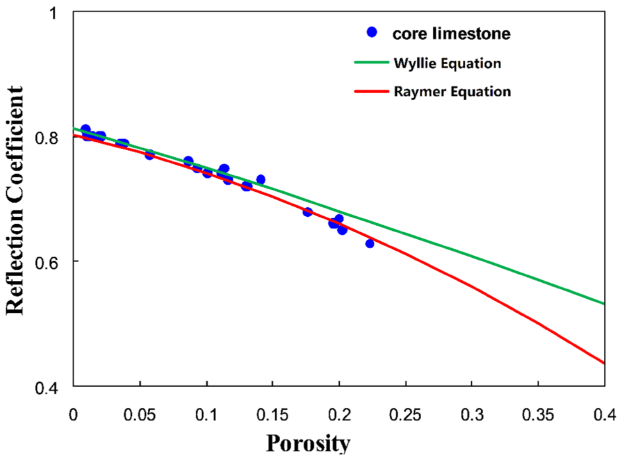


Fig. 5 Comparison diagram of the relationship between reflection coefficient and porosity deduced using Wyllie and Raymer equations in pure calcite matrix model. Blue dots are from experimental data (Chelini et al. 1998). Green and red curves are, respectively, obtained by Wyllie and Raymer equations ($m = 2.5$). Drilling mud wave impedance is 1850 g/cm^3 m/s

Comparison studies and applicable condition analysis

In this section, different application effects of the above three conversion models (i.e. Wyllie and Raymer equations, and argillaceous matrix model) are studied, and by processing real data, the effect of converted ultrasonic images to porosity images and spectra are verified by comparison to micro-resistivity imaging logging.

Comparison of conversion models

As Fig. 5 shows, when reflection coefficient is small, the gap between the porosity converted by Wyllie and Raymer equations is big. It indicates that when the porosity is less than 15%, the Wyllie and Raymer equations are both suitable for the calculation in limestone formation with pure calcite matrix, but when the porosity is more than 15%, the porosity calculated by Wyllie equations is larger than Raymer equations. It reveals that the Raymer equation has a wider range of application in porosity calculation. Figure 6 shows an actual case of the difference between the two models. In this interval, there are two vertical fractures. The color of cracks in the first porosity image is darker than that in the second, and that means the porosity values of the fractures calculated from Wyllie equation are

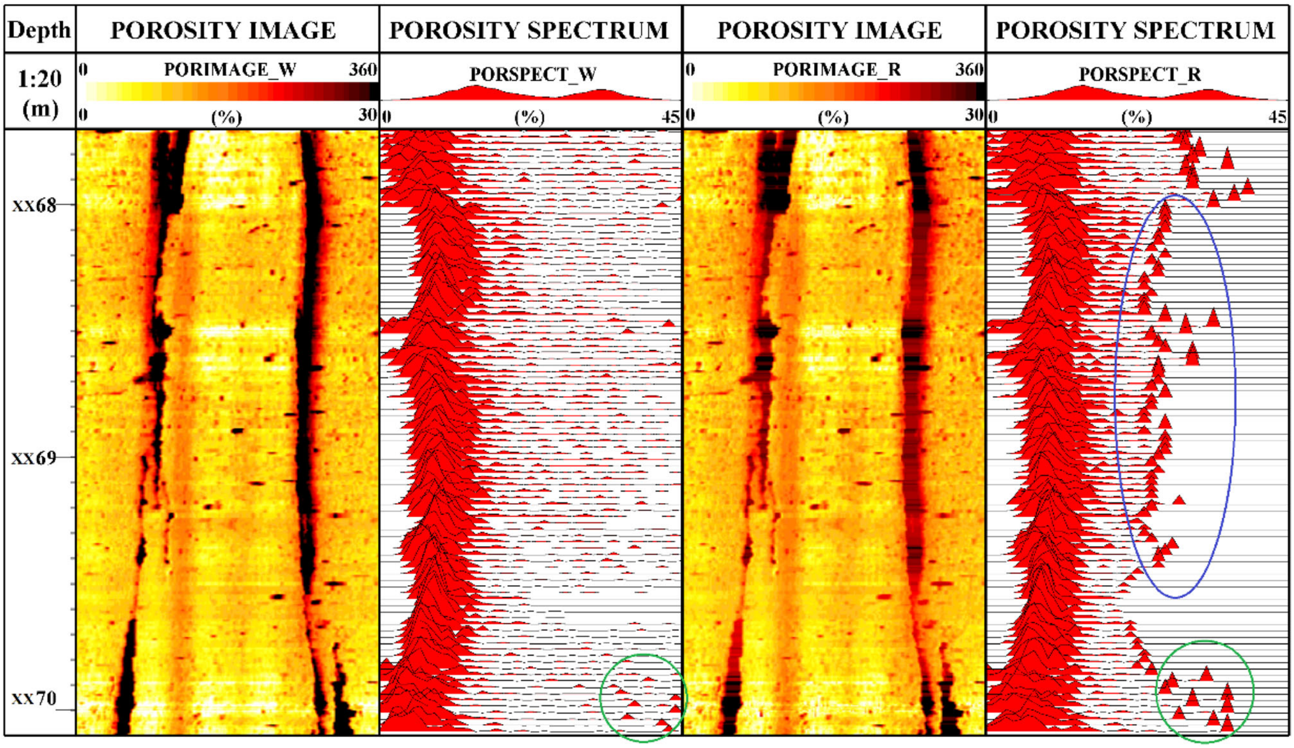


Fig. 6 The porosity images and spectra converted using Wyllie and Raymer models. From left: depth, porosity image computed using Wyllie equation, the spectrum of the Wyllie porosity image, porosity image computed using Raymer equation, and the spectrum of the Raymer porosity image

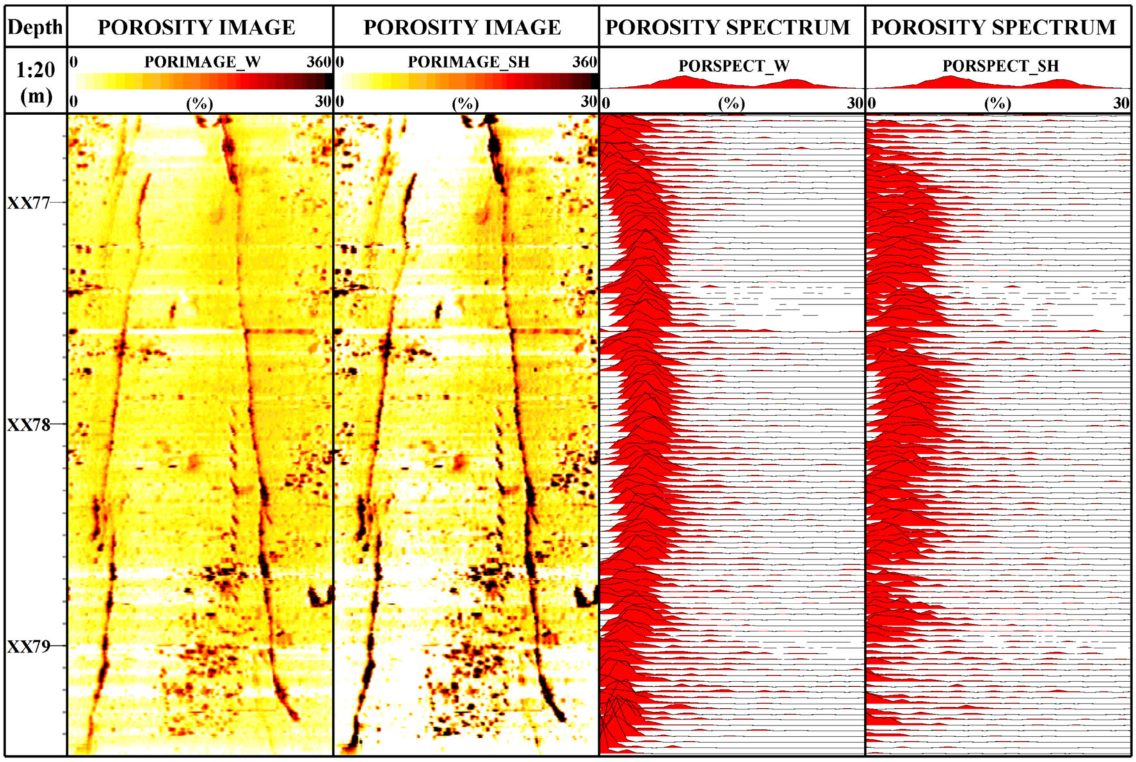


Fig. 7 The porosity images and spectra converted based on bulk-volume rock models without and with shale correction. From left: depth, porosity image without correction, porosity image with correction, the spectrum of the porosity image without correction, and the spectrum of the porosity image with correction

higher. This results in the porosity value of the peaks appearing in the right (circled by green ellipses) in the Raymer result are smaller than the ones in the Wyllie result. And because the higher fracture porosities in Wyllie result are more concentrated, the count is low. Thus, the peaks in the Raymer results (circled by blue ellipse) are not so obvious in the Wyllie result.

Comparison of pre- and post-shale correction

In addition, the models converted by Wyllie and Raymer equations are only suitable in the formations with pure calcite and dolomite matrix, but not with too much shale. Figure 7 shows the conversion results of models of bulk-volume rock with or without shale correction. In the porosity images, the shade of color represents the magnitude of porosity. After shale correction, the color of porosity image is lighter, and the spectra move left along with lower peaks. The shale content (V_{sh}) is calculated by natural gamma ray logs (GR) using follow equations (Larionov 1969):

$$\mathrm{SH} = (\mathrm{GR} - \mathrm{GR}_{\min})/(\mathrm{GR}_{\max} - \mathrm{GR}_{\min}), \tag{9a}$$

$$V_{sh} = (2^{\mathrm{gcur}\times \mathrm{SH}} - 1)/(2^{\mathrm{gcur}} - 1), \tag{9b}$$

where the SH is the original shale content, $\mathrm{GR}_{\min}$ and $\mathrm{GR}_{\max}$ are, respectively, the GR values of carbonate stone and pure mudstone formation. The gcur is the correction coefficient, which is 2 in older strata and 3.7 in newer strata.

Comparison of ultrasonic and micro-resistivity imaging logging

Thanks to the works of predecessors, the technology of converting micro-resistivity images to porosity images is proved to be mature and valid (Tyagi and Bhaduri 2002). In this section, we compare the conversion results of ultrasonic and micro-resistivity images. In Fig. 8, the secondary pores are very clear in ultrasonic porosity image, but not in the resistivity one. Anyway, the trends of them and the modes reflecting formation information are similar except the distribution range of ultrasonic spectrum is wider. It is conspicuous that the secondary porosities calculated by ultrasonic and resistivity spectra are near except the part circled by green ellipse. In this part, the secondary porosity of ultrasonic image is significantly larger than resistivity image because of the dissolved pores measured by ultrasonic imaging logging are more obvious (black dots in the image). Therefore, the method of converting

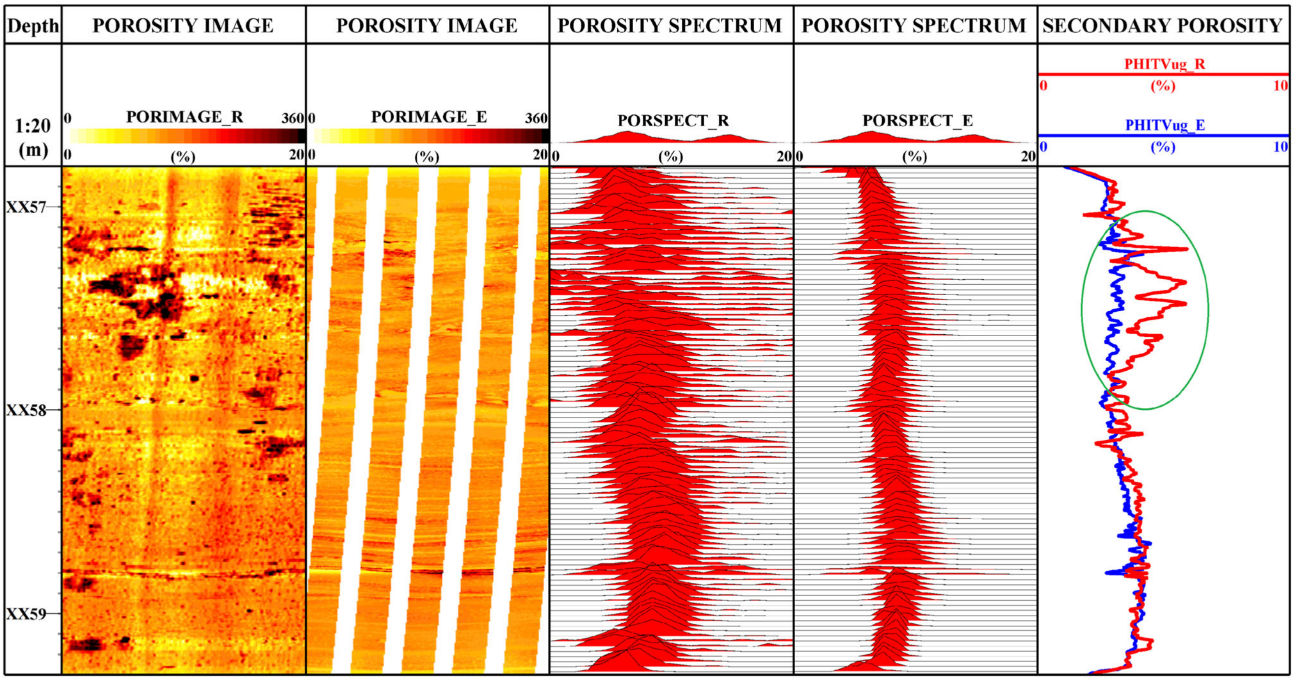


Fig. 8 The ultrasonic and micro-resistivity imaging porosity images and spectrum analysis. From left: depth, ultrasonic borehole wall porosity image, unfilled micro-resistivity borehole porosity image, the spectrum of ultrasonic porosity image, the spectrum of micro-resistivity porosity image, the secondary porosity curves calculated from the two images (red for ultrasonic and blue for micro-resistivity, the porosity range is 0–10%)

ultrasonic images to porosity images and spectra, even to calculate the primary and secondary porosity is workable.

Analysis of applicable conditions

The measurement of ultrasonic imaging logging tools is based on the transducer that is located in the center of a round borehole of which the caliper is regular. Due to the influence of the borehole shape and tool eccentricity, acoustic signal propagation time in the drilling mud will vary with location. Even if the well wall medium is uniform, it can also result in a heterogeneous imaging logging image. When some or the entire reflected acoustic wave cannot be received by the transducer, the echo amplitude will decline seriously, so that the imaging logging image will present vertical black stripes significantly.

Porosity obtained via ultrasonic imaging logs introduce echo amplitude and the echo time as important inputs, therefore its transformation effect will be affected by the borehole shape and tool eccentricity significantly. When the tool is in an elliptical borehole, ultrasonic logging image will appear as two dark black stripes, which will affect the identification of cracks; when the tool is not in the hole center, black vertical stripes will appear on imaging logging image, and they also appear on porosity image, which will hide the details on borehole wall, so porosity spectrum can show up abnormal, such as loss of spectrum, etc. Therefore, the quality of imaging logging is very important for applying this conversion method.

Case study

In this study, our data for application are from a well LXX in the Tianhuan Depression, Ordos Basin in Northwestern China. The application result is shown in Fig. 9.

Figure 9 is the porosity spectrum analysis results image, generated from the ultrasonic imaging log from the interval of XX43–XX63m in LXX well using the workflow presented in this paper. Porosity images in track 6 and track 7 can reflect the size of the porosity from the shade in the image. They can not only help us distinguish pore type of the reservoir, but also reveal the size change of the primary porosity intuitively and clearly. Track 8 is ultrasonic imaging spectrum transformed from pure calcite matrix porosity model. The forms of porosity spectrum curve are unimodal, bimodal or multi-peak, associated with the reservoir heterogeneity. Compared to resistivity porosity spectrum in track 9, the shape of ultrasonic porosity spectrum track 8 is similar, but its spectral peak is wider relatively. This may be caused by the lower sensitivity of the porosity to the wave attenuation.

The secondary porosity in depth XX47–XX48m is poor developed, so the porosity spectrum is narrower and

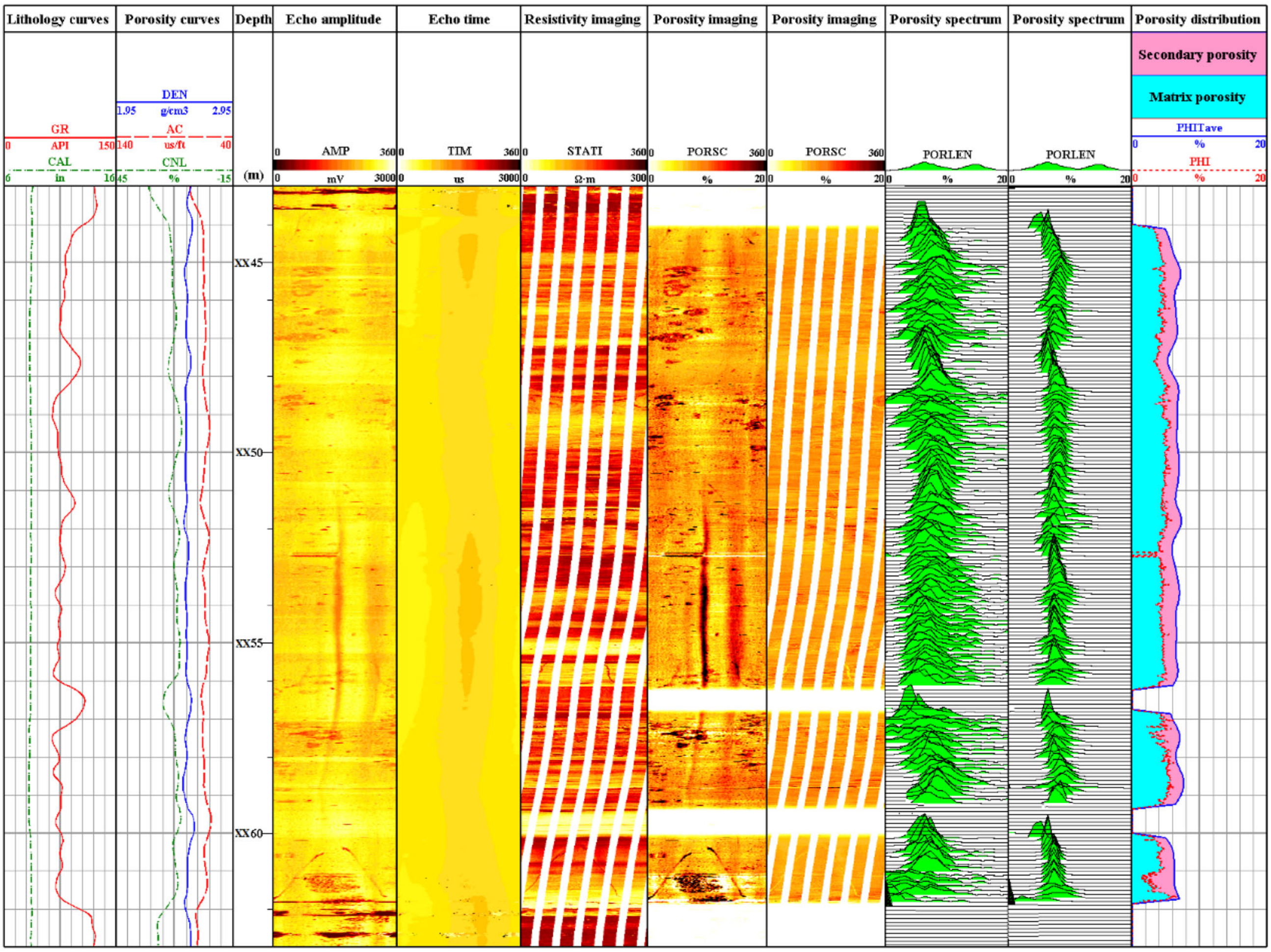


Fig. 9 Well LXX (XX43m–XX63m) porosity spectrum analysis results. Track 1: GR 0–150 API, CAL 6–16 in.; track 2: DEN 1.95–2.95 g/cm^3, AC 140–40 μs/ft, CNL 45–15%; track 3 and track 4 are ultrasonic echo amplitude and echo time image; track 5 is static resistivity imaging; track 6 and track 7 are ultrasonic imaging and resistivity porosity image, 0–20%; track 8 and track 9 are ultrasonic imaging and resistivity porosity image, 0–20%; track 10 is porosity distribution, PHI and PHITave represent the primary porosity and total average porosity, respectively, 0–20%

unimodal; In depth intervals XX45–XX47m, XX57–XX57.8m and XX61–XX61.6m, different scales of secondary pores resulted from dissolution are visible, therefore, the porosity spectrum shows wider bimodal or multimodal spectra with lower peaks; In depth interval XX60.3–XX61.6m high conductivity fractures caused by drilling mud invasion or argillaceous filling develop, and their porosity presents wider unimodal or bimodal spectra. In depth interval XX53.7–XX56.2m, there is not any vertical fracture in the resistivity image, but two dark vertical stripes are found in the ultrasonic image. That means the ultrasonic tool eccentric position may occur at this depth section. The stripes are mistaken for vertical fractures in analysis workflow, and that causes the ultrasonic porosity spectrum wider than resistivity porosity spectrum. This reveals that the measurement quality of ultrasonic imaging affect the effect of porosity spectrum directly.

Track 10 in Fig. 9 is primary and secondary porosity calculated by porosity image quantitatively based on OTSU method. It is obvious that the matrix porosity of the reservoir section is very stable, about 5–6%; Secondary porosity is about 2%, increasing obviously in intervals where different scale dissolved pores develop.

In general, the results show that the method is applicable and the application result depends on the logging quality.

Conclusions

In this paper, we provide a new way to evaluate the dual-porosity systems in carbonate formation using ultrasonic borehole images. This is a new integrated workflow to establish the relationship between the reflected wave amplitude and reflection coefficient which carry information about formation porosity. Based on models of bulk-volume rock and Raymer equation, three porosity conversion models suitable for pure and shaly matrix have been provided and verified. From the whole research, we can come to several conclusions as follows:

1. Based on bulk-volume rock models, we establish the conversion relationship between reflection coefficient of drilling mud and formation interface and the formation porosity, discovering that the reflection coefficient decreases in an inverse proportion function as the porosity increases.
2. Based on reflected wave amplitude attenuation equation and the bulk-volume rock models, we can use conventional well logging curves such as the formation density and acoustic interval transit time and ultrasonic imaging data such as echo amplitude and traveling time to get porosity distribution image and porosity spectrum.
3. OTSU method can be effectively used to determine the threshold of the primary and secondary porosity and precisely calculate the primary and secondary porosity of reservoir, which is meaningful to quantitative analysis of the fracture development degree of the reservoir.
4. The spectrum analysis method for ultrasonic imaging is similar to the one used in resistivity imaging. Under the same geological condition, the shapes of porosity spectra from ultrasonic and resistivity imaging logs are similar, but the spectrum peak range is relatively wider in results from ultrasonic imaging which may occur due to the lower sensitivity of the porosity to the wave attenuation.
5. Compared to micro-resistivity borehole images, ultrasonic borehole images can cover 100% of borehole wall, so it can provide more complete strata porosity information (e.g. vugs and fractures, etc.). However, because of the sensitivity of ultrasonic imaging tools, the imaging result can be affected by many factors such as the borehole shape condition or tool position. Therefore, the application condition of our workflow is very rigorous and we should consider the imaging quality first, and use the pore spectrum result with caution. It is better to use the spectra obtained from both ultrasonic and resistivity imaging logging data together to analyze the porosity.

Acknowledgements This research is supported by National Natural Science Foundation of China (Grant nos. 41504094 and 41404084). The authors thank the editor and the two reviewers for their constructional comments and suggestions. Dr. Xin Nie's visiting research in Georgia Institute of Technology is supported by China Scholarship Council (CSC).

Compliance with ethical standards

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

References

Akbar M, Chakravorty S, Russell SD, et al (2000) Unconventional approach to resolving primary and secondary porosity in Gulf carbonates from conventional logs and borehole images. In: SPE paper 87297-MS presented at the Abu Dhabi international petroleum conference and exhibition

Chelini V, Meazza O, and Verhoeff EK (1998) Relations between ultrasonic amplitude and petrophysical characteristics. In: SPE paper 50606-MS presented at the European petroleum conference

Ghafoori MR, Roostaeian M, Sajjadian VA (2009) Secondary porosity: a key parameter controlling the hydrocarbon production in heterogeneous carbonate reservoirs (case study). Petrophysics 50(1):68–78

Gu Y, Bao Z, Lin Y, Qin Z, Lu J, Wang H (2017) The porosity and permeability prediction methods for carbonate reservoirs with extremely limited logging data: stepwise regression vs. N-way analysis of variance. J Nat Gas Sci Eng 42:99–119. https://doi.org/10.1016/j.jngse.2017.03.010

Hurley NF, Zimmermann RA, and Pantoja D (1998) Quantification of vuggy porosity in a dolomite reservoir from borehole images and core, Dagger Draw field, New Mexico. In: SPE paper 49323-MS presented at the SPE international technical conference and exhibition

Larionov VV (1969) Radiometry of boreholes. NEDRA, Moscow **(in Russian)**

Newberry BM, Grace LM, and Stief DD (1996) Analysis of carbonate dual porosity system from borehole electrical images. In: SPE paper 35158-MS presented at the Permian basin oil and gas recovery conference

Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9:62–66. https://doi.org/10.1109/TSMC.1979.4310076

Raymer LL, Hunt ER, and Gardner JS (1980) An improved sonic transit time-to-porosity transform. In: SPWLA 21st annual logging symposium, paper P

Tetsushi Y, Daniel Q, Arnaud E et al (2013) Revisiting porosity analysis from electrical borehole images: integration of advanced texture and porosity analysis. In: SPWLA 54th annual logging symposium, paper E

Tyagi AK and Bhaduri A (2002) Porosity analysis using borehole electrical images in carbonate reservoirs. In: SPWLA 43rd annual logging symposium, paper KK

Urick RL (1948) The absorption of sound in suspensions of irregular particles. J Acoust Soc Am 20(3):283–289

Urick RL, Ament WS (1949) The propagation of sound in composite media. J Acoust Soc Am 21(1):62

Wang Z (2011) Application researching of the quantitative interpretation of borehole resistivity and acoustic imaging logging in fractured reservoir. Dissertation, China University of Petroleum, East China

Wang H, Tao G (2011) Wavefield simulation and data-acquisition-scheme analysis for LWD acoustic tools in very slow formations. Geophysics 76(3):E59–E68

Wang H, Fehler MC, Miller D (2017) Reliability of velocity measurements made by monopole acoustic logging-while-drilling tools in fast formations. Geophysics 82(4):D225–D233. https://doi.org/10.1190/geo2016-0387.1

Wyllie MRJ, Gregory AR, Gardner LW (1956) Elastic wave velocities in heterogeneous and porous media. Geophysics 21(1):41–70

Wyllie MRJ, Gregory AR, Gardner LW (1958) An experimental investigation of factors affecting elastic wave velocities in porous media. Geophysics 23(3):459–493

Xu C, Richter P, Russell D, Gournay J (2006) Porosity partitioning and permeability quantification in vuggy carbonates using wireline logs, Permian basin, west Texas. Petrophysics 47(1):13–22

Acta Geophysica (2018) 66:203–222
https://doi.org/10.1007/s11600-018-0121-6

RESEARCH ARTICLE - HYDROLOGY

Applicability of TOPMODEL in the mountainous catchments in the upper Nysa Kłodzka river basin (SW Poland)

Justyna Jeziorska[1,2] · Tomasz Niedzielski[1]

Received: 14 September 2017 / Accepted: 9 February 2018 / Published online: 9 March 2018

Abstract

River basins located in the Central Sudetes (SW Poland) demonstrate a high vulnerability to flooding. Four mountainous basins and the corresponding outlets have been chosen for modeling the streamflow dynamics using TOPMODEL, a physically based semi-distributed topohydrological model. The model has been calibrated using the Monte Carlo approach—with discharge, rainfall, and evapotranspiration data used to estimate the parameters. The overall performance of the model was judged by interpreting the efficiency measures. TOPMODEL was able to reproduce the main pattern of the hydrograph with acceptable accuracy for two of the investigated catchments. However, it failed to simulate the hydrological response in the remaining two catchments. The best performing data set obtained Nash–Sutcliffe efficiency of 0.78. This data set was chosen to conduct a detailed analysis aiming to estimate the optimal timespan of input data for which TOPMODEL performs best. The best fit was attained for the half-year time span. The model was validated and found to reveal good skills.

Keywords TOPMODEL · Hydrologic model · Discharge · Poland · Kłodzko land

Introduction

Better understanding of watershed dynamics is one of the key factors in solving water-related scientific and practical problems. This role becomes crucial for effective planning and management of water resources (Beven and Freer 2001a; Bastola et al. 2008) in areas endangered by extreme hydrological events. Key variables influencing the hydrological response of the catchment need to be estimated based on data recorded on gauges, and the limitations occur when data are poor or insufficient. Thus, the hydrological modeling of water cycle components is the essential tool which becomes necessary for extending hydrological data both in space and time (Bastola et al. 2008). Many researchers attempted to develop solutions that would model the complexity of processes and heterogeneity of factors influencing the hydrological system dynamics. The availability of spatial characteristics of the catchment that rose with the advent of geographic information systems (GIS) shifted the research interests from the traditional lumped models towards more complicated, distributed ones. The advantage of the latter is the possibility of having a spatial pattern of the modeling outputs, such as soil moisture or saturation zone extent (Sun and Deng 2004). One of the semi-distributed and physically based conceptual models is the TOPography-based hydrological MODEL, known also as TOPMODEL (Beven and Kirkby 1979; Beven et al. 1995; Beven and Freer 2001a). It combines the advantages of two above-mentioned approaches. The number of parameters is reduced, but they maintain their physical interpretation. Simplified model structure diminishes the data requirements. Thus, this conceptual model integrates the ability to simulate the spatial distribution of its results at any time step (Choi and Beven 2007) with the computational efficiency that allows multiple simulations (Peters et al. 2003). These features

---

✉ Justyna Jeziorska
jajezior@ncsu.edu

Tomasz Niedzielski
tomasz.niedzielski@uwr.edu.pl

[1] Faculty of Earth Sciences and Environmental Management, Institute of Geography and Regional Development, University of Wrocław, Wrocław, Poland

[2] Center for Geospatial Analytics, North Carolina State University, Raleigh, USA

contribute to the model popularity and successful applications in numerous studies. Furthermore, recent increase of the TOPMODEL application is caused by access to more detailed data describing the watershed characteristics.

According to Beven et al. (1995), the TOPMODEL concept should not be considered as a hydrologic modeling package, but rather as a set of conceptual tools that can describe the catchment behavior. Hence, since its introduction in 1979 (Beven and Kirkby 1979), many versions have been developed and numerous studies have applied the TOPMODEL approach to a wide range of hydrology-related topics. The research problems investigated with the use of TOPMODEL include: flood frequency analysis (Beven 1986; Cameron et al. 1999), scaling theory in hydrology (Wood et al. 1988), examination of the influence of the digital elevation model (DEM) resolution on the simulation results (Brasington and Richards 1998), analysis of climate change scenarios (Romanowicz 2007), water table estimation (Merot et al. 1995; Moore and Thompson 1996; Lamb et al. 1997), testing the applicability to water quality problems (Wolock et al. 1990; Robson et al. 1992), and uncertainty analysis (Freer et al. 1996; Choi and Beven 2007; Bastola et al. 2008; Fisher and Beven 1996).

Although the initial TOPMODEL applications concentrated on examining the catchment dynamics in the humid temperate climate zone in the UK (Beven and Kirkby 1979; Beven et al. 1984; Quinn and Beven 1993), in the eastern USA (Beven and Wood 1983; Hornberger et al. 1985), the capability of providing good simulation results have been proven in the variety of environments in basins all over the globe. TOPMODEL has been successfully used in temperate cold climate in Norway (Lamb et al. 1997). Furthermore, the model performance has been investigated in drier Mediterranean regimes (Durand et al. 1992; Piñol et al. 1997), and in the monsoon region of China (Chen and Wu 2012). The research was also carried out in the humid temperate (Cameron et al. 1999; Bárdossy 2007; Choi and Beven 2007; Furusho et al. 2014) and Mediterranean (Gallart et al. 2008) climate zones. TOPMODEL has also been successfully utilized in the tropical climate zone in French Guiana (Molicova et al. 1997) and within the same climate zone, but in its antipodes, in one of the most hydrologically responsive forested headwater catchments, in Maimai in New Zealand (Freer et al. 2003). Sigdel et al. (2011) applied the model to watersheds in the Bagmati River basin in Nepal. It is worth noting that many researchers focused their studies on the TOPMODEL applications in the Nepal region: Brasington and Richards (1998) applied the TOPMODEL to a small headwater catchment in the Nepal Middle Hills, Shrestha et al. (2007) investigated its performance in different physiographic regions of Nepal, and Bastola et al. (2008) conducted a comparative study of 26 catchments across the globe, including 4 basins in Nepal. The vast majority of the TOPMODEL applications concern small- or medium-sized catchments (up to several dozen of square kilometers); however, the model demonstrated good performance for: very small basins of 0.75 ha (Lamb et al. 1997) and of 1.57 ha (Molicova et al. 1997) as well as very large basins of over 25 000 km^2 (Sun and Deng 2004; Chen and Wu 2012). High attention has been paid to the TOPMODEL capability of rainfall-runoff modeling in the mountainous regimes. Holko and Lepistö (1997) applied the model to the Jalovecky Creek catchment in Western Tatra Mountains, Blazkova and Beven (1997) investigated mountain wetlands in the Czech-Moravian highlands, and the research area for Nourani and Mano (2007) was watershed located 800–2178 m a.s.l. in the western Iran.

The model has been successfully applied to several catchments in Central Europe. Emphasis should be placed on the above-mentioned research of Holko and Lepistö (1997), focusing on the mountainous watershed in Slovakia, and on the investigation by Bárdossy (2007) into 16 lowland catchments in the German part of the Rhine basin. Applications of TOPMODEL for Polish basins were rare and focused rather on selected episodes than on the entire hydrologic years (Table 1). Szalińska et al. (2014) applied the model in question to simulate discharges in selected gauges installed within upper Nysa Kłodzka and Soła basins, with the emphasis placed only on selected high-flow events. Recent developments carried out in the Institute of Meteorology and Water Management—National Research Institute, Poland (Instytut Meteorologii i Gospodarki Wodnej—Państwowy Instytut Badawczy, IMGW-PIB) include the Hydropath framework, the platform that includes TOPMODEL (Orczykowski and Tiukało 2016). More recently, Niedzielski and Miziński (2017) used TOPMODEL as one of ensemble members in the multimodel hydrologic ensemble prediction, within the HydroProg system implemented in Kłodzko Land.

Even though these studies demonstrated good modeling results, the TOPMODEL performance has never been tested in Poland in longer time perspectives than single high-flow episodes. To fill this gap and fulfill a need of detailed analyses of the model in a variety of environments, postulated by Durand et al. (1992), in this study, TOPMODEL is applied to a few medium-sized catchments in the flood-prone areas of the Sudety Mountains in SW Poland. Hence, the aim of the modeling experiment was to test the applicability of the TOPMODEL for the purpose of discharge simulation at the outlets of four contributing catchments of Nysa Kłodzka river (gauges: Bystrzyca Kłodzka, Kłodzko, Bardo) and the river of Biała Lądecka (gauge: Żelazno). In addition, the aim of the study is also to estimate the time span of data (discharge/rainfall/evapotranspiration) for

Table 1 Applications of TOPMODEL for Polish basins and their sampling intervals

Basin	River	Institution	References	Data	Sampling interval
Nysa Kłodzka	Nysa Kłodzka (above Miedzylesie)	IMGW-PIB	Szalińska et al. (2014)	Selected episodes from 2010 to 2012	1 h
	Nysa Kłodzka (above Bystrzyca Kłodzka)	UWr	This paper	Hydrologic years 2009–2012	15 min
	Nysa Kłodzka (above Kłodzka)	UWr	Niedzielski and Miziński (2017)	Selected episodes from 2014 to 2015	15 min
	Nysa Klodzka (above Kłodzka)	UWr	This paper	Hydrologic years 2009–2012	15 min
	Nysa Kłodzka (above Bardo)	UWr	This paper	Hydrologic years 2009–2012	15 min
	Bystrzyca (outlet unknown)	IMGW-PIB	Szalińska et al. (2014)	Selected episodes from 2010 to 2012	1 h
	Biała Lądecka (outlet unknown)	IMGW-PIB	Szalińska et al. (2014)	Selected episodes from 2010 to 2012	1 h
	Biała Lądecka (above Żelazno)	UWr	This paper	Hydrologic years 2009–2012	15 min
	Bystrzyca Dusznicka (above Szalejów Dolny)	IMGW-PIB	Szalińska et al. (2014)	Selected episodes from 2010 to 2012	1 h
Sola	Sola (above Rajcza)	IMGW-PIB	Szalińska et al. (2014)	Selected episodes from 2010 to 2012	1 h
	Woda Ujsolska (outlet unknown)	IMGW-PIB	Szalińska et al. (2014)	Selected episodes from 2010 to 2012	1 h
	Zabniczanka (above Żabnica)	IMGW-PIB	Szalińska et al. (2014)	Selected episodes from 2010 to 2012	1 h
	Bystra (outlet unknown)	IMGW-PIB	Szalińska et al. (2014)	Selected episodes from 2010 to 2012	1 h
	Koszarawa (outlet unknown)	IMGW-PIB	Szalińska et al. (2014)	Selected episodes from 2010 to 2012	1 h
Kamienna	Kamieinna (above Jakuszyce)	IMGW-PIB	Orczykowski and Tiukało (2016)	Unknown choice of data span	1 h

IMGW-PIB Instyhit Mereorologii i Gospodarki Wodnej—Panstwowy Instytut Badawczy (Institute of Meteorology and Water Management—National Research Institute), *UWr* Univtersytet Wroclawski (University of Wroclaw)

which the TOPMODEL performance is optimal. Therefore, the novelty of the work presented in this paper is twofold: (1) it presents new results on discharge modelling in the upper Nysa Kłodzka river basin in time scales of a few hydrologic years, hence not only for selected high-flow episodes; (2) it shows first attempts to relate modelling skills with time span of input data.

The paper is organized as follows: the next section presents a simplified description of the model (its concept and features), based on the extended explanations of TOPMODEL theory given by Beven and Wood (1983) and Beven (1986). The second section also contains overview of methods used for the model assessment. The study area is presented in the third section, followed by the fourth section that focuses on description of data used in the research. The results of land surface parameterization, model calibration, sensitivity analysis, and validation of the TOPMODEL for four contributing basins located within the upper Nysa Kłodzka basin (SW Poland) are included in the fifth section. They are concluded in the last section, which also offers an overview of potential future research activities.

Methods

Concept of TOPMODEL

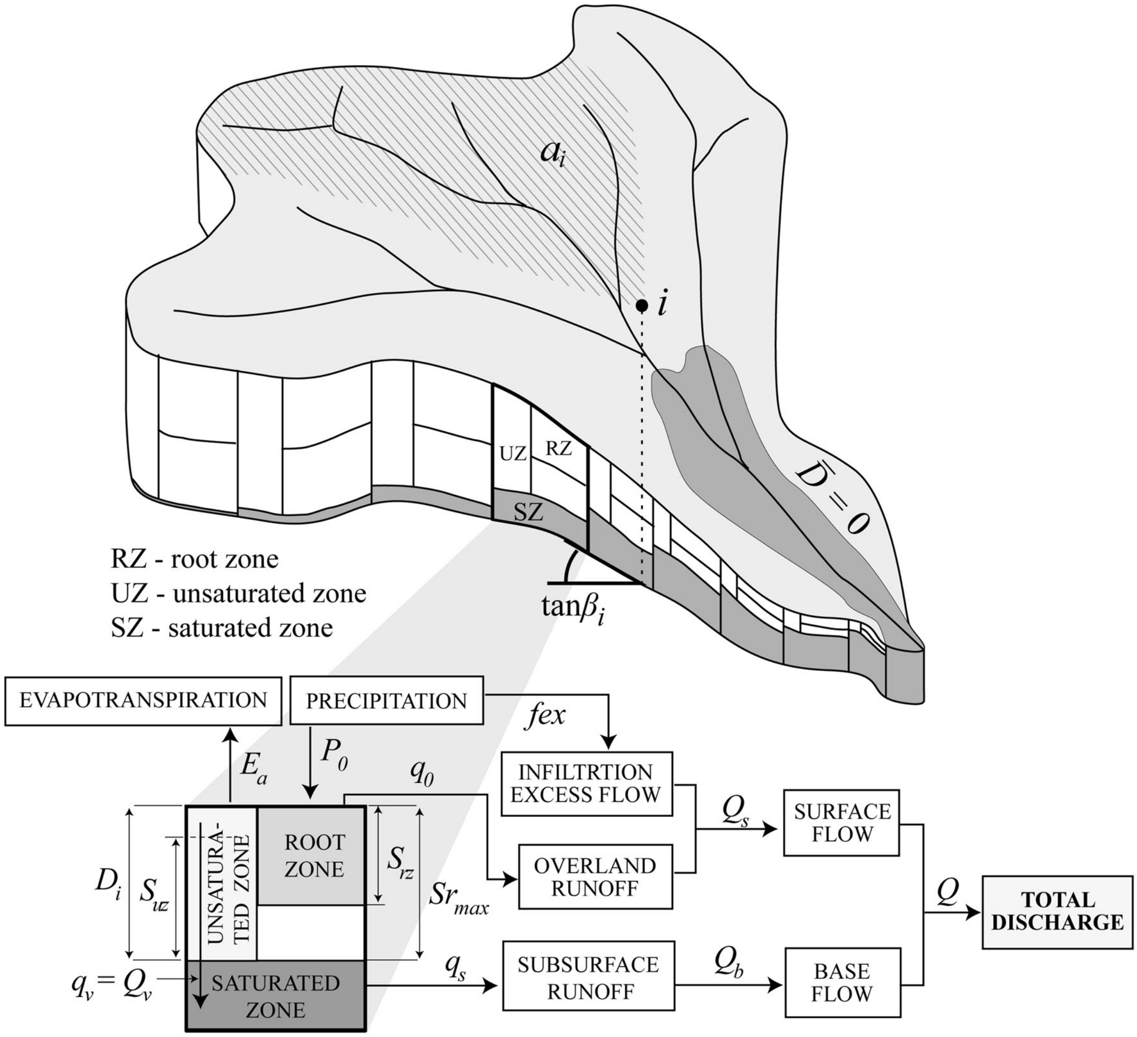
The idea that runoff is primarily a result of overland flow generated by rainfall when infiltration capacity of the soil is exceeded is known as Hortonian theory of infiltration excess overland flow. However, in recent studies, it is commonly superseded by the contrary concept that emphasizes the significance of saturation-excess overland flow and subsurface runoff generation. On the contrary to the Hortonian concept, stating that the occurrence of surface overland flow is possible when the soil is not fully saturated, the saturated-excess overland flow theory assumes that the overland flow is generated when the soil is fully saturated to the surface or if subsurface flow returns to the surface in saturated areas (Nourani et al. 2011). Among different applications of this concept in hydrological modeling, one of the most widely used is TOPMODEL (Beven and Kirkby 1979; Beven 1997; Beven and Freer

2001a). The predominant factors influencing the discharge generation in the model are topography of the basin and soil characteristics (Franchini et al. 1996). The topography is quantitatively expressed by the topographic index (also known as Topographic Wetness Index, TWI). Its value is computed from the basin topography using the following expression:

$$\mathrm{TWI}_i = \ln \frac{a_i}{\tan \beta_i}, \tag{1}$$

where a_i is the upslope contributing area of ith basin and $\tan \beta_i$ is slope of the ground surface of this basin. Upslope contributing area (Fig. 1) represents the area that can potentially produce runoff to the location of interest, i.e., to the outlet from the contributing basin (Erskine et al. 2006). In the raster representation of the terrain, it should be replaced by the upslope drainage area per unit of contour length (Moore and Wilson 1992; Moore and Burch 1986; Desmet and Govers 1996), which is equivalent to DEM grid cell size (Mitasova et al. 1996). Areas associated with high TWI values tend to saturate first and will, therefore, constitute potential subsurface or surface contributing areas (Gumindoga 2010; Beven 1997).

TWI refers to variable source area concept of runoff generation (Hewlett and Hibbert 1967) and is based on the following three simplifying assumptions regarding the



Q_v – total flux, q_v – the flux of water entering the water table locally (per unit area), A_i – area associated with topographic index class, E_a – actual evapotranspiration, Sr_{max} – maximum root zone deficit, S_{rz} – root zone deficit, S_{uz} – local water storage in the unsaturated zone, q_0 – overland runoff, P_0 – precipitation per unit width, fex – infiltration excess flow, Q_s – surface flow, Q_b – base flow, q_sz – subsurface saturated zone flux, Q – total discharge

Fig. 1 Basic concept of TOPMODEL scheme. Based and combined from schemes by Nourani et al. (2011); Franchini et al. (1996); Gumindoga (2010) and Fisher and Beven (1996) hydrologic system (Nourani et al. 2011; Gumindoga 2010; Brasington and Richards 1998; Beven et al. 1995; Holko and Lepistö 1997):

- dynamics of the saturated zone can be approximated by successive steady-state representations;
- hydraulic gradient of the saturated zone can be approximated by the local surface topographic slope ($\tan \beta_i$), and thus, the groundwater table and saturated flow are parallel to the local surface slope;
- distribution of downslope transmissivity with depth is an exponential function of storage deficit or depth to the water table.

This approach implies that all points with the same value of TWI respond in the same way (Fisher and Beven 1996). Calculations need to be performed only for representative values of the index, what greatly simplifies the procedure and reduces the computational cost while maintaining the capability of the identification of water table levels and soil moisture within the catchment (Chairat and Delleur 1993; Fisher and Beven 1996). The results may be mapped back into space using knowledge of the pattern of TWI derived from a topographic analysis (Beven 1997).

Sigdel et al. (2011) pointed out that the above assumptions may be valid for small and medium catchments, with shallow soils and moderate topography, which do not experience excessively long dry periods. The quasi-steady-state dynamics concept has been criticized (Barling et al. 1994; Beven 1997; Peters et al. 2003), and it cannot be always safely accepted (Beven 1997).

According to the TOPMODEL concept, there are two main factors that account for runoff generation, namely the catchment topography and the transmissivity that diminishes with depth (Beven and Kirkby 1979). A soil column in TOPMODEL is defined as a set of three stores: root zone, unsaturated zone, and saturated zone. They behave like three interdependent repositories (Fig. 1). The detailed explanation of physical processes taken into account with TOPMODEL calculations can be found in Peters et al. (2003), Taschner (2003), Sun and Deng (2004) and Sigdel et al. (2011).

The extended interpretation of the TOPMODEL theory is given by Beven and Wood (1983) and Beven (1986).

Model performance measures

Assessment of the efficiency of the hydrological model is necessary not only for the estimation of its ability to reproduce catchment behavior, but also for modifying model structure. Apart from the subjective visual inspection of the simulated and observed hydrographs, there are numerous statistical measures which can be used for hydrological model assessment (Krause et al. 2005). In this study, two model performance measures have been chosen: root-mean-square error (RMSE) and Nash–Sutcliffe efficiency (NSE). The latter was created specifically to assess hydrological models (Nash and Sutcliffe 1970) and can be calculated using the formula:

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{N} (\mathrm{Qobs}_t - \mathrm{Qsim}_t)^2}{\sum_{t=1}^{N} (\mathrm{Qobs}_t - \overline{\mathrm{Qobs}_t})^2}, \quad (2)$$

where Qobs_t is observed discharge, Qsim_t is simulated discharge at time step t, and N is number of observations/simulations. Although its value has been questioned (Criss and Winston 2008), it is still most widely used criterion for evaluating TOPMODEL performance (Beven and Binley 1992). The main critic relates to its sensitivity to extreme values and, since the hydrological data often contain outliers, the measure can be misleading. Values of NSE vary from $-\propto$ (strong misfit) to 1 (perfect fit), and the situation $\mathrm{NSE} = 0$ occurs when model predictive skills are similar to performance of extrapolated mean of observations. According to Moriasi et al. (2007), there are several classifications of NSE and the corresponding interpretations, with a range of NSE intervals to describe a satisfactory model performance. Since the intervals vary significantly, NSE values greater than 0.6 are hereinafter assumed to describe a satisfactory fit (Beven and Freer 2001b).

Study area

The TOPMODEL performance was investigated using data from four gauges located in the Sudety Mountains in upper Nysa Kłodzka basin (SW Poland and the border region of Czech Republic). The following reasons led to the choice of the study area.

- The present research is associated with the HydroProg system (Niedzielski et al. 2014) in which TOPMODEL is used for predicting river stages in real time.
- Rainfall and water level data sets are available, and the access to date is courtesy of the authorities of Kłodzko County, the owner of the Local System for Flood Monitoring (Lokalny System Osłony Przeciwpowodziowej, LSOP).
- TOPMODEL has not been applied in the upper Nysa Kłodzka basin so far.
- TOPMODEL has been already shown to work well in small mountainous catchments in different parts of the world (e.g. Bastola et al. 2008; Cameron et al. 1999).

To investigate the performance of the model in different conditions, four gauges and the associated contributing basins were delimited (Bardo, Kłodzko, Bystrzyca Kłodzka, and Żelazno). The main features of the watersheds are presented in Fig. 2 and listed in Table 2.

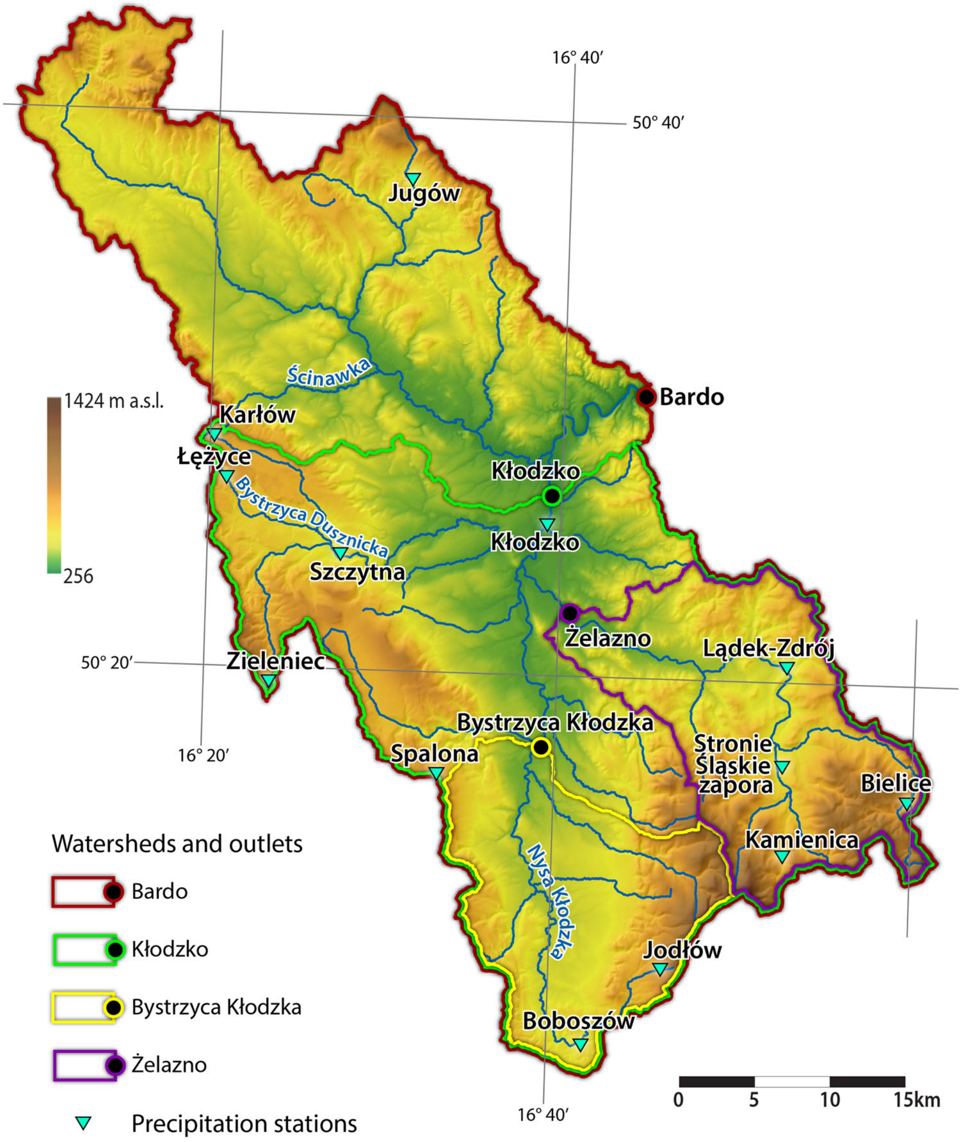


Fig. 2 Investigated watersheds and precipitation stations

Table 2 Main characteristics of investigated watersheds

Catchment outlet (gauge)	Bardo	Klodzko	Zelazno	Bygtrzyca Klodzka
Altitude (m a.s.l.)	259	285	341	319
Area (km^2)	1744.1	1079.31	259.985	302.94
Mean altitude (m a.s.l.)	534.79	567.45	576.53	648.06
Elevation difference within the catchment (m)	1165	1139	1024	1106
Mean slope (%)	8.09	8.24	7.45	11.17
Flow distance from source (km)	78	58	47	36
Mean TWI	7.37	6.91	6.81	6.34
River	Nysa Kłodzka	Nysa Kłodzka	Biała Lądecka	Nysa Kłodzka
Mean observed flow (m^3/s)	19.19[a]	13.31[a]	4.8[a]	4.29[a]

[a] Based on: http://bip.umwd.pl/fileadmin/user_upload/woda_i_melioracja/Program_Malej_Retencji_Wodnej_tekst_jednolity_10_2006.pdf

The Sudety Mountains are a medium–high mountain range spread along the Polish-Czech boundary in Central Europe. A maximum elevation of the mountains is equal to 1603 m a.s.l. Sudety Mountains are an example of fault-block mountains which are characterized by fault-generated mountain fronts and structural basins attributed to up- and downfaulting in the late Cenozoic (Migoń and Placek 2014). Diverse tectonic structure is additionally enhanced by the lithological complexity. Highly complex and diversified landscape is a result of the mosaic of underlying geology and its polygenetic origin (Wieczorek and Migoń 2014). The structure has also significant impact on extreme events: meteorological, hydrological, and geomorphological ones (Migon and skutki 2010). Large elevation differences strengthen foehn effects and reinforce orographic rainfall that can lead to flood wave formation in mountainous steep-slope streams (Migon and skutki 2010).

The main river of the study area is Nysa Kłodzka, the left tributary of the Odra River. It is characterized by rapid water supplies in the spring and summer as a result of concentric arrangement of numerous tributaries, which are mostly mountain streams. Table 2 juxtaposes main characteristics of the studied basins, along with mean discharges measured at four outlets under study. Three of the investigated basins have outlet located along Nysa Kłodzka. The largest one, with the area of 1744 km^2, is the contributing basin above the gauge in Bardo. Further southward, 16 km upstream, located is the Kłodzko gauge which closes the second biggest of the investigated basins. The smallest one, still located along Nysa Kłodzka, is the basin above the gauge in Bystrzyca Kłodzka, with the area of nearly 260 km^2. However, the Żelazno gauge is located along Biała Lądecka which is the longest right tributary of Nysa Kłodzka in the study area. The contributing basin above Żelazno is a subcatchment of Bardo and Kłodzko basins.

The study area belongs almost entirely to Poland; only the NW parts of the Bardo catchment belong to Czech Republic. That division makes the data related to geographical characteristics of the entire investigated area (i.e., soil cover and land use) not compatible due to different national classification criteria. Thus, the detailed statistical data, that can be assumed as representative for the study area, are available for Kłodzko County which is the administrational unit covering 84.17% of the area of interest.

The topography of the research area, i.e., Kłodzko Valley and the upper Nysa Kłodzka basin surrounded by mountain ranges, is responsible for its distinct microclimate. Although the investigated catchments, as the entire Sudety Mountains, are located in the cool temperate climate zone with marked maritime influences (Schmuck 1969), diversity between the main climate components is noticeable. Average annual air temperature calculated for the entire area of Kłodzko County is 6.3 °C (Geographic Characteristic of Counties 2004). Allowing for the temperature drop with altitude, it is lower than 1 °C at the summits (Waroszewski et al. 2013). In contrast, in Kłodzko Valley (most of Bystrzyca Kłodzka and Żelazno subcatchments),, the annual average temperature rises to approximately 7.4 °C (Schmuck 1969). The altitude influence is also clearly seen in the annual precipitation sums, which vary from 590 mm in the lower parts of Kłodzko Valley to about 1500 mm at the summits (Godek et al. 2015). The mean annual precipitation rate calculated for the entire county is 803 mm (Geographic Characteristic of Counties 2004), and at the altitude of 800 m a.s.l., annual rainfall varies between 800 and 1000 mm (Latocha and Migoń 2006). While the storms with rainfall intensity 20–50 mm/day are not considered to be abnormal (Piasecki 1996), catastrophic rainfall episodes with daily precipitation exceeding 50 mm are observed rarely—a few times per decade (Pawlik et al. 2013). Snow cover is present in the Kłodzko station for average of 63 days (Bednorz 2011) with the first occurrence in November and last in April.

In the river valleys of Nysa Kłodzka, Biała Lądecka, Ścinawka, and lower reach of Bystrzyca Dusznicka, the groundwater level does not exceed 2 m (Geographic Characteristic of Counties 2004). It changes with the distance from the river, and 2–6 km from the channel, it deepens to 10 m. The lowest water levels occur within the mountain massifs and can reach depth of several tens of meters.

The soil pattern is homogenous in the studied catchments. The soil types of the region are mainly Brown Earths and Podzoles. The river valleys are covered with Fluvisoles (Geographic Characteristic of Counties 2004).

Data

The data which become inputs to TOPMODEL can be divided into two groups, i.e., hydrometeorological time series (temporal variability) and terrain characteristics (spatial variability). The next two subsections correspond to this classification. For the purpose of the experiment, 4 consecutive hydrologic years have been selected, abbreviated hereinafter as HYs (note that in Poland hydrologic year begins on 1 November and finishes on 31 October). These HYs are: 2009, 2010, 2011, and 2012. Only four contributing basins which are mentioned and characterized above are the focus of the study.

Hydrometeorological time series

Time series of discharge, rainfall, and potential evapotranspiration are needed to calibrate TOPMODEL. They all should be calculated in m/m^2 per time step. The observed river flow and precipitation data, sampled every 15 min, are obtained from the above-mentioned Local System for Flood Monitoring, known also as LSOP, courtesy of Kłodzko County. However, potential evapotranspiration is modelled empirically, and the same 15-min time step is kept.

Since LSOP observes only water level, and thus, no discharge is measured, there is a need to use rating curves to calculate discharge time series. This has been done using the tabulated rating curves for four gauges under study, obtained courtesy of IMGW-PIB. The validity periods of the rating curves were the following: for Bystrzyca Kłodzka (06/11/2012), for Kłodzko (01/02/2013), for Bardo (30/01/2013), and for Żelazno (09/12/2012). For a few model calibration exercises, the curves were slightly newer, but the adequacy of their performance in calculating discharges from LSOP-based water levels was confirmed by the comparison with discharges measured by IMGW-PIB. The tabulated rating curves, after applying the square-root transformation to the discharge data, have been approximated with high-order polynomials which have been fitted using the least-squares method. Attention has been paid to the uniqueness of the model solution, specifically for low flows. The models have used to compute discharge time series, expressed in m^3/s, for four sites under scrutiny, and sampling every 15 min have been inherited from river stages. Since discharge data should be expressed in m/m^2 per time step, for each gauge, so calculated discharge was multiplied by 15×60 (15 min between consecutive discharge data times 60 s in a minute) and divided it by basin area (expressed in m^2).

Rainfall is measured in LSOP at 13 automatic weather stations, and hourly precipitation rate is recorded every 15 min. The data have been recalculated to fit the 15-min time step, and millimeters of rainfall have been converted to m/m^2 per time step. Since no continuous information on precipitation field is provided, the Thyssen polygons have been applied to relate rainfall to a given contributing basin.

Potential evapotranspiration has been computed empirically, i.e., an averaged time series have been constructed. It is assumed to be valid for every year. In our exercise, the evapotranspiration data set is a combination (sum) of: the daily-averaged potential evapotranspiration data for the entire year (computed for each day of year as a mean potential evapotranspiration on the corresponding days in many years) and the diurnal harmonic variation computed on the basis of the true sunrise and sunset times.

Due to the fact that the daily evapotranspiration data are not available for Kłodzko Land, the daily potential evapotranspiration data for the nearest German site of Goerlitz have been utilized, they were computed using the Turc–Wendling method in frame of the NEYMO project. Since Goerlitz is located approximately 140 km from the center of Kłodzko Land, a scaling approach is proposed to account for change in evapotranspiration between Goerlitz and the considered sites in Kłodzko Land. Thus, the daily-averaged evapotranspiration data for Goerlitz have been multiplied by a constant number which was a ratio of: (1) mean annual potential evapotranspiration between 1966 and 1995 in a single site in Kłodzko Land (499 mm for Bystrzyca Kłodzka as well as 516 mm for Kłodzko, Bardo, and Żelazno); and (2) mean annual potential evapotranspiration between 1966 and 1995 in Zgorzelec which is equal to 570 mm (Zgorzelec is a Polish town that forms an entity with German town of Goerlitz). The resulting ratios, computed using values published by Drabiński et al. (2006), led to the calculation of daily-averaged potential evapotranspiration data for basins above gauges in Bystrzyca Kłodzka, Kłodzko, Bardo and Żelazno.

Diurnal harmonic variation was simulated on a basis of true sunrise and sunset times. It was assumed that evapotranspiration is equal to zero in the night and starts growing non-linearly after the sunrise, which reaches the daily maximum and declines to zero at sunset. In this paper, the day segment was modelled with a sinusoid and replicated many times to reveal the same length as the above-mentioned daily evapotranspiration data. The similar, but not entirely identical approach, was applied by Liu et al. (2005). In our exercise, the mean value of the diurnal components was subtracted, and such a procedure prevented extra evapotranspiration values to occur when integration over time was performed.

Subsequently, the data-based daily-averaged potential evapotranspiration data were added to the mean-corrected diurnal evapotranspiration component, leading to the ultimate estimate of potential evapotranspiration.

A note should be given here on why the above-mentioned approach for estimating potential evapotranspiration has been selected. The TOPMODEL configured as described in this paper is used in the experimental multi-model hydrologic ensemble prediction system, known as HydroProg (Niedzielski et al. 2014). It is, therefore, natural that for reference to the system itself and to the previously published work on the use of TOPMODEL in HydroProg (Niedzielski and Miziński 2017), the same settings of the model should be kept to allow comparisons. However, it is likely that the empirical method for estimating averaged time series which is assumed to be valid for every year may become one of the sources of calibration errors.

Terrain characteristics

Topography of the investigated watersheds, as a major factor influencing the TOPMODEL performance, was, in this study, a subject of scrutiny. Digital Elevation Model for the watersheds was interpolated to 25 m grid based on LiDAR data. The grid size was this is a compromise between using a lower resolution that causes a loss of information about hillslope flow pathways, and a high resolution that is influenced by noise (Nourani and Zanardo 2014). Based on the flow accumulation maps generated using GIS tools, the TWI value was assigned to each raster cell in the digital elevation models of the watersheds. The classification of the values into 16 classes (Blazkova and Beven 1997; Holländer et al. 2009) was performed to reduce the computation time, and thus, the calculations were carried out for the topographic index class instead of each individual value. Figure 3 shows the frequency and the spatial distribution of the TWI values. It is clearly seen that the highest values—which indicate potential subsurface or surface contributing areas (Beven 1997)—are associated with the bases of slopes and the stream valleys. These areas are characterized by large contributing areas and relatively flat slopes. The mean TWI for the watersheds varies from 6.34 for Bystrzyca Kłodzka catchment, through 6.81 and 6.91 for Żelazno and Kłodzko, respectively, to the highest mean value of 7.37 obtained for the Bardo catchment.

DEM was also used for the estimation of the delay function which represents the time for a particle of water to travel to the outlet. Based on the flow length map, each watershed was divided into five zones, grouping the similar values of the function. The simple matrix was prepared for the further computations, with the first column indicating average distance to the outlet and the second column representing relative cumulative area of each zone.

Results

Parameter estimation

Although the TOPMODEL assumptions require relatively small number of parameters that need to be estimated, the

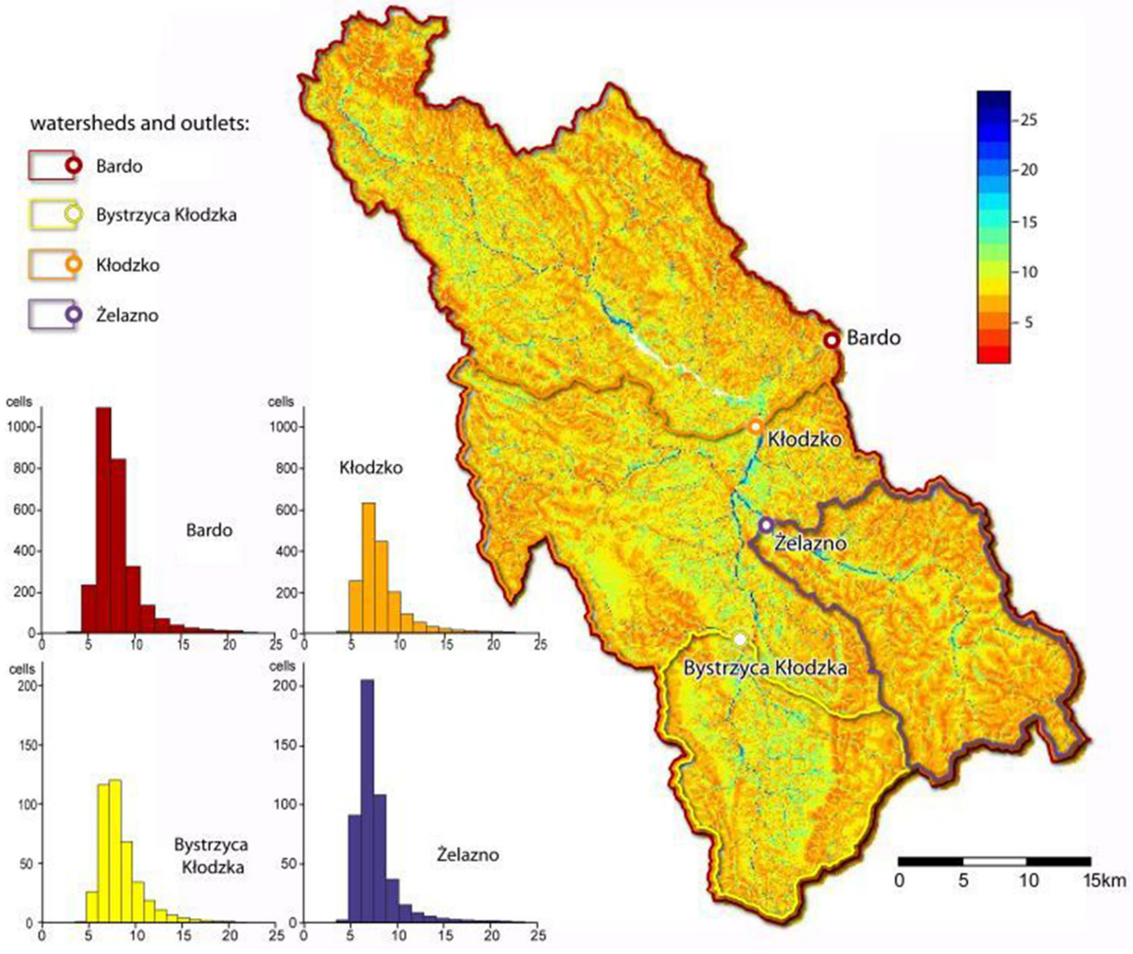


Fig. 3 TWI values and its distribution in investigated catchments

difficulty of the calibration is caused by the uncertainty of the parameters (Kuczera and Mroczkowski 1998). In addition, as Beven and Freer (2001b) stress out, a diverse set of possible parameter values can produce similar modeling results.

The Monte Carlo procedure, that has been proven to be particularly useful for hydrological studies (Romanowicz and Beven 2003), was carried out to estimate a set of parameters that offer the best model performance. In this paper, the Monte Carlo approach is used in association with the uniform distribution, i.e., random sampling across the specified parameter range is performed, assuming the same probability of sampling each element. Table 3 shows the ranges applied for each parameter based on the previous studies and manual calibration. The ranges of the parameters were kept wider than the expected possible values for the catchment (Freer et al. 1996). To enhance the certainty, the number of simulations was set to 10,000. Further increase of this number did not improve the final result and required a more time-consuming computation. The procedure was carried out for all above-mentioned catchments, in each case study for the period of 1 hydrological year, and the data from HY 2010, HY 2011, and HY 2012 were taken as an input. Table 3 confirms the statement that constraining the perfect parameter set is not possible and the modeling needs to rely on the best performing, not necessary actual, values. In the experiment, three independent calibration exercises (HY 2010, HY 2011, and HY 2012) are carried out to show the level of variability of model parameters and its impact on model performance.

Each parameter is equally important during the Monte Carlo sampling, although the manual calibration showed that four of the parameters—m, lnTe, $\mathrm{Sr}_{\max}$, and vch—are more meaningful, i.e., variation in their values influences the model performance and the shape of the simulated hydrograph most significantly. The highest sensitivity is associated with m parameter, which represents the change in the saturate hydraulic conductivity with depth. Small values of m imply the quick flow and insignificant subsurface runoff, while large values indicate that more rainfall can infiltrate the soil, and thus, less water reaches the outlet via surface route (Sigdel et al. 2011). For the investigated catchments, m parameter range was kept wide, assuming that the well-vegetated, deep-seated catchments of the study area can be well characterized by the large values of m. The next highly sensitive parameter, lnTe, influences directly the shape of hydrograph. The quick recession is associated with small values of the lnTe parameter, while low values result in gradual fall of the hydrograph limb after the peak, as a result of increasing saturated transmissibility that may cause runoff delay. This parameter draws a special attention in this study, since the shape of the recession limb in the modelled hydrographs often did not resemble the observed ones. Constrained allowable range for this parameter was set to be between − 2 and 1, and its value for the most efficient runs varies from − 1.37 (Bystrzyca Kłodzka catchment, simulation for 2011) to 0.99 (Żelazno catchment, simulation for 2012). The third parameter that was found to be sensitive, although not as much as the previous ones, is the maximum root zone deficit $\mathrm{Sr}_{\max}$. The value of this parameter indicates the influence of evapotranspiration on the hydrological behavior of the catchment. Small root zone deficit (low $\mathrm{Sr}_{\max}$ value) allows less water to be stored in the root zone and hence available for evapotranspiration (Sigdel et al. 2011) what can lead to the increased runoff. An extended knowledge of the catchment vegetation is necessary for the $\mathrm{Sr}_{\max}$ calculation. The difference of the water contents at field capacity and the permanent wilting point needs to be multiplied by the rooting depth of the soil (Beven and Freer 2001a). Due to the lack of such detailed data, $\mathrm{Sr}_{\max}$ parameter ranges were very wide (0–3 m). The best performing parameter sets for the simulations for different HY in the same catchment contained $\mathrm{Sr}_{\max}$ parameters with

Table 3 Parameter ranges applied for random sampling in the Monte Carlo procedure and their significance

Parameter		Range	Significance
qs0	Initial subsurface flow per unit area (m)	0 to 0.00004	Insensitive
lnTe	Log of the areal average of $T0$ (m^2/h)	− 2 to 1	More sensitive
m	Model parameter controlling the rate of decline of transmissivity in the soil profile	0 to 2	Highly sensitive
Sr0	Initial root zone storage deficit (m)	0 to 0.02	Insensitive
$\mathrm{Sr}_{\max}$	Maximum root zone storage deficit (m)	0 to 3	Sensitive
td	Unsaturated zone time delay per unit storage deficit (h/m)	0 to 3	Less sensitive
vr	Channel flow inside catchment (m/h)	800 to 1000	Sensitive
$k0$	Surface hydraulic conductivity (m/h)	0 to 0.01	Less sensitive
CD	Capillary drive	0 to 5	Insensitive

various values, and the estimation of the right span was difficult. Constraining the ranges for the last highly sensitive parameter did not cause such problems, since channel flow velocity can be estimated by dividing the observed discharge by the cross-sectional area of the stream. The channel velocity varies along the stream, but for all the investigated catchments, the best performed parameter sets were generated when the each parameter range was set to 800–1000 m/s.

There final parameter values' ranges presented in Table 4 which yielded the best overall fit of the model when executed for entire hydrograph for a HY.

Model performance

The calibration of the model was performed on all four watersheds for HY 2010, HY 2011, and HY 2012. The rainfall, discharge, and evapotranspiration data were converted to 15-min time steps, and the Monte Carlo procedure was performed to generate the best performing parameters set out of 10,000 individual sets. Table 4 shows the modeling results and the calibrated parameter values. Obtained efficiency statistics as well as the parameter ranges are not consistent for all catchments and all simulation periods. For all of the catchments, TOPMODEL achieved the best fit modeling the discharge in HY 2011. The topography, land use, soils, and other characteristics of terrain influencing the runoff generation were relatively steady, but model efficiency statistics varied between the modeling time periods, which proves that the model performance measures are strongly dependent on weather conditions.

The best performance of TOPMODEL was found for Bystrzyca Kłodzka catchment in HY 2011, with $NSE = 0.78$ and the correlation between observed and simulated discharge of 0.89 (Fig. 4). Slightly less skillful was TOPMODEL in Kłodzko catchment, with values of the above-mentioned statistics of 0.66 and 0.83, respectively. The model performance expressed by $NSE > 0.6$ is considered as satisfactory, also named as behavioral (Beven and Freer 2001b). For two remaining catchments, Bardo and Żelazno, TOPMODEL was unable to simulate the hydrograph with fair accuracy. In Bardo basin, the modelled discharge was, for each time period, less accurate as the mean of the observed data ($NSE < 0$). For Żelazno catchment, this situation occurred for the HY 2012 data, and the results for the remaining modeling periods, HY 2010 and HY 2011, were also not satisfactory, with $NSE = 0.31$ and 0.42, respectively.

Based on the results of calibration using the yearly data, the watersheds were categorized into three categories: "good"—all obtained $NSE > 0$, "acceptable"—all obtained $NSE \geq 0$, and "unacceptable"—some obtained $NSE < 0$ (Blazkova et al. 2002). The third category, containing two watersheds—Żelazno and Bardo—was excluded from further analysis and the emphasis was put on finding the underlying causes of model bad performance in these basins.

Bardo is the biggest of the investigated catchments (Fig. 2) and includes the basin of the left Nysa Kłodzka river tributary—i.e., Ścinawka. Within this watershed, there is only one meteorological station with rain gauge. Hence, due to numerous local anomalies in the precipitation field in the study area (orographic precipitation, rain shadows), the precipitation measurements weighted by Thiessen polygons are imprecise representation of real spatial variability of rainfall. Both dense distribution of measurements in the mountainous areas and relatively small number of stations located on the plains and in the NW part of Bardo watershed lead to the difficulty in adjusting model parameters and, as result, in accurate simulating discharge. Żelazno is a watershed with the most diverse topography; hence, the local anomalies in precipitation occur more frequently and have greater impact on model misrepresentation of the spatial rain pattern. It is also considered that the soil properties in this forested watershed (forests compose over 63% of land use) can exhibit seasonal variability which is more significant than in remaining watersheds. Further investigation into the latter problem is needed, because Polish digital soil maps provide information about soil properties only for agricultural land (Drzewiecki et al. 2014), and thus, there are no data for forested areas.

For Bystrzyca Kłodzka and Kłodzko watersheds, TOPMODEL performed better in predicting discharge than the observed mean. The only exception is associated with Kłodzko watershed in HY 2010, for which the model predictions were exactly as accurate as the mean observed discharge ($NSE = 0.03$). In the watershed where TOPMODEL performance was superior over the remaining basins, i.e., Bystrzyca Kłodzka, the model obtained the highest efficiency for HY 2011, what is consistent with other watersheds investigated in HY 2011. It may hypothesized that meteorological conditions in this HY differed significantly from the remaining calibration HYs, and the processes involved in these conditions can be better represented by the model. This may be confirmed by the analysis of rainfall and discharge patterns in Kłodzko station (Table 5).

The primary difference in the shapes of hydrographs for HYs 2010–2012 is the existence of a large mid-summer peak as a result of major storm event occurring on 24/07/2011. The peak was reconstructed well by the model—its underestimation by over 20% is acceptable taking into

Table 4 Modeling results and the calibrated parameter values

	Bardo			Bystrzyca Kłodzko			Kłodzko			Żelazno		
	2010	2011	2012	2010	2011	2012	2010	2011	2012	2010	2011	2012
NSE	− 0.49	− 0.01	− 0.10	0.29	0.78	0.32	0.03	0.66	0.32	0.31	0.42	− 0.18
RMSE	23.884	20.561	14.458	4.502	3.485	3.264	16.103	7.831	6.974	5.644	2.947	2.889
Cor	0.07	0.29	0.14	0.59	0.89	0.58	0.59	0.83	0.64	0.70	0.66	0.06
Best performing parameter set												
qs0	3.99E−05	3.99E−05	3.95E−05	1.75E−05	3.99E−05	3.21E−05	5.79E−06	3.91E−05	3.96E−05	2.04E−05	3.90E−05	2.47E−05
lnTe	0.898	0.400	0.872	− 0.345	− 1.964	− 0.542	0.366	− 0.576	0.278	− 1.332	0.023	0.989
m	1.932	0.956	0.453	0.022	1.605	0.096	0.037	1.875	0.175	0.030	1.268	1.971
Sr0	0.004	0.018	0.011	0.007	0.009	0.002	0.007	0.008	0.001	0.018	0.015	0.006
$\mathrm{Sr}_{\mathrm{max}}$	2.647	2.810	2.619	0.002	0.464	2.409	0.000	0.275	2.979	0.019	0.021	0.355
td	0.839	2.391	1.948	0.855	1.212	0.485	1.127	1.371	2.446	1.211	1.731	0.502
vch	811.5	860.2	895.4	877.0	876.0	851.7	825.3	898.3	815.7	824.5	965.6	848.4
vr	891.9	934.5	980.7	958.0	841.5	903.5	987.8	999.1	970.1	809.3	878.5	983.0
*k*0	0.006	0.006	0.003	0.005	0.003	0.005	0.003	0.008	0.007	0.004	0.009	0.003
CD	2.471	4.101	1.256	1.038	4.434	2.313	3.332	1.852	1.577	4.921	2.209	1.735

NSE Nash–Sutcliffe efficiency, *RMSE* root-mean-square error (m^3/s), *cor* Pearson's correlation coefficient, *qs0* initial subsurface flow per unit area (m), *lnTe* log of the aerial average of *T*0 (m^2/h), *m* model parameter controlling the rate of decline of transmissivity with the soil profile, *Sr0* initial root zone storage deficit (m), Sr_{max} maximum root zone storage deficit (m), *td* unsaturated zone time delay per unit storage deficit (h/m), *vr* channel flow inside the catchment (m/h), *k0* surface hydraulic conductivity (m/h), *CD* capillary drive

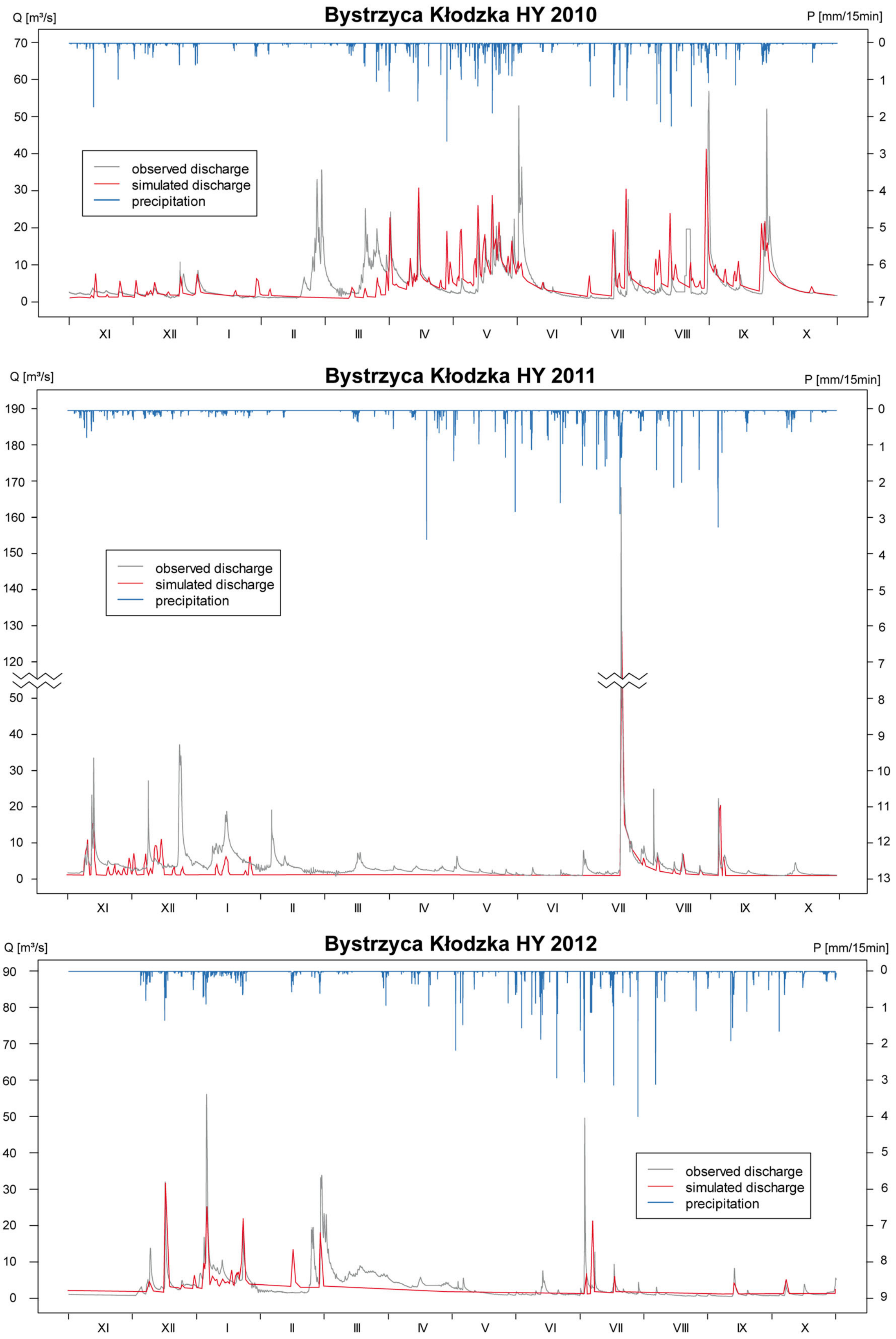


Fig. 4 TOPMODEL simulation results for the Bystrzyca Kłodzka watershed

account the magnitude of the event. The well-fitted parts of the hydrograph include also recession curve after the main peak, and modelling such situations is perceived as one of the most problematic responses to be reconstructed by the TOPMODEL (Sigdel et al. 2011). The model was able to predict smaller peaks after the event and the estimation of the base flow also improved after the main peak. The biggest discrepancies between observed and simulated runoff occur in the winter season. This is due to the limitation of this simple version of TOPMODEL which does not account for the water accumulated in snow cover. Because the model uses the same parameters to estimate the discharge during the whole simulation period, its values need to be calibrated to produce the smallest overall prediction error, for different hydrological settings. In such a case, the model seems to provide superior fit to the large peaks rather than to other hydrological situations. The evaluation of model performance on a basis of NSE itself can be misleading due to the inclination of this measure to place emphasis on the larger errors, while the smaller ones tend to be neglected. The acceptable performance of the model during one extreme event contributes to good statistical performance for entire hydrograph and poor results in the representation of the base flow. Although this discrepancy occurs during long periods of simulation for low flows, its impact on the overall efficiency measure is rather small. The model parameters estimated with support of this criterion produce a hydrograph that recreates peaks with reasonable accuracy, but fails to match the observed hydrograph during low flows. In 2011, despite the high 0.78 NSE, the model underestimated the mean discharge by over 40%, similar to the HY 2012 with much lower NSE of 0.29.

The same pattern can be observed in Kłodzko catchment. In this case, the recession curve was not reproduced as accurately and the overall model performance is lower (NSE = 0.66) than in Bystrzyca Kłodzka catchment (Fig. 5). The model does not provide a good representation of hydrograph during the winter season, when discharge is impacted by snow-melt and water can be stored in snow cover. Similar situation was detected also on other hydrographs for all the catchments, i.e., simulations for period December–April were found to be inaccurate. This leads to the conclusion that snow component should be included in the model structure to properly reconstruct the hydrological behavior of the investigated catchments in winter seasons. To confirm the impact of this misrepresentation of the hydrograph, the model was tested on the shorter periods and the following section contains the results of this simulations.

Table 5 Rainfall and discharge patterns in Kłodzko station and the NSE values obtained in the simulations

	2010	2011	2012
Mean observed discharge (m^3/s)	3.11	3.82	4.81
Mean simulated discharge (m^3/s)	3.61	2.27	2.92
NSE	0.29	0.78	0.32
Precipitation (mm)	748	758	836

Optimal time span for simulations

The best performing watershed—Bystrzyca Kłodzka—was chosen to conduct a detailed analysis of the model ability to reproduce hydrological response during periods of different lengths. It was assumed, based on the shape of the simulated hydrograph for entire year in relation to the observed one, that the model performance during the winter season will exhibit the lowest accuracy expressed by the NSE. To determine the most optimal time span for the model simulations, the periods of 1 week, 2 weeks, 3 weeks, 1 month, 2 months, 3 months, and 6 months were taken into account. Table 6 shows the results for HY 2011, where a given time span (e.g., 1 week of data) was iteratively moved forward by 1 day (which corresponds to 96 observations) to rerun simulations and get a set of statistics. Figure 6 depicts the variability in achieved NSE values for 2-week, 1-month, and 6-month periods.

The overall performance of a model for particular time period was judged subjectively by comparing the efficiency measures obtained by each simulation independently. To shorten the time-consuming calculations, the number of simulations in the Monte Carlo procedure has been reduced from 10,000 to 1000. It was impossible to find the timespan that would give satisfactory results for all hydrological settings, and this is clearly seen when analyzing winter seasons. The mean NSE, calculated for a window of a given length moved in a stepwise way through the entire year, is severely impacted by the low values representing the periods in winter season. Most sensitive to this effect, and thus producing dispersed values, were shorter periods—standard deviation of NSE for the 1-week period exceeds 2 m^3/s. The NSE of the best-fitted parameter set in all periods is very high, namely of 0.95–0.97, but these values relate mostly to the time spans when the discharge rate is stable and this stability is expressed properly by the model. To avoid the positive bias, percentage of the parameter sets that can be considered behavioral ($NSE > 0.6$) and percentage of the parameter sets that perform better than the mean of observed values ($NSE > 0$) have been calculated. The variability of the mean NSE among the investigated periods is high, but the ratio of behavioral parameter sets and especially parameter sets with $NSE > 0$ is much more stable. The NSE decreases rapidly after each peak caused by the snow-melt

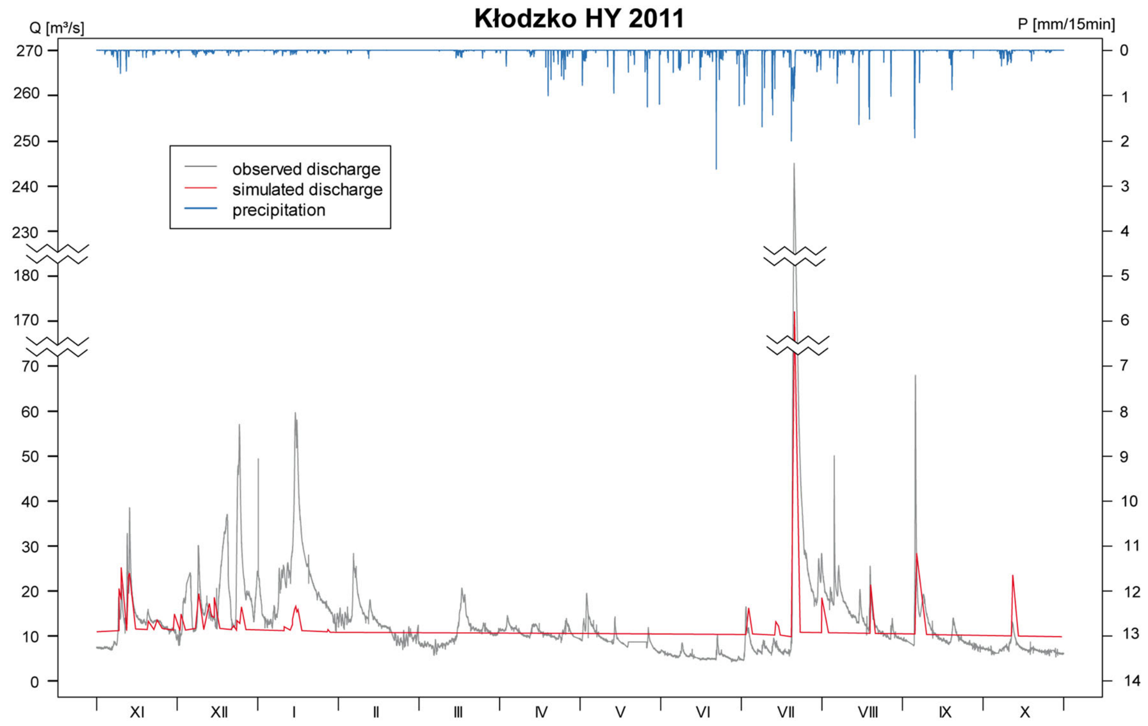


Fig. 5 TOPMODEL simulation results for the Kłodzko watershed in HY 2011

Table 6 Comparison of simulation results for chosen period lengths for Bystrzyca Kłodzka catchment in HY 2011

Time span	1 week	2 weeks	3 weeks	1 month	2 months	3 months	6 months
Number of simulations	359	352	345	336	306	276	246
Number of time steps in each simulation	672	1344	2016	2880	5760	8640	11,520
Mean NSE	− 0.52	0.09	0.22	0.28	0.34	0.39	0.41
Std dev. NSE	2.22	5.25	0.53	0.48	0.43	0.44	0.44
Best NSE	0.98	0.97	0.95	0.96	0.96	0.96	0.95
NSE > 0.6 (%)	25.07	26.42	25.79	26.78	30.06	32.97	42.27
NSE > 0 (%)	62.67	67.04	70.14	72.32	70.26	73.18	75.61
Mean RMSE	1.69	2.00	2.18	2.34	2.74	2.90	2.92

NSE Nash–Sutcliffe efficiency, *RMSE* root-mean-square error (m^3/s)

as indicated in Fig. 6. Another major dip corresponds to the storm from 24 July 2011, when the underestimation of the main peak flow affects the NSE value. The best performance of the model was noticed for the 6-month period. Just after the winter season, the NSE values rise gradually and reach a plateau of $NSE > 0.9$ beginning with the simulations starting at the end of March. It has been noticed that for the longer time spans, the model was able to simulate the major peak from 24 July 2011 with higher accuracy.

Further analysis of the impact of the snow component was performed by limiting the modeling to period without snow cover: April to October for HY 2010, 2011, and 2012. Univariate analysis of the simulated discharge in comparison to the calculations for entire HY as well as NSE values confirms that the model performance is highly influenced by the snow-melt component (Table 7). For the best performing HY 2011, as predicted in optimal time span analysis, the NSE value was as high as 0.93 for the months without snow cover in comparison to 0.78 for entire year. The goodness-of-fit for simulations in HY 2010 and 2012 was also significantly enhanced by limiting modeling to the April–October period. NSE value for HY 2010 increased from 0.29 to 0.62 reaching the behavioral threshold; and in HY 2012, the value of NSE raised from

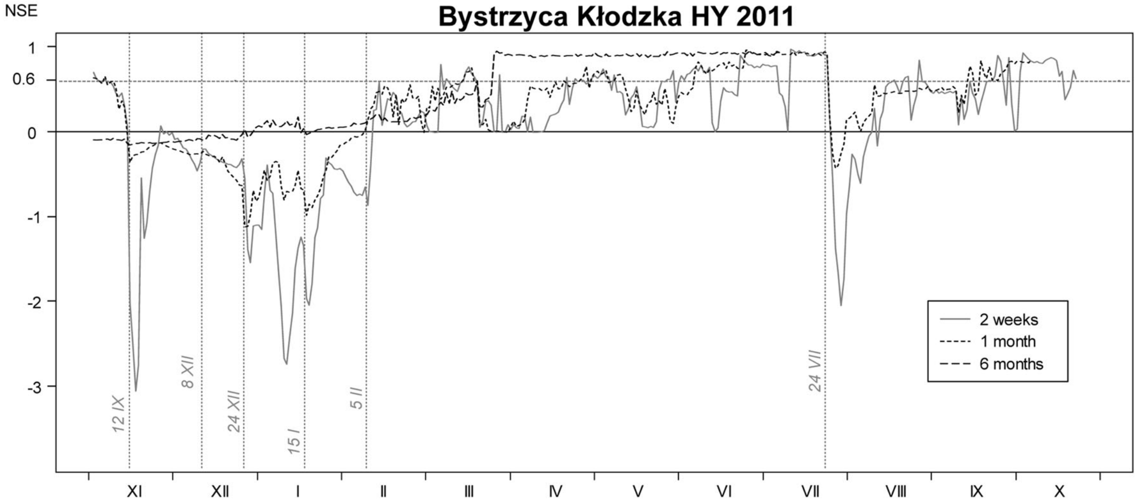


Fig. 6 Performance of TOPMODEL as a function of span of data used to calibrate the model

Table 7 Comparison of statistics for model performance in Bystrzyca catchment for entire HY and for months without snow cover (April–October)

	HY			IV–X		
	2010	2011	2012	2010	2011	2012
NSE	0.29	0.78	0.32	0.62	0.93	0.57
Standard dev.	3.19	5.91	4.01	6.12	7.32	7.01

0.32 to 0.57. It should also be noted that for the experiment limited to the April–October period (Table 7).

Finally, it is worth noting that there were applications of TOPMODEL in temperate cold climate for which snow-melt component was not included (Lamb et al. 1997). This supports our approach and the comparison presented in Table 7.

Model validation

The optimized parameter set from HY 2011 in the best performing watershed Bystrzyca Kłodzka was applied to the same watershed for the HY 2009 to validate the model. In the validation exercise, the 6-month time span was used as it was shown to be the data length for which model skills were the superior over all the studied cases (Table 6). Thyssen polygons for spatial distribution of rainfall data were recalculated because of the lack of precipitation data in one of the station used for HY 2010–2012. Precipitation data from station located 8 km away were used as a substitute.

Validation performed for entire 2009 HY showed that the hydrograph fails to match with the observed one at the beginning of the year, during the snow-melting period. Following the procedure described in "Optimal time span for simulations", the simulations were limited to the shorter time spans, focusing on the period without snow cover and using the best performing parameter set calibrated for Bystrzyca catchment for the months without snow cover.

Model performed best for the summer and fall months, with the exception of the period from mid-June to mid-July, where the simulated hydrograph failed to match the observed one. Figure 7 shows the results of the validation preformed on the late summer months of the 2009 HY. The NSE value obtained for depicted time span was 0.73; and limiting the simulation period to only the months of August and September, the NSE value reaches 0.87 and predicts accurately the recession curve. Despite the high value of NSE and relatively accurate representation of the first peak, the hydrograph significantly underestimates the second peak in the period of simulation.

Validation showed satisfactory results also for the month of May 2009 (Fig. 8). The NSE of 0.63 is above the 0.6 threshold for being classified as behavioral. Similar to the above described period, the peaks of the hydrograph are underestimated and the model fails to predict accurately the last peak in the period showing only a slight rise in the discharge rates, when the observed values form the largest peak in this time span.

The poor performance of the model for the validation period of entire 2009 HY as well as the months without snow cover can be explained by the uniqueness of conditions in the best performing 2011 HY, when the intense precipitation and very high discharge rates in the major peak from 24 July 2011 resulted in calibration of the parameters performing best in this unique conditions and failing to simulate the hydrograph in less extreme circumstances. Despite that, there were time spans in the validation period that the model demonstrated satisfactory results, only slightly lower than the values obtained during calibration period.

Fig. 7 Validation of TOPMODEL for the Bystrzyca Kłodzka watershed

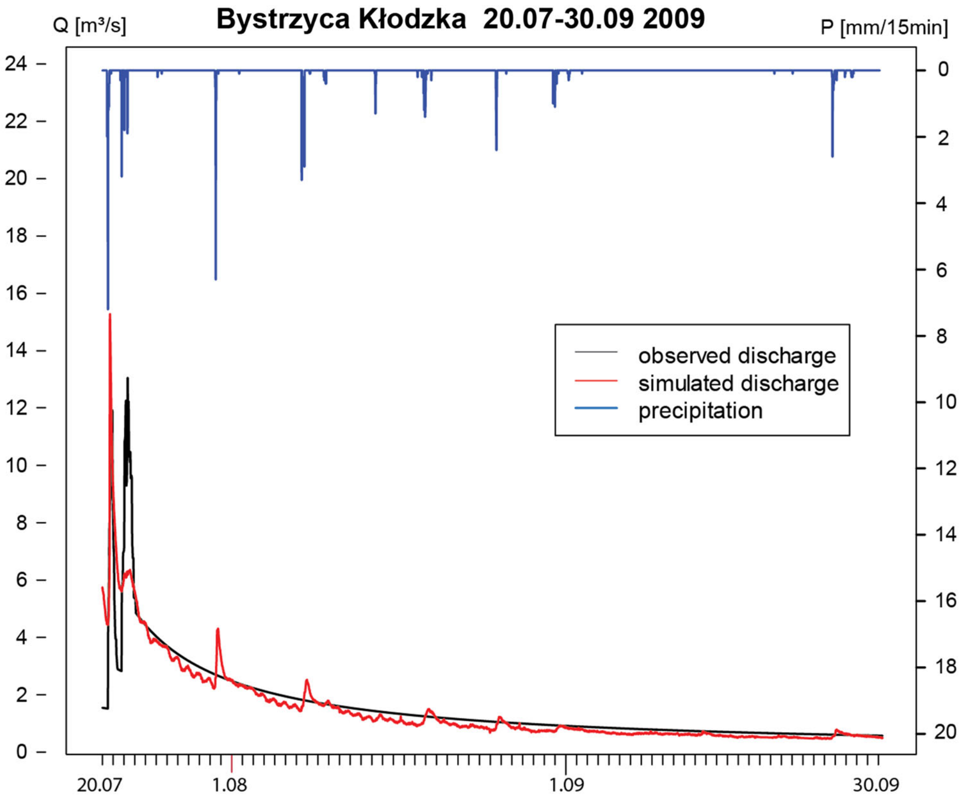
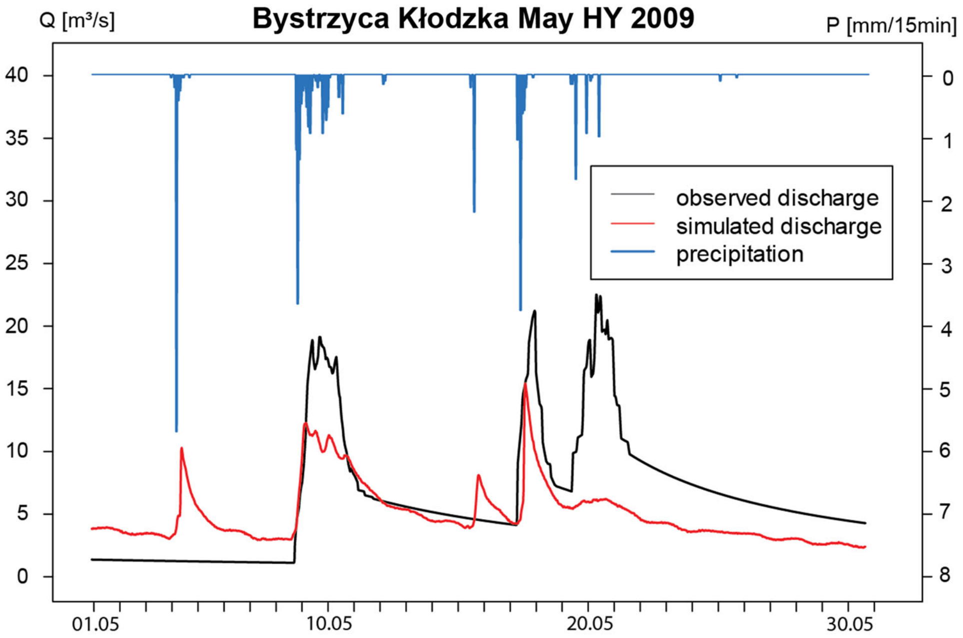


Fig. 8 Validation of TOPMODEL for the Bystrzyca Kłodzka watershed in the month of May 2009



Conclusions

TOPMODEL was successfully applied to four subcatchments of the upper Nysa Kłodzka river basin, but was able to reproduce the main pattern of the hydrograph with acceptable accuracy only for two of them. The conclusions are the following.

1. Poor performance of the model in two catchments can have variety of reasons, including input data error, calibration inaccuracy, parameter uncertainty, and model structure. The most probable cause of

misrepresentation of hydrograph lies in the snow-melt component that is not included in this basic version of TOPMODEL. A more sophisticated structure of the response function needs to be used to improve the TOPMODEL performance in all of the investigated watersheds. Low accuracy of the model can also be effect of the model inability to represent distributed rainfall pattern.

2. Complicated environment and lack of soil data make the calibration of parameters challenging. The Monte Carlo simulation produces the most suitable parameter sets, but they may not correspond to the actual conditions in the watershed.
3. It has been found that the goodness-of-fit increases along with time span of data used for TOPMODEL calibration, and among the studied periods, the half-year solution produces the best agreement between data and model simulations. However, such estimates cannot be treated as global ones, since they are highly dependent on hydrological settings and weather conditions.
4. Simulations that do not include winter season provided promising results, the NSE for nearly half of the simulations using 6-month time span of data for Bystrzyca Kłodzka catchment are higher than 0.6.
5. Snow cover was found to impact the model performance, i.e., when the analysis is limited to snow-free months, the NSE values are considerably higher than for the entire year which includes periods of snow cover occurrence.
6. Validation performed using the best set of parameters obtained during calibration for the best performing watershed was found to demonstrate satisfactory results (obtained just slightly lower NSE values than during calibration period), but only for shorter time spans and failed to simulate the hydrograph for the entire HY used as validation period.

Acknowledgements The research has been financed by the National Science Centre (Poland), research Project no. 2011/01/D/ST10/04171 under leadership of Dr. hab. Tomasz Niedzielski, Professor at the University of Wrocław (Poland). The authors thank the authorities of the County Office in Kłodzko for productive partnership and providing us with the data of the Local Flood Monitoring System (Lokalny System Osłony Przeciwpowodziowej–LSOP). The tabulated rating curves have been acquired from the Institute of Meteorology and Water Management, National Research Institute (Poland). Sincere gratitude needs to be expressed to Prof. Renata Romanowicz for the guidance and support in clarifying the model concepts and nuances. Further thanks go to Prof. Helena Mitasova who provided comments and valuable suggestions during the last stage of project. We also thank Dr. Wouter Buytaert for unveiling the details of the source code. The daily evapotranspiration data for Goerlitz were obtained courtesy of the KLAPS/NEYMO projects, and we wish to thank Mr. Andreas Völlings, Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie (Germany), for his approval for use of the aforementioned data set. We are also grateful to Mrs. Magdalena Stec for preparing evapotranspiration time series. We are also indebted to Dr. Danuta Trojan for discussions on LSOP and on water management problems in Kłodzko County. We thank Dr. hab. Mariusz Szymanowski, Dr. Małgorzata Wieczorek, and Dr. Waldemar Spallek for preparing the Digital Elevation Model that has been used in this study. Last but not least, we express our thanks to Mr. Bartłomiej Miziński who kindly helped to verify the rating curve models.

Compliance with ethical standards

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

References

Bárdossy A (2007) Calibration of hydrological model parameters for ungauged catchments. Hydrol Earth Syst Sci 11(2):703–710

Barling RD, Moore ID, Grayson RB (1994) A quasi-dynamic wetness index for characterizing the spatial distribution of zones of surface saturation and soil water content. Water Resour Res 30(4):1029–1044

Bastola S, Ishidaira H, Takeuchi K (2008) Regionalisation of hydrological model parameters under parameter uncertainty: a case study involving TOPMODEL and basins across the globe. J Hydrol 357:188–206

Bednorz E (2011) Synoptic conditions of the occurrence of snow cover in central European lowlands. Int J Climatol 13(8):1108–1118

Beven K (1986) Runoff production and flood frequency in catchments of order n: an alternative approach. In: Gupta VK, Rodríguez-Iturbe I, Wood EF (eds) Scale problems in hydrology. Reider, Dordrecht, pp 107–131

Beven K (1997) TOPMODEL: a critique. Hydrol Process 11(9):1069–1085

Beven K, Binley A (1992) The future of distributed models: model calibration and uncertainty prediction. Hydrol Process 6(3):279–298

Beven K, Freer J (2001a) A dynamic topmodel. Hydrol Process 15(10):1993–2011

Beven K, Freer J (2001b) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J Hydrol 249:11–29

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. Hydrol Sci 24:43–69

Beven K, Wood EF (1983) Catchment geomorphology and the dynamics of runoff contributing areas. J Hydrol 65:139–158

Beven KJ, Kirkby MJ, Schofield N, Tagg AF (1984) Testing a physically-based flood forecasting model (TOPMODEL) for three U.K. catchments. J Hydrol 69:119–143

Beven KJ, Quinn PF, Lamb R, Romanowicz R, Freer J (1995) TOPMODEL. In: Singh VP (ed) Computer models of watershed hydrology. Water Resources Publications, Highlands Ranch, pp 627–668

Blazkova S, Beven K (1997) Flood frequency prediction for data limited catchments in the Czech Republic using a stochastic rainfall model and TOPMODEL. J Hydrol 195:256–278

Blazkova S, Beven KJ, Kulasova A (2002) On constraining TOPMODEL hydrograph simulations using partial saturated area information. Hydrol Process 16(2):441–458

Brasington J, Richards K (1998) Interactions between model predictions, parameters and DTM scales for TOPMODEL. Comput Geosci 24:299–314

Cameron DS, Beven KJ, Tawn J, Blazkova S, Naden P (1999) Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty). J Hydrol 219:169–187

Chairat S, Delleur JW (1993) Effects of the topographic index distribution on predicted runoff using grass. J Am Water Resour Assoc 29:1029–1034

Chen J, Wu Y (2012) Advancing representation of hydrologic processes in the soil and water assessment tool (SWAT) through integration of the TOPographic MODEL (TOPMODEL) features. J Hydrol 420–421:319–328

Choi HT, Beven K (2007) Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. J Hydrol 332:316–336

Criss RE, Winston WE (2008) Do nash values have value? Discussion and alternate proposals. Hydrol Processes 22(14):2723–2725

Desmet PJJ, Govers G (1996) A GIS procedure for automatically calculating the USLE LS factor on topographically complex landscape units. J Soil Water Conserv 51:427–433

Drabiński A, Radczuk L, Nyc K, Mokwa M, Olearczyk D, Markowska J, Bac-Bronowicz J, Chmielewska I, Jawecki B, Gromada O, Pikul K, Malczewska B, Goździk M (2006) Program Małej Retencji Wodnej w województwie dolnośląskim. Sejmik Województwa Dolnośląskiego, Wrocław

Drzewiecki W, Wężyk P, Pierzchalski M, Szafrańska B (2014) Quantitative and qualitative assessment of soil erosion risk in Małopolska (Poland), supported by an object-based analysis of high-resolution satellite images. Pure Appl Geophys 171(2014):867–895

Durand P, Robson A, Neal C (1992) Modelling the hydrology of submediterranean montane catchments (Mont-Lozère, France) using TOPMODEL: initial results. J Hydrol 139:1–14

Erskine RH, Green TR, Ramirez JA, MacDonald LH (2006) Comparison of grid-based algorithms for computing upslope contributing area. Water Resour Res 42:1–9

Fisher JI, Beven KJ (1996) Modelling of streamflow at Slapton Wood using TOPMODEL within an uncertainty estimation framework. Field Stud 8:577–584

Franchini M, Wendling J, Obled C, Todini E (1996) Physical interpretation and sensitivity analysis of the TOPMODEL. J Hydrol 175:293–338

Freer J, Beven K, Ambroise B (1996) Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. Water Resour Res 32(7):2161–2173

Freer J, Beven K, Peters N (2003) Multivariate seasonal period model rejection within the generalised likelihood uncertainty estimation procedure. In: Duan Q, Gupta H, Sorooshian S, Rousseau AN, Turcotte R (eds) Calibration of watershed models. AGU Books, Washington, pp 69–87

Furusho C, Andrieu H, Chancibault K (2014) Analysis of the hydrological behaviour of an urbanizing basin. Hydrol Process 28(4):1809–1819

Gallart F, Latron J, Llorens P, Beven KJ (2008) Upscaling discrete internal observations for obtaining catchment-averaged TOPMODEL parameters in a small Mediterranean mountain basin. Phys Chem Earth 33:1090–1094

Geographic Characteristic of Counties (2004) IUNG—Institute of Soil Science and Plant Cultivation, Puławy, Poland

Godek M, Sobik M, Błaś M, Polkowska Ż, Owczarek P, Bokwa A (2015) Tree rings as an indicator of atmospheric pollutant deposition to subalpine spruce forests in the Sudetes (Southern Poland). Atmos Res 151:259–268

Gumindoga W (2010) Hydrologic impacts of landuse change in the Upper Gilgel Abay River Basin, Ethiopia: TOPMODEL Application., Ph.D. Thesis, University of Twente, Faculty of Geo-Information and Earth Observation (ITC), Enschede, Netherlands

Hewlett JD, Hibbert AR (1967) Factors affecting the response of small watersheds to precipitation in humid areas. Forest hydrology. Pergamon Press, New York, pp 275–290

Holko L, Lepistö A (1997) Modelling the hydrological behaviour of a mountain catchment using TOPMODEL. J Hydrol 196:361–377

Holländer HM, Blume T, Bormann H, Buytaert W, Chirico GB, Exbrayat J-F, Gustafsson D, Hölzel H, Kraft P, Stamm C, Stoll S, Blöschl G, Flühler H (2009) Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data. Hydrol Earth Syst Sci 13:2069–2094

Hornberger GM, Beven KJ, Cosby BJ, Sappington DE (1985) Shenandoah watershed study: calibration of a topography-based, variable contributing area hydrological model to a small forested catchment. Water Resour Res 21(12):1841–1850

Krause P, Boyle DP, Bäse F (2005) Comparison of different efficiency criteria for hydrological model assessment. Adv Geosci 5:89–97

Kuczera G, Mroczkowski M (1998) Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. Water Resour Res 34(4):751–763

Lamb R, Beven KJ, Myrabø S (1997) Discharge and water table predictions using a generalized TOPMODEL formulation source. Hydrol Process 11(9):1145–1167

Latocha A, Migoń P (2006) Geomorphology of medium-high mountains under changing human impact, from managed slopes to nature restoration: a study from the Sudetes, SW Poland. Earth Surf Proc Land 31(13):1657–1673

Liu S, Graham WD, Jacobs JM (2005) Daily potential evapotranspiration and diurnal climate forcings: influence on the numerical modelling of soil water dynamics and evapotranspiration. J Hydrol 309:39–52

Merot P, Ezzahar B, Walter C, Aurousseau P (1995) Mapping waterlogging of soils using digital terrain models. Hydrol Process 9(1):27–34

Migoń P, Placek A (2014) Litologiczno-strukturalne uwarunkowania rzeźby Sudetów (Lithological and structural control on the relief of the Sudetes). Przegląd Geologiczny 62(1):36–43

Migoń P, Wyjątkowe zdarzenia przyrodnicze na Dolnym Śląsku i ich skutki (2010) In: Rozprawy Naukowe Instytutu Geografii i Rozwoju Regionalnego Uniwersytetu Wrocławskiego, Uniwersytet Wrocławski, Wrocław, Poland, pp 35–80

Mitasova H, Hofierka J, Zlocha M, Iverson LR (1996) Modeling topographic potential for erosion and deposition using GIS. Int J Geogr Inf Syst 10:629–641

Molicova H, Grimaldi M, Bonell M, Hubert P (1997) Using TOPMODEL towards identifying and modelling the hydrological patterns within a headwater, humid, tropical catchment. Hydrol Process 11(9):1169–1196

Moore ID, Burch GJ (1986) Physical basis of the length-slope factor in the universal soil loss equation. Soil Sci Soc Am J 50:1294–1298

Moore RD, Thompson JC (1996) Are water table variations in a shallow forest soil consistent with the TOPMODEL concept? Water Resour Res 32:663–669

Moore ID, Wilson JP (1992) Length-slope factors for the revised universal soil loss equation: simplified method of estimation. J Soil Water Conserv 47:423–428

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. Trans ASABE 50(3):885–900

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—a discussion of principles. J Hydrol 10:282–290

Niedzielski T, Miziński B (2017) Real-time hydrograph modelling in the upper Nysa Kłodzka river basin (SW Poland): a two-model hydrologic ensemble prediction approach. Stoch Env Res Risk Assess 31:1555–1576

Niedzielski T, Miziński B, Kryza M, Netzel P, Wieczorek M, Kasprzak M, Kosek W, Migoń P, Szymanowski M, Jeziorska J, Witek M (2014) HydroProg: a system for hydrologic forecasting in real time based on the multimodelling approach. Meteorol Hydrol Water Manag Res Oper Appl 2:65–72

Nourani V, Mano A (2007) Semi-distributed flood runoff model at the subcontinental scale for southwestern Iran. Hydrol Process 21(23):3173–3180

Nourani V, Zanardo S (2014) Wavelets-based regularization of the extracted topographic index from high-resolution topography for hydro-geomorphic applications. Hydrol Process 28:1345–1357

Nourani V, Roughani A, Gebremichael M (2011) Topmodel capability for rainfall-runoff modeling of the Ammameh watershed at different time scales using different terrain algorithms. J Urban Environ Eng 5:1–14

Orczykowski T, Tiukało A (2016) Retention of afforestation areas as part of flood protection—research site and methodology for headwater watershad in Poland/Retencja Leśna Zlewni Jako Element Ochrony Przeciwpowodziowej. Civ Environ Eng Rep 20:59–70

Pawlik Ł, Migoń P, Owczarek P, Kacprzak A (2013) Surface processes and interactions with forest vegetation on a steep mudstone slope, Stołowe Mountains, SW Poland. CATENA 109:203–216

Peters NE, Freer J, Beven K (2003) Modelling hydrologic responses in a small forested catchment (Panola Mountain, Georgia, USA): a comparison of the original and a new dynamic TOPMODEL. Hydrol Process 179(2):345–362

Piasecki J (1996) Wybrane cechy klimatu Masywu Śnieżnika. In: Jahn A, Kozłowski S, Pulina M (eds) Masyw Śnieżnika—zmiany w środowisku przyrodniczym. PAE, Warszawa, pp 189–218

Piñol J, Beven K, Freer J (1997) Modelling the hydrological response of mediterranean catchments, Prades, Catalonia. The use of distributed models as aids to hypothesis formulation. Hydrol Process 11(9):1287–1306

Quinn PF, Beven KJ (1993) Spatial and temporal predictions of soil moisture dynamics, runoff, variable source areas and evapotranspiration for plynlimon, mid-wales. Hydrol Process 7(4):425–448

Robson A, Beven K, Neal C (1992) Towards identifying sources of subsurface flow: a comparison of components identified by a physically based runoff model and those determined by chemical mixing techniques. Hydrol Process 6(2):199–214

Romanowicz RJ (2007) Data based mechanistic model for low flows: implications for the effects of climate change. J Hydrol 336:74–83

Romanowicz R, Beven K (2003) Estimation of flood inundation probabilities as conditioned on event inundation maps. Water Resour Res 39(3):1073–1085

Schmuck A (1969) Meteorologia i klimatologia dla WSR. PWN, Warszawa

Shrestha S, Bastola S, Babel MS, Dulal KN, Magome J, Hapuarachchi HAP, Takeuchi K (2007) The assessment of spatial and temporal transferability of a physically based distributed hydrological model parameters in different physiographic regions of Nepal. J Hydrol 347:153–172

Sigdel A, Jha R, Bhatta D, Abou-Shanab RAI, Sapireddy VR, Jeon B-H (2011) Applicability of TOPMODEL in the catchments of Nepal: Bagmati River Basin. Geosyst Eng 14(4):181–190

Sun S, Deng H (2004) A study of rainfall-runoff response in a catchment using TOPMODEL. Adv Atmos Sci 21(1):87–95

Szalińska W, Tokarczyk T, Jełowicki J, Chorążyczewski A, Michalski A, Tiukało A, Ostojski M (2014) Środowisko obliczeniowe operacyjnego modelu typu opad-odpływ. Monografie Komitetu Gospodarki Wodnej PAN XX, pp 293–306

Taschner S (2003) Flood modelling in the Ammer watershed using coupled meteorological and hydrological models, Ph.D. Thesis, Ludwig-Maximilians-Universität München, Germany

Waroszewski J, Kalinski K, Malkiewicz M, Mazurek R, Kozlowski G, Kabala C (2013) Pleistocene–holocene cover-beds on granite regolith as parent material for Podzols—an example from the Sudeten Mountains. CATENA 104:161–173

Wieczorek M, Migoń P (2014) Automatic relief classification versus expert and field based landform classification for the medium-altitude mountain range, the Sudetes, SW Poland. Geomorphology 206:133–146

Wolock DM, Hornberger GM, Musgrove TJ (1990) Topographic effects on flow path and surface water chemistry of the Llyn Brianne catchments in Wales. J Hydrol 115(1–4):243–259

Wood EF, Sivapalan M, Beven K, Band L (1988) Effects of spatial variability and scale with implications to hydrologic modeling. J Hydrol 102:27–47

Acta Geophysica (2018) 66:223–232
https://doi.org/10.1007/s11600-018-0130-5

RESEARCH ARTICLE - HYDROLOGY

Climate-driven seasonal geocenter motion during the GRACE period

Hongyue Zhang[1,2] · **Yu Sun**[3,4]

Received: 13 October 2017 / Accepted: 16 March 2018 / Published online: 23 March 2018

Abstract

Annual cycles in the geocenter motion time series are primarily driven by mass changes in the Earth's hydrologic system, which includes land hydrology, atmosphere, and oceans. Seasonal variations of the geocenter motion have been reliably determined according to Sun et al. (J Geophys Res Solid Earth 121(11):8352–8370, 2016) by combining the Gravity Recovery And Climate Experiment (GRACE) data with an ocean model output. In this study, we reconstructed the observed seasonal geocenter motion with geophysical model predictions of mass variations in the polar ice sheets, continental glaciers, terrestrial water storage (TWS), and atmosphere and dynamic ocean (AO). The reconstructed geocenter motion time series is shown to be in close agreement with the solution based on GRACE data supporting with an ocean bottom pressure model. Over 85% of the observed geocenter motion time series, variance can be explained by the reconstructed solution, which allows a further investigation of the driving mechanisms. We then demonstrated that AO component accounts for 54, 62, and 25% of the observed geocenter motion variances in the *X*, *Y*, and *Z* directions, respectively. The TWS component alone explains 42, 32, and 39% of the observed variances. The net mass changes over oceans together with self-attraction and loading effects also contribute significantly (about 30%) to the seasonal geocenter motion in the *X* and *Z* directions. Other contributing sources, on the other hand, have marginal (less than 10%) impact on the seasonal variations but introduce a linear trend in the time series.

Keywords Geocenter motion · Mass transportation · GRACE · Degree 1 coefficients

Introduction

Geocenter motion is usually defined as the relative movement of the center-of-mass (CM) of the entire Earth system with respect to the center-of-figure (CF) of the solid Earth surface (e.g., Petit and Luzum 2010; Ray 1999). Its non-tidal portion is primarily driven by redistributing masses both on the Earth's surface and inside the solid Earth. Detectable solid Earth contributions including those due to glacial isostatic adjustment (GIA) and very large earthquakes only affect the linear trend estimates of the geocenter motion time series. Mass changes in the Earth's outer fluid layers, such as atmosphere, oceans, terrestrial water, glaciers, and ice sheets, are responsible for both seasonal and linear variations in the geocenter motion time series.

In the CF reference frames, geocenter motion is equivalent to the degree 1 terms (C_{10}, C_{11}, and S_{11}) of the temporal variations of the gravity field. It can be determined with traditional geodetic techniques such as the satellite laser ranging (SLR), the global navigation satellite system (GNSS), and the Doppler orbiography and radiopositioning integrated by satellite (DORIS). These techniques are called direct methods, since they use ground tracking stations anchored to the solid Earth to observe satellites orbiting about the CM and can directly build a link between the CM and the CF (e.g., Cheng et al. 2013; Meindl et al. 2013; Feissel-Vernier et al. 2006). However, the quality of the corresponding products is limited by the deficiency in modeling the satellite dynamics and the so-called network effect. The network effect affects all direct

---

✉ Yu Sun
jade.yusun@outlook.com

[1] Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China

[2] University of Chinese Academy of Sciences, Beijing, China

[3] Key Lab of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University, Fuzhou, China

[4] Geoscience and Remote Sensing Department, Delft University of Technology, Delft, The Netherlands

methods, since the employed ground tracking networks are sparse or not evenly distributed that they are realizing the center-of-network (CN) rather than the CF (Wu et al. 2003). It is reported recently that the network effect can be as large as the geocenter motion itself (Zannat and Tregoning 2017a, b). Geocenter motion can also be inverted if the global solid Earth deformation is measured at GPS sites (Blewitt et al. 2001). This is because the translation of the geocenter is always companioned with a unique degree 1 load induced solid Earth deformation (Blewitt 2003). Other variants of the GPS inversion method are emerged to overcome several problems with this approach. For example, GRACE data are incorporated to better isolate the degree 1 related signals and ocean bottom pressure (OBP) model predictions are used as pseudo observations to improve the data coverage (e.g., Kusche and Schrama 2005; Wu et al. 2006). Swenson et al. (2008) proposed to estimate degree 1 coefficients by combing GRACE data with an OBP model. Such a method has been further developed by Sun et al. (2016a) to also estimate J_2 coefficients. The resulting J_2 solution is well compared to the SLR solutions, which are currently the most accurate source for J_2 (Meyrath et al. 2017). Sun et al. (2016b) further verified and refined the estimates by implementing a simulation study and optimizing the parameter settings. The geocenter motion estimates are proved to be accurate in terms of estimating surface mass changes when companioned with GRACE data (Sun et al. 2017). Hereafter, such a method is called the GRACE-OBP approach.

Finally, geocenter motion can be reconstructed with geophysical models. Many researchers have compared observed geocenter motion time series with their reconstructed solutions based on modeled mass variations of atmosphere, oceans, terrestrial water, glaciers and ice sheets, etc. (e.g., Dong et al. 1997; Chen et al. 1999; Dong et al. 2014; Collilieux et al. 2010). However, all the former efforts compared modeled geocenter motion with those based on solutions from the direct method, such as SLR and DORIS. The geocenter motion time series is poorly compared due to both systematic errors in the observations and deficiencies in the less constrained geophysical models. In this study, we reconstructed a new geocenter motion time series based on available state-of-the-art geophysical models and compared it with the GRACE-OBP solution. By doing so, we obtain in-depth knowledge of the geophysical interpretation of geocenter motion. According to Sun et al. (2016b), the linear trends in geocenter motion time series from the GRACE-OBP method are largely dependent on the applied GIA model, which still carry large uncertainties. Therefore, the comparison is focused on the seasonal geocenter motion arising from the present-day mass transport (PDMT). Linear trends in the geocenter motion time series originate from the solid Earth, and the mass loading is not considered at the time being.

Geocenter motion time series from the GRACE-OBP method

We have implemented the GRACE-OBP approach using the implementation parameters suggested by Sun et al. (2016b) in order to optimize the estimates of the seasonal variations in the resulting geocenter motion time series. In that study, the authors identified those parameters with an end-to-end simulation. That is, self-attraction and loading (SAL) effects are activated when distributing ocean waters; a buffer zone of 200 km wide is used to avoid signal leakage from continent to oceans due to the coarse spatial resolution of the GRACE data; the input GRACE solutions are truncated at degree 45. These parameter settings are applicable for all GRACE solutions.

The GRACE-OBP method requires two data sets including the GRACE data and the degree 1 component of an OBP model. The input GRACE data are in the form of Stoke coefficients directly taken from the level-2 product (known as GSM). To obtain those GSM coefficients, high-frequency atmospheric and dynamic ocean effects are modeled and subtracted from raw GRACE data at early stages using the Atmosphere and Ocean Dealiasing level-1B (AOD1B) product (Flechtner and Dobslaw 2013). Therefore, GSM coefficients should contain no atmospheric and dynamic ocean (AO) effects provided that the AOD1B product is perfect. Monthly averages of AOD1B coefficients are stored in the GAC files. The projection of GAC on the oceans is an OBP model and stored in the GAD files. According to Swenson et al. (2008), when working with GSM coefficients, the degree 1 coefficients of the GAD need to be removed from the input degree 1 coefficients of the applied OBP model for consistency. Therefore, if we apply the OBP model extracted from the AOD1B product as the input of the GRACE-OBP method, the input OBP degree 1 coefficients becomes zeros. At the same time, the obtained geocenter motion contains no AO effects, and need to add back the GAC degree 1 coefficients to obtain the full geocenter motion.

In this study, we obtain geocenter motion time series based on the latest GRACE solutions from three official centers, namely, the Center for Space Research (CSR) RL05 solution (Bettadpur 2012), the Geo Forschungs Zentrum (GFZ) RL05a solution (Dahle et al. 2013), and the Jet Propulsion Laboratory (JPL) RL05 solution (Watkins 2012) (see Fig. 1). All three solutions are using the same AOD1B product, and the input OBP degree 1 coefficients are also the same. Therefore, the resulting geocenter motion time series are somewhat correlated. The

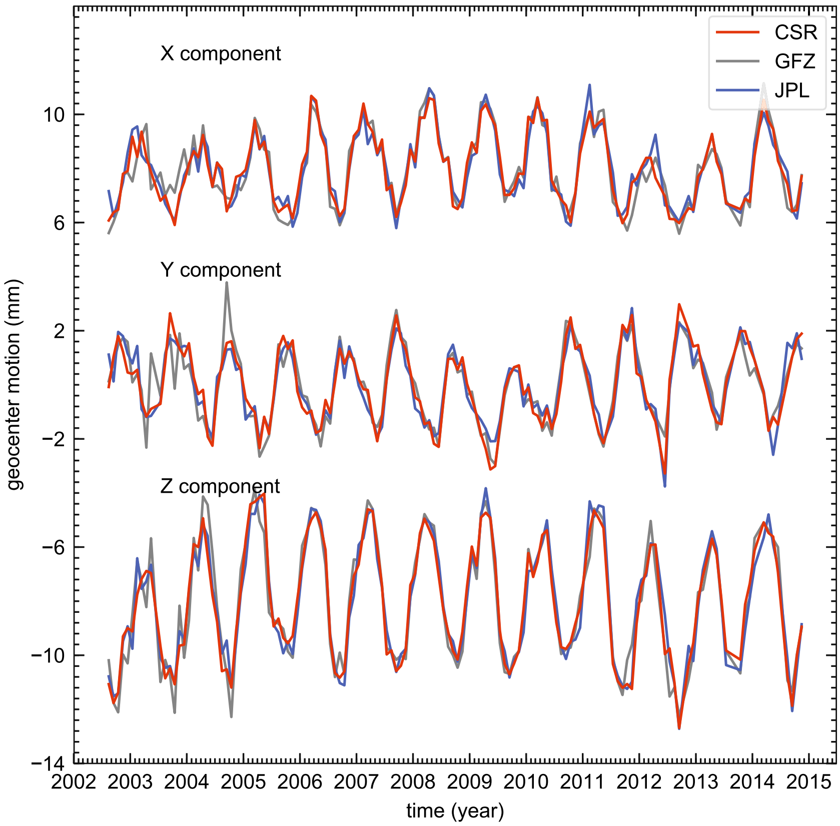


Fig. 1 Geocenter motion time series estimated from the GRACE-OBP approach. The gray, blue, and red time series are results based on GFZ RL05, JPL RL05, and CSR RL05 GRACE solutions, respectively. The linear trends are removed from all time series, which implies that the GIA signals are not included. The AO effects predicted by the dealiasing product AOD1B are not restored

good comparison of the results, however, show that the geocenter motion estimates are not sensitive to the choice of GRACE solutions. We believe that the average of the three geocener motion time series should be sufficiently representative.

It is now clear that the AO mass variations are the monthly averages derived from the AOD1B product. The AOD1B product is based on the European Center for Medium-Range Weather Forecasts (ECMWF) (Dee et al. 2011) and the Ocean Model for Circulation and Tides (OMCT) (Thomas 2002), which is driven by the ECMWF. The contributions of the AO mass variations are shown in Fig. 2. As mentioned above, the total geocenter motion time series can be found by restoring these contributions to the GRACE-OBP solution. The AO mass variations contribute significantly to the climate-driven seasonal

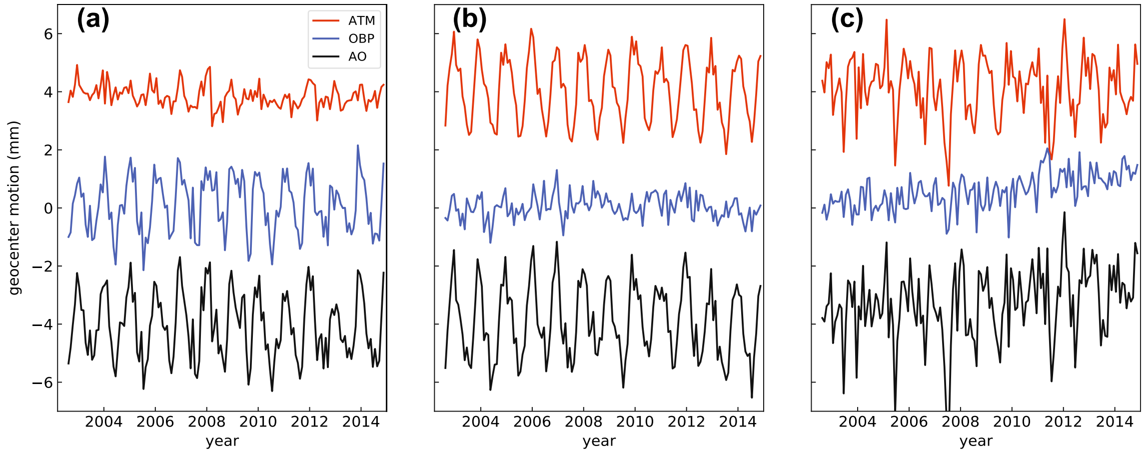


Fig. 2 Geocenter motion due to AO. **a**–**c** *X*, *Y*, and *Z* geocenter motion components, respectively. OBP denotes the ocean bottom pressure, which is based on the GAD product. Note that the atmospheric contribution over oceans is also included due to the inverted barometer effect. ATM represents the GAC-GAD and represent the contribution of continental atmosphere. AO is the combination of OBP and ATM

geocenter motion. The GRACE-OBP approach assumes that the AO-induced geocenter motion is known and accurate. This, however, is not necessarily true. Sun et al. (2016b) implemented the GRACE-OBP method while using an alternative OBP model, i.e., the estimating the circulation and climate of the ocean (ECCO) (Fukumori 2002; Kim et al. 2007). They found that the annual amplitudes of degree 1 coefficients are different within 15%. For annual phase estimates, the largest differences are 11 days. The differences caused using two OBP models are not significant, but still need to be further analyzed in future studies. This, however, is not the primary purpose of this study.

Geocenter motion predicted by geophysical models

It is well known that seasonal geocenter motion is primarily driven by surface mass loading, which can be attributed to multiple sources. The AO effects have already been given in "Geocenter motion time series from the GRACE-OBP method". Here, we consider the contributions of mass variations over both Antarctica and Greenland ice sheets, continental glaciers, terrestrial water storage, as well as total ocean mass changes. By summing up all these contributors, one should be able to reconstruct the GRACE-OBP-based geocenter motion solution.

We first obtain mass loading fields represented in terms of equivalent water height from geophysical models on a monthly basis. Then, the degree 1 mass coefficients are estimated by performing spherical harmonic analysis and converted to geocenter motion estimates. In the following, we give a brief introduction of the geophysical models used.

Ice sheets and continental glaciers

Mass changes over ice sheets as well as continental glaciers can be obtained from several methodologies. It can be estimated using GRACE gravimetry measurements (Jacob et al. 2012), altimetry measurements (Zwally et al. 2005) and the input–output method (Rignot et al. 2008). However, solutions from all techniques have reconciled estimates of ice-sheet mass balance recently (Shepherd et al. 2012).

Here, for both Antarctica and Greenland ice sheets, we use the input–output method to estimate their mass variations following Frederikse et al. (2016). The input comes from the Regional Atmospheric Climate Model (RACMO) 2.3 (van de Berg and Medley 2016; Noël et al. 2015; Wessem et al. 2014) surface mass balance (SMB), while the output is the ice discharge acceleration which can be taken from van den Broeke et al. (2016) for Greenland (6.6 Gt year^{-1}) and Rignot et al. (2011) for Antarctica (9.0 Gt year^{-1}). The contribution of ice sheets is further validated by the estimates from the Ice-sheet Mass Balance Intercomparison Exercise (IMBIE) project and GRACE data. The Antarctica and Greenland ice-sheet mass variations are denoted as ANT and GRE hereafter.

For continental glaciers mass balance (GLA), we use the estimates based on Marzeion et al. (2015). So far, only annual observations in the glacier mass balance are available. Although monthly data are obtained based on linear interpolation, it hardly contributes to any seasonal changes of the geocenter motion. However, we still decided to include it for completeness.

Terrestrial water storage

Terrestrial water storage (TWS) includes the contribution from soil moisture, snow, surface water, and groundwater, etc. It is critical in the global hydrology system. Due to the lack of global-scale data constraints, the TWS is currently modeled with data assimilation techniques. As a result, the large-scale mass redistribution derived from a TWS model may not be accurate. Here, we choose the Global land data assimilation system (GLDAS) products, which currently contains only the snow, soil moisture, and canopy water. The groundwater as well as separate surface water components such as rivers and lakes, on the other hand, are not included. GLDAS has four different models including Noah model (NOAH), the community land model (CLM), the Mosaic model (MOS), and the variable infiltration capacity model (VIC) (Rodell et al. 2004). We integrated the snow, canopy water, and all soil moisture layers for each of the four models to obtain the total water content. The uncertainties of these models are not available and we expect using the average of them allows for the removal of some random noises in the geocenter motion time series. However, the CLM model failed to predict the seasonal variabilities in the geocenter motion time series (see Fig. 3). Therefore, we take the average of NOAH, VIC, and MOS models as the final solution of the GLDAS model.

Another commonly used model, the WaterGAP Global Hydrology Model (WGHM), is also used for the purpose of mutual comparisons. In this study, the WGHM TWS is kindly provided by Müller Schmied (personal communication) based on the latest WaterGAP 2.2 c model (Müller Schmied et al. 2014, 2016). Different from the GLDAS model, the WGHM model also computes the variations in groundwater storage.

It is also worth noting that Greenland and its periphery area (300 km) are masked out from both models as snow, and ice dynamics are poorly modeled there.

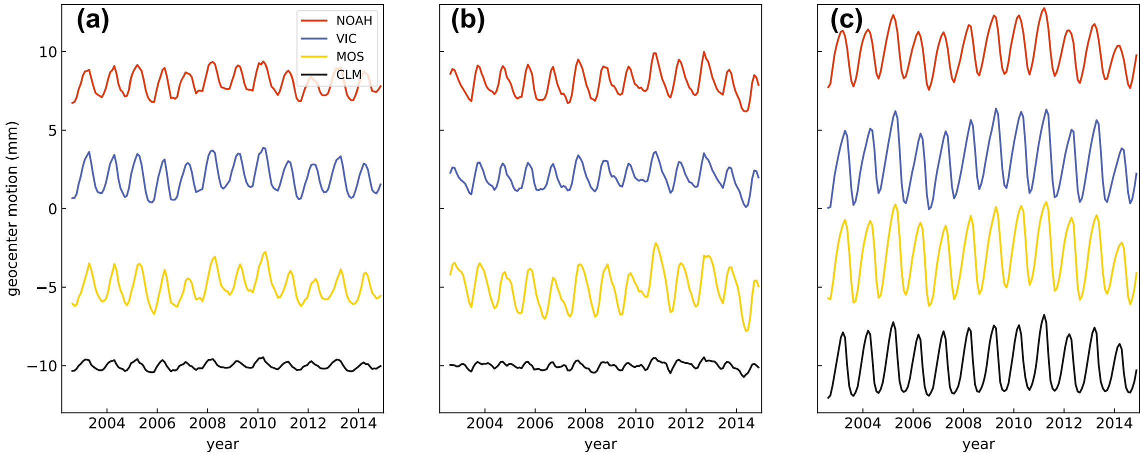


Fig. 3 Geocenter motion due to GLDAS NOAH, VIC, MOS, and CLM models

Self-attraction and loading effects

The geophysical models adopted in this study are not coupled meaning that the sum of all these models does not ensure the mass conservation of the entire Earth system. This problem was usually ignored in former efforts and is recently solved by applying the so-called Greatbatch correction. This correction simply adds/removes a thin uniform layer of water from oceans to balance the mass variations from all other components combined (e.g., Swenson et al. 2008; Dong et al. 2014). However, such a uniform distribution is not realistic as it ignores the impact of SAL effects. Regional mass redistributes modify the Earth's gravity field and ocean waters will passively redistribute to align with the newly formed equipotential surface. The water flows in to or out from the ocean basins will not be distributed over oceans as a uniform layer but following certain patterns known as fingerprints (e.g., Tamisiea et al. 2010), which can be computed through the sea-level equation (e.g., Farrell and Clark 1976; Mitrovica et al. 2001). Therefore, we take into account these effects to obtain more realistic geocenter motion estimates due to total ocean mass variations (OCN).

Results

In Fig. 4, we show the five considered contributing sources of geocenter motion excluding the AO component. The most important contributing source of the seasonal signals in geocenter motion is the TWS and the OCN. As discussed above, the OCN needs to balance out the sum of mass changes from all other components to ensure the total mass conservation in the Earth system, the OCN-induced geocenter motion is thus strongly correlated with the combination of other contributing sources. we show the geocenter motion time series derived from GLDAS and WGHM models (see Fig. 5). Both solutions are featured with clear and similar seasonal variations. Switching only the hydrology models used for the TWS component while keeping all other geophysical models unchanged will also change the OCN-induced geocenter motion accordingly (see Fig. 6).

In Fig. 7, we show the reconstructed geocenter motion time series by combining all the components shown in Fig. 4. On top of the reconstructed time series, the GRACE-OBP solution is also shown as the reference. We further estimate the amplitude and phase of the annual variations of the geocenter motion to facilitate a more detailed comparison between the reconstructed and observed solutions (see Table 1). In Table 2, we calculate the variance of the observed geocenter motion explained by individual components of the considered contributing sources. The explained variance (R) is defined as $R = 1 - \mathrm{var} < \mathrm{obs} - \mathrm{rec} > / \mathrm{var} < \mathrm{obs} >$, where obs represent the GRACE-OBP solution after restoring the GAC product (the AO effects), rec denotes the reconstructed solution).

In the upper row of Fig. 7, we show the reconstructed geocenter motion solutions with (Panel a) and without (Panel b) adding back the AO effects. The TWS component is estimated from the GLDAS model. In the lower row, the only difference from the upper ones is that the TWS component is modeled using the WGHM model. Without considering the AO effects, the observed GSM-like geocenter motion time series are well recovered by the selected geophysical contributions in the X and Z directions (Panels a, c), with the reconstructed solution explaining about 80% of the variance in the observed one (see Table 2). All the main features in the observed geocenter motion time series are well explained. The annual amplitude and phase are also relatively close. The mean annual

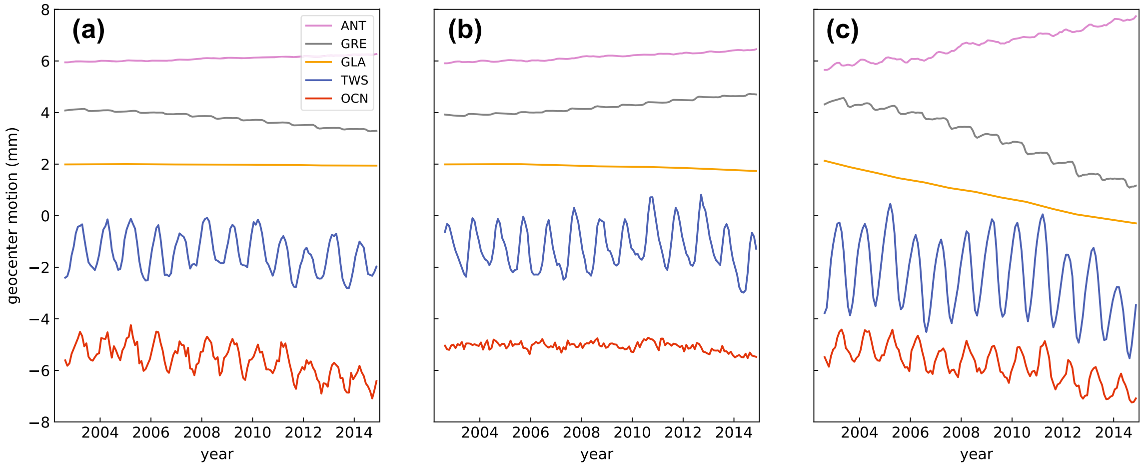


Fig. 4 Contributing sources of the GSM-like geocenter motion. GSM-like means that the AO component is not included

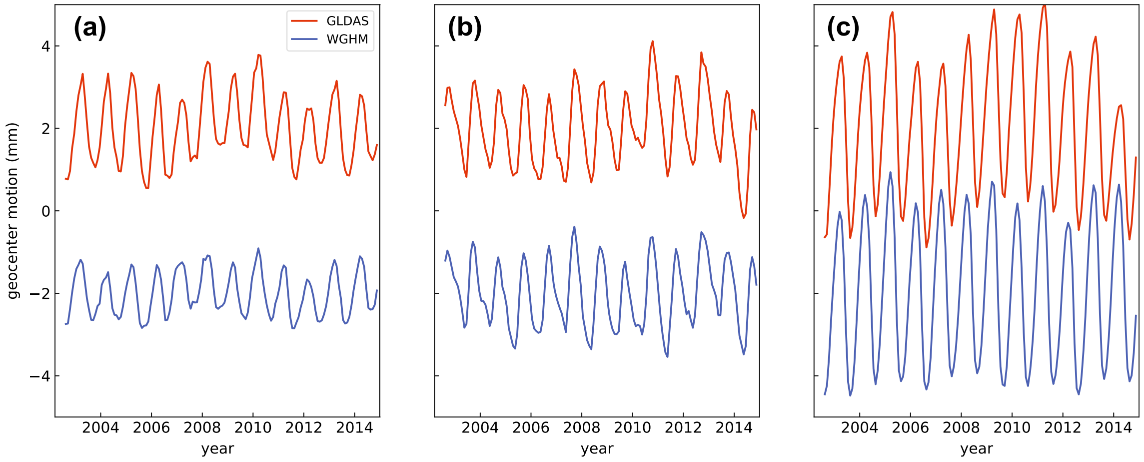


Fig. 5 Geocenter motion due to GLDAS and WGHM models

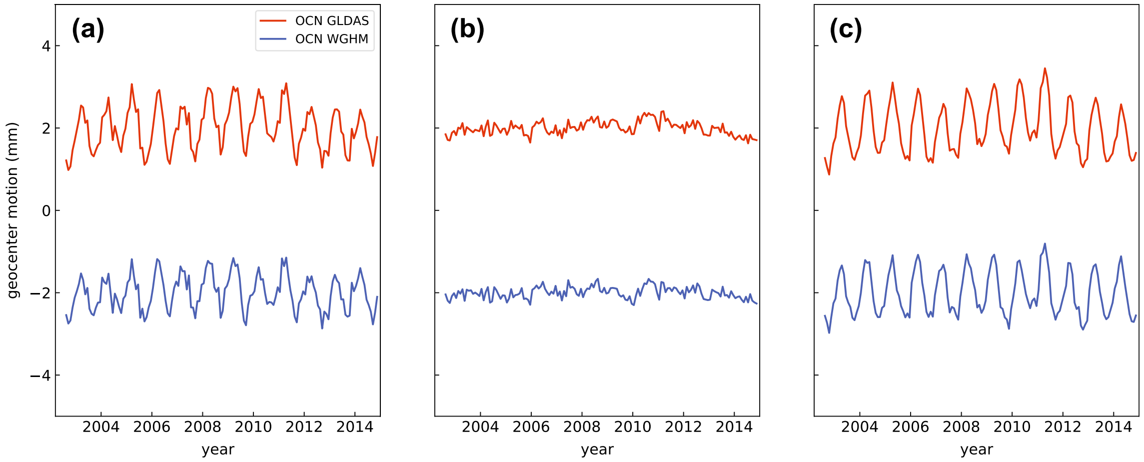


Fig. 6 Geocenter motion due to OCN component

amplitudes of the two reconstructed solutions are 1.47 and 3.04 mm, respectively, for the *X* and *Z* component, which are quite close to those from the observed solution (1.56 and 2.82 mm). The annual phases of either of the two reconstructed solutions are also close to the observed solution (within 2 weeks). Unfortunately, the agreement is

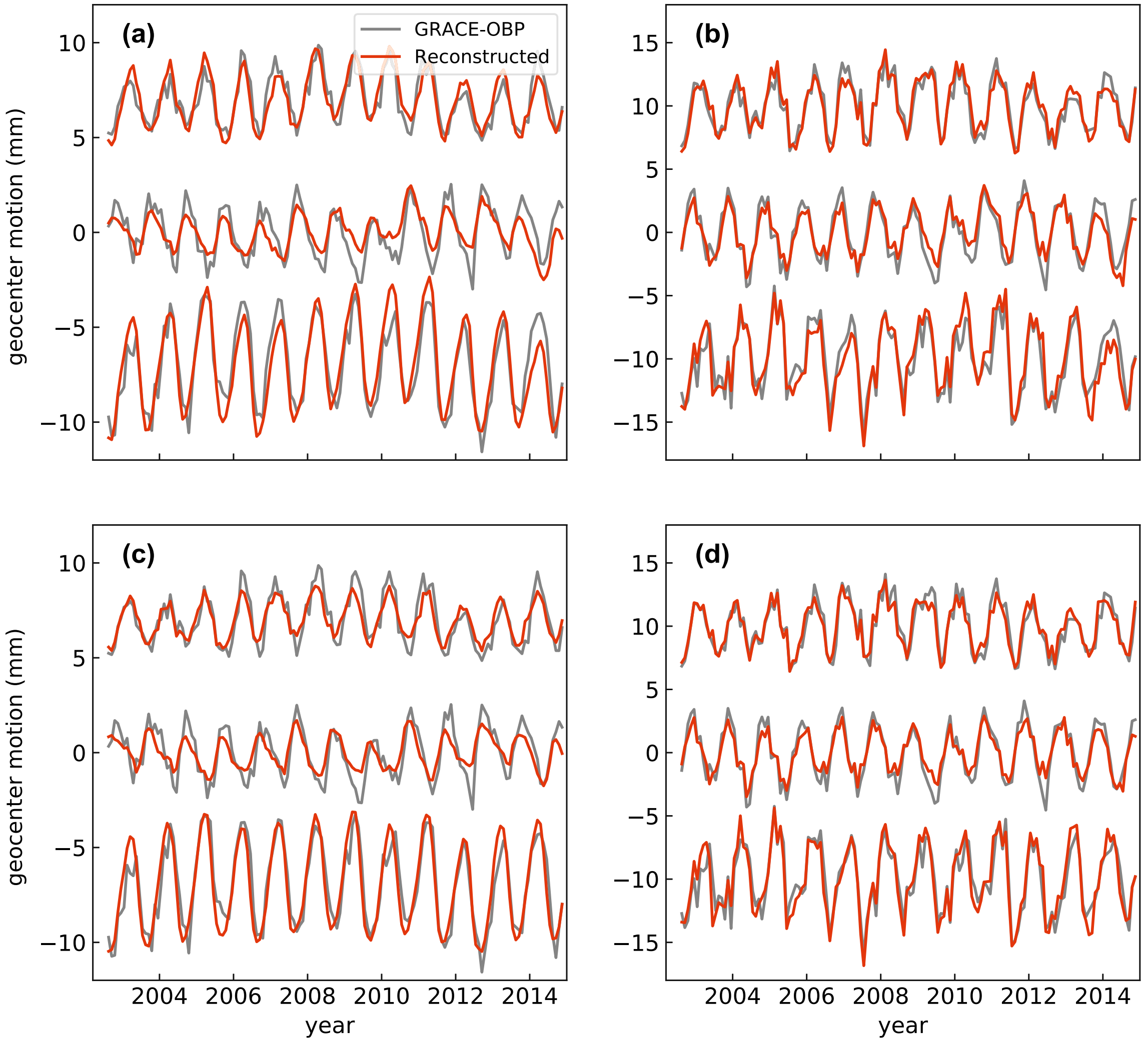


Fig. 7 Comparison between the reconstructed and the estimated geocenter motion time series. Geocenter components *X*, *Y* and *Z* are shown in the top, middle, and bottom of the panels. **a**, **c** The GSM-like time series which do not include the AO effects. **b**, **d** The total geocenter motion time series, which are the GSM-like geocenter motion time series plus the AO effects. The only difference between the upper panels and lower panels is the TWS component used to obtain the reconstructed solutions. The GLDAS model is used for the plots in the upper panels, while the WGHM model is used in the lower ones

Table 1 Annual variations of the reconstructed and GRACE-OBP-based geocenter motion time series

	X (C_{11})		Y (S_{11})		Z (C_{10})	
	Amp (mm)	Pha (day)	Amp (mm)	Pha (day)	Amp (mm)	Pha (day)
GRACE-OBP	1.56 ± 0.08	86 ± 3	1.60 ± 0.07	301 ± 3	2.82 ± 0.09	86 ± 2
Reconstructed GLDAS	1.72 ± 0.06	83 ± 2	1.00 ± 0.05	294 ± 4	2.96 ± 0.08	81 ± 2
Reconstructed WGHM	1.22 ± 0.04	72 ± 2	0.98 ± 0.05	279 ± 3	3.12 ± 0.05	72 ± 1
TOT GRACE-OBP	2.28 ± 0.13	52 ± 3	2.79 ± 0.09	327 ± 2	2.92 ± 0.18	69 ± 3
TOT reconstructed GLDAS	2.47 ± 0.10	51 ± 2	2.22 ± 0.09	331 ± 2	3.12 ± 0.16	65 ± 3
TOT reconstructed WGHM	2.19 ± 0.10	39 ± 3	2.04 ± 0.07	327 ± 2	3.41 ± 0.14	57 ± 2

Note that the uncertainties shown are formal errors, which are estimated from the post-fit residuals and do not reflect the real uncertainty of the estimates

Table 2 Variance of the observed geocenter motion explained by the reconstructed solutions and contributing sources

	GLDAS			WGHM		
	X (%)	*Y* (%)	*Z* (%)	*X* (%)	*Y* (%)	*Z* (%)
GSM	78	61	80	79	67	85
TOT	90	85	85	90	88	88
AO	54	62	25	–	–	–
ANT	1	0	3	–	–	–
GRE	2	1	7	–	–	–
GLA	0	0	0	–	–	–
TWS	42	42	39	36	33	67
OCN	36	−4	29	32	−2	29

less well in the *Y* direction as only about 60% of the variance has been explained. This is also reflected in the annual amplitude estimate (1.6 mm for the observed solution and only about 1 mm for the reconstructed solutions). Since the seasonal variations mainly come from the TWS component in this direction, the discrepancies can be largely attributed to the model deficiencies in both GLDAS and WGHM models. After restoring the AO effects (Panel b, d), we notice that the full geocenter motion time series variance can be explained by over 85% using either of the two hydrology models. The annual amplitudes and phases are in line with solutions based on other techniques and thus can be considered as reasonable.

In view of the good agreement of the reconstructed geocenter motions and the observed solution, it is possible for us to quantify the contribution of each component in the Earth system to the total geocenter motion. We show this result also in terms of explained variance (see Table 2), but based on a slightly different definition, i.e., $R = 1 - \mathrm{var} < \mathrm{obs} - \mathrm{com} > / \mathrm{var} < \mathrm{obs} >$, with com the individual contributing sources. The melting of the continental glaciers and the polar ice sheets dominate the linear trend estimates in the geocenter motion time series (see Fig. 4). They, however, have only minor contributions to the seasonal signals in the geocenter motion time series. Glaciers contribute negligibly to all three components as expected. Ice sheets have a slightly larger contribution in the *Z* direction, but still explain less than 10% of the observed variance.

Atmosphere and dynamic ocean is the largest contributing source for seasonal geocenter motion in both *X* and *Y* components. When combined, they explain 54 and 62% of the variance in those two directions, respectively. In the *Z* direction, it accounts for 25% of the variance. TWS is another major contributing source. Based on the GLDAS model, it explains about 40% of the observed variances in the *X*, *Y*, and *Z* directions. However, the *Z* geocenter motion is much better explained when using the WGHM model (67%). This is probably because the WGHM model also models the groundwater component. Finally, ocean mass variations due to ocean–land mass exchanges also play a critical role in recovering the observed *X* and *Z* geocenter time series. Such a contribution is responsible for about 30% of the observed *X* and *Z* component variances, respectively. The minor impact on the *Y* component is expected, since the *Y* component of geocenter motion is mostly influenced by continental mass changes rather than oceanic ones (Chen et al. 1999).

Conclusions and discussion

In this study, we use geophysical models to reproduce the seasonal geocenter motion time series estimated from the GRACE-OBP approach. For the time being, AO effects modeled by the GAC products are assumed to be accurate. Under this assumption, we conclude that the AO and TWS components and the associating ocean responses are the main contributors to the seasonal variations in geocenter motion. They combined explain almost the entire seasonal variations in geocenter motion time series. The contribution from ANT and GRE, on the other hand, is rather limited. Currently, we are still not able to accurately determine the contribution of continental glaciers due to the low temporal resolution of the observations.

The geocenter motion in the *Y* direction, unlike those in the other two directions, is not that well explained by the selected geophysical models. This is probably due to the lack or mis-modeling of groundwater component in the GLDAS and WGHM models. Note that WGHM claims to contain the groundwater component and works indeed better than that of the GLDAS model. This becomes evident as the reconstructed geocenter motion involving the WGHM model explains the observed solution slightly better. However, the groundwater component, at least its low-frequency components, are not accurate enough, since the large discrepancy between the reconstructed and observed *Y* direction geocenter motion time series is still exists. Such groundwater component significantly affects the estimation of the geocenter motion in the *Y* direction, whereas it has only secondary effects on the other two directions. This is expected, since the groundwater is trapped over land areas and may only affect the OCN component through the SAL effects, which are likely to be minor. Anthropological impacts, such as groundwater depletion and building dams, could also have some effects on the geocenter motion, but likely restricted to the linear trend estimates. To a lesser extent, the discrepancies may also relate to the omission of the seasonal variabilities in the GLA component. Mass variations in the GLA component will directly affect the total ocean mass changes and thus affects the geocenter motion in all three directions. The good agreement in the *X* and *Z* of the reconstructed and the observed geocenter motion time series indicates the GLA has a rather limited contribution to the seasonal signals in the geocenter motion time series.

The modeled AO effects are used in both the GRACE-OBP solution and the reconstructed solution. Therefore, this study is not able to conclude if the observed AO-induced geocenter motion and the remodel-predicted solution are reconciled or not. To do so, a geocenter motion solution independent from any geophysical models is better suited for such a comparison study. We recently noted that Wu et al. (2017) provided a geocenter motion time series largely based on real data which may be used in future investigations.

In this study, we exclude the solid Earth contributions to geocenter motions by removing the trend in the time series. The side effect of doing so is that we also remove the trend due to the surface mass transport. Since GIA models are very uncertain, it is premature to discuss the geocener motion trend using the GRACE-OBP approach. However, there are already approaches that allow us to obtain promising geocenter motion trend (e.g., Wu et al. 2010; Rietbroek et al. 2012, 2016), which may be suited to study the contributing sources of geocenter motion trend in the next step.

References

Bettadpur S (2012) UTCSR Level-2 Processing Standards Document. Technical Version 4, Univ. Texas, Austin

Blewitt G (2003) Self-consistency in reference frames, geocenter definition, and surface loading of the solid Earth. J Geophys Res Solid Earth 108(B2):2103. https://doi.org/10.1029/2002JB002082

Blewitt G, Lavallée D, Clarke P, Nurutdinov K (2001) A new global mode of Earth deformation: seasonal cycle detected. Science (New York, NY) 294(5550):2342–2345. https://doi.org/10.1126/science.1065328

van den Broeke M, Enderlin E, Howat I, Kuipers Munneke P, Noël B, van de Berg WJ, van Meijgaard E, Wouters B (2016) On the recent contribution of the Greenland ice sheet to sea level change. Cryosphere Discuss 2016:1–26. https://doi.org/10.5194/tc-2016-123

Chen JL, Wilson CR, Eanes RJ, Nerem RS (1999) Geophysical interpretation of observed geocenter variations. J Geophys Res Solid Earth 104(B2):2683–2690. https://doi.org/10.1029/1998JB900019

Cheng MK, Ries JC, Tapley BD (2013) Geocenter variations from analysis of SLR data. In: Altamimi Z, Collilieux X (eds) Reference frames for applications in geosciences, no. 138 in International Association of Geodesy Symposia, Springer, Berlin, pp 19–25

Collilieux X, Altamimi Z, Coulot D, van Dam T, Ray J (2010) Impact of loading effects on determination of the International Terrestrial Reference Frame. Adv Space Re 45(1):144–154. https://doi.org/10.1016/j.asr.2009.08.024

Dahle C, Flechtner F, Gruber C, König D, König R, Michalak G, Neumayer KH, GFZ DG (2013) GFZ GRACE level-2 processing standards document for level-2 product release 0005. Deutsches GeoForschungsZentrum GFZ

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597. https://doi.org/10.1002/qj.828

Dong D, Dickey JO, Chao Y, Cheng MK (1997) Geocenter variations caused by atmosphere, ocean and surface ground water. Geophys Res Lett 24(15):1867–1870. https://doi.org/10.1029/97GL01849

Dong D, Qu W, Fang P, Peng D (2014) Non-linearity of geocentre motion and its impact on the origin of the terrestrial reference frame. Geophys J Int 198(2):1071–1080. https://doi.org/10.1093/gji/ggu187

Farrell WE, Clark JA (1976) On postglacial sea level. Geophys J R Astron Soc 46(3):647–667. https://doi.org/10.1111/j.1365-246X.1976.tb01252.x

Feissel-Vernier M, Bail KL, Berio P, Coulot D, Ramillien G, Valette JJ (2006) Geocentre motion measured with DORIS and SLR, and predicted by geophysical models. J Geod 80(8–11):637–648. https://doi.org/10.1007/s00190-006-0079-z

Flechtner F, Dobslaw H (2013) AOD1b Product Description Document for Product Release. 05. GFZ German Research Centre for Geosciences

Frederikse T, Riva R, Kleinherenbrink M, Wada Y, van den Broeke M, Marzeion B (2016) Closing the sea level budget on a regional scale: trends and variability on the Northwestern European continental shelf. Geophys Res Lett 43(20):10864–10872. https://doi.org/10.1002/2016GL070750

Fukumori I (2002) A Partitioned Kalman filter and smoother. Mon Weather Rev 130(5):1370–1383. https://doi.org/10.1175/1520-0493(2002)130<1370:APKFAS>2.0.CO;2

Jacob T, Wahr J, Pfeffer WT, Swenson S (2012) Recent contributions of glaciers and ice caps to sea level rise. Nature 482(7386):514–518. https://doi.org/10.1038/nature10847

Kim SB, Lee T, Fukumori I (2007) Mechanisms controlling the interannual variation of mixed layer temperature averaged over the Niño-3 region. J Clim 20(15):3822–3843. https://doi.org/10.1175/JCLI4206.1

Kusche J, Schrama EJO (2005) Surface mass redistribution inversion from global GPS deformation and gravity recovery and climate experiment (GRACE) gravity data. J Geophys Res Solid Earth 110(B9). https://doi.org/10.1029/2004JB003556

Marzeion B, Leclercq PW, Cogley JG, Jarosch AH (2015) Brief communication: global reconstructions of glacier mass change during the 20th century are consistent. Cryosphere 9(6):2399–2404. https://doi.org/10.5194/tc-9-2399-2015

Meindl M, Beutler G, Thaller D, Dach R, Jäggi A (2013) Geocenter coordinates estimated from GNSS data as viewed by perturbation theory. Adv Space Res 51(7):1047–1064. https://doi.org/10.1016/j.asr.2012.10.026

Meyrath T, Rebischung P, van Dam T (2017) GRACE era variability in the Earth's oblateness: a comparison of estimates from six different sources. Geophys J Int 208(2):1126–1138. https://doi.org/10.1093/gji/ggw441

Mitrovica JX, Tamisiea ME, Davis JL, Milne GA (2001) Recent mass balance of polar ice sheets inferred from patterns of global sea-level change. Nature 409(6823):1026–1029. https://doi.org/10.1038/35059054

Müller Schmied H, Eisner S, Franz D, Wattenbach M, Portmann FT, Flörke M, Döll P (2014) Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. Hydrol Earth Syst Sci 18(9):3511–3538

Müller Schmied H, Adam L, Eisner S, Fink G, Flörke M, Kim H, Oki T, Portmann FT, Reinecke R, Riedel C et al (2016) Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. Hydrol Earth Syst Sci 20(7):2877–2898

Noël B, van de Berg WJ, van Meijgaard E, Kuipers Munneke P, van de Wal RSW, van den Broeke MR (2015) Evaluation of the updated regional climate model RACMO2.3: summer snowfall impact on the Greenland Ice Sheet. Cryosphere 9(5):1831–1844. https://doi.org/10.5194/tc-9-1831-2015.

Petit G, Luzum B (2010) IERS conventions (2010) Technical Note 36. Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main

Ray J (1999) IERS Analysis Campaign to Investigate Motions of the Geocenter. Technical Note 25, Central Bureau of IERS Observatioire de Paris, Paris

Rietbroek R, Fritsche M, Brunnabend SE, Daras I, Kusche J, Schröter J, Flechtner F, Dietrich R (2012) Global surface mass from a new combination of GRACE, modelled OBP and reprocessed GPS data. J Geodyn 59–60:64–71. https://doi.org/10.1016/j.jog.2011.02.003

Rietbroek R, Brunnabend SE, Kusche J, Schröter J, Dahle C (2016) Revisiting the contemporary sea-level budget on global and regional scales. Proc Natl Acad Sci 113(6):1504–1509. https://doi.org/10.1073/pnas.1519132113

Rignot E, Box JE, Burgess E, Hanna E (2008) Mass balance of the Greenland ice sheet from 1958 to 2007. Geophys Res Lett 35(20):L20,502. https://doi.org/10.1029/2008GL035417

Rignot E, Velicogna I, van den Broeke MR, Monaghan A, Lenaerts JTM (2011) Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise. Geophys Res Lett 38(5):L05,503. https://doi.org/10.1029/2011GL046583

Rodell M, Houser PR, Jambor U, Gottschalck J, Mitchell K, Meng CJ, Arsenault K, Cosgrove B, Radakovich J, Bosilovich M, Entin* JK, Walker JP, Lohmann D, Toll D (2004) The global land data assimilation system. Bull Am Meteorol Soc 85(3):381–394. https://doi.org/10.1175/BAMS-85-3-381

Shepherd A, Ivins ER, AG, Barletta VR, Bentley MJ, Bettadpur S, Briggs KH, Bromwich DH, Forsberg R, Galin N, Horwath M, Jacobs S, Joughin I, King MA, Lenaerts JTM, Li J, Ligtenberg SRM, Luckman A, Luthcke SB, McMillan M, Meister R, Milne G, Mouginot J, Muir A, Nicolas JP, Paden J, Payne AJ, Pritchard H, Rignot E, Rott H, Sørensen LS, Scambos TA, Scheuchl B, Schrama EJO, Smith B, Sundal AV, Angelen JHv, Berg WJvd, Broeke MRvd, Vaughan DG, Velicogna I, Wahr J, Whitehouse PL, Wingham DJ, Yi D, Young D, Zwally HJ (2012) A reconciled estimate of ice-sheet mass balance. Science 338(6111):1183–1189. https://doi.org/10.1126/science.1228102

Sun Y, Ditmar P, Riva R (2016a) Observed changes in the Earth's dynamic oblateness from GRACE data and geophysical models. J Geod 90(1):81–89. https://doi.org/10.1007/s00190-015-0852-y

Sun Y, Riva R, Ditmar P (2016b) Optimizing estimates of annual variations and trends in geocenter motion and J2 from a combination of GRACE data and geophysical models. J Geophys Res Solid Earth 121(11):8352–8370. https://doi.org/10.1002/2016JB013073

Sun Y, Ditmar P, Riva R (2017) Statistically optimal estimation of degree-1 and C20 coefficients based on GRACE data and an ocean bottom pressure model. Geophys J Int. https://doi.org/10.1093/gji/ggx241

Swenson S, Chambers D, Wahr J (2008) Estimating geocenter variations from a combination of GRACE and ocean model output. J Geophys Res Solid Earth 113(B8):B08,410. https://doi.org/10.1029/2007JB005338

Tamisiea ME, Hill EM, Ponte RM, Davis JL, Velicogna I, Vinogradova NT (2010) Impact of self-attraction and loading on the annual cycle in sea level. J Geophys Res Oceans 115(C7):C07,004. https://doi.org/10.1029/2009JC005687

Thomas M (2002) Ocean induced variations of Earth's rotation results from a simultaneous model of global circulation and tides. PhD thesis, Univ. of Hamburg, Germany

van de Berg WJ, Medley B (2016) Brief communication: upper-air relaxation in RACMO2 significantly improves modelled interannual surface mass balance variability in Antarctica. Cryosphere 10(1):459–463. https://doi.org/10.5194/tc-10-459-2016

Watkins M (2012) JPL Level-2 Processing Standards Document. Technical Version 4, Jet Propulsion Laboratory

Wessem JMV, Reijmer CH, Morlighem M, Mouginot J, Rignot E, Medley B, Joughin I, Wouters B, Depoorter MA, Bamber JL, Lenaerts JTM, Berg WJVD, Broeke MRVD, Meijgaard EV (2014) Improved representation of East Antarctic surface mass balance in a regional atmospheric climate model. J Glaciol 60(222):761–770. https://doi.org/10.3189/2014JoG14J051

Wu X, Heflin MB, Ivins ER, Argus DF, Webb FH (2003) Large-scale global surface mass variations inferred from GPS measurements of load-induced deformation. Geophys Res Lett 30(14). https://doi.org/10.1029/2003GL017546

Wu X, Heflin MB, Ivins ER, Fukumori I (2006) Seasonal and interannual global surface mass variations from multisatellite geodetic data. J Geophys Res Solid Earth 111(B9). https://doi.org/10.1029/2005JB004100

Wu X, Heflin MB, Schotman H, Vermeersen BLA, Dong D, Gross RS, Ivins ER, Moore AW, Owen SE (2010) Simultaneous estimation of global present-day water transport and glacial isostatic adjustment. Nat Geosci 3(9):642–646. https://doi.org/10.1038/ngeo938

Wu X, Kusche J, Landerer FW (2017) A new unified approach to determine geocentre motion using space geodetic and GRACE gravity data. Geophys J Int 209(3):1398–1402. https://doi.org/10.1093/gji/ggx086

Zannat UJ, Tregoning P (2017a) Estimating network effect in geocenter motion: applications. J Geophys Res Solid Earth 122(10):2017JB014,247. https://doi.org/10.1002/2017JB014247

Zannat UJ, Tregoning P (2017b) Estimating network effect in geocenter motion: Theory. J Geophys Res Solid Earth 122(10):2017JB014,246. https://doi.org/10.1002/2017JB014246

Zwally HJ, Giovinetto MB, Li J, Cornejo HG, Beckley MA, Brenner AC, Saba JL, Yi D (2005) Mass changes of the Greenland and Antarctic ice sheets and shelves and contributions to sea-level rise: 1992–2002. J Glaciol 51(175):509–527

Acta Geophysica (2018) 66:233–241
https://doi.org/10.1007/s11600-018-0129-y

RESEARCH ARTICLE - ATMOSPHERIC AND SPACE SCIENCES

Satellite orbit determination using quantum correlation technology

Bo Zhang[1] · **Fuping Sun**[1] · **Xinhui Zhu**[1] · **Xiaolin Jia**[2]

Received: 4 September 2017 / Accepted: 16 March 2018 / Published online: 23 March 2018

Abstract

After the presentation of second-order correlation ranging principles with quantum entanglement, the concept of quantum measurement is introduced to dynamic satellite precise orbit determination. Based on the application of traditional orbit determination models for correcting the systematic errors within the satellite, corresponding models for quantum orbit determination (QOD) are established. This paper experiments on QOD with the BeiDou Navigation Satellite System (BDS) by first simulating quantum observations of 1 day arc-length. Then the satellite orbits are resolved and compared with the reference precise ephemerides. Subsequently, some related factors influencing the accuracy of QOD are discussed. Furthermore, the accuracy for GEO, IGSO and MEO satellites increase about 20, 30 and 10 times, respectively, compared with the results from the resolution by measured data. Therefore, it can be expected that quantum technology may also bring delightful surprises to satellite orbit determination as have already emerged in other fields.

Keywords Quantum correlation · Entanglement · Satellite · Orbit determination

Introduction

With the development of quantum mechanics in recent decades, more and more countries and individuals have paid great attention to quantum technology. Considering the large amount of significant discoveries and applications it has already brought to human beings, there are grounds to believe that quantum technology will bring us more exciting surprises in the next 100 years. The fundamental topic of this paper is quantum entanglement, which is described as a technology of "ghost" that can instantly interrelate and interact with each other over very long distances. Giovannetti et al. (2001) proposed a positioning method utilizing quantum correlation techniques based on quantum entanglement. Later research (D'Ariano et al. 2001; Ozawa 2001; Giovannetti et al. 2002) shows that compared with traditional navigation and positioning technologies, quantum correlation techniques can not only substantially improve the security, channel capacity and efficiency of the information transmission, but also achieve higher system accuracy in a faster and more simplified process. Therefore, it is expected to become a strategic direction for the future development of satellite navigation technologies. On 16 August, 2016, the world's first quantum science experimental satellite "Mo-tse", independently developed by China, was successfully launched into orbit, which will help to realize the first quantum communication between the satellite and the ground, and make further studies on quantum correlation technology possible.

In the field of satellite navigation and positioning, the realization of quantum communication between satellites and the ground makes the development of quantum orbit determination (QOD) techniques possible. The precision of satellite orbits directly affects the positioning and navigation accuracy for the user terminals. The observation equation for orbit resolving is established based on measurements of distances between stations and satellites, therefore, distance measurements of high precision are essential for improving the accuracy of satellite orbits. As far as the current distance measurement approaches for orbit determination are concerned, there are pseudo range, carrier phase and satellite laser ranging (SLR). The pseudo range, as an observation, is easy to revise the correction models of itself but the ranging precision is relatively low

---

✉ Bo Zhang
linjiak@163.com

[1] Zhengzhou Institute of Surveying and Mapping, Zhengzhou 450001, China

[2] Research Institute of Surveying and Mapping, Xi'an 710000, China

in the decimetre level. In comparison, carrier phase measurement accuracy can reach millimetres, but it needs precise corrections of systematic errors for ambiguity fixing. SLR observations collected from the International Laser Ranging Service (ILRS) (Pearlman et al. 2002) network are with the accuracy of centimetre level (Degnan 1993). However, it is difficult to further upgrade the accuracy and extend the distance of measurements substantially due to the restrictive repetition frequency of laser pulses. Compared with these three traditional measurement approaches, distance measurement by quantum correlation technology, referred to as quantum ranging, is based on the characteristics of entangled photons, and can surpass the energy and bandwidth limitation of classical measurement. After quantum ranging using entangled photons was first introduced as a newly developed ranging technique, Valencia et al. (2004) proposed a scheme of ranging with a high accuracy using the second-order correlation between a pair of entangled photons. Based on Valencia's work, Xiao et al. (2013) performed a proof-of-principle experiment. The experimental results showed a high precision of 200 µm at a distance of 1043.3 m. Quantum ranging is performed by measuring the second-order correlation of entangled photon pairs (Glauber 1997). The accuracy of quantum ranging is determined by the full-width at half-maximum of the second-order correlation, which depends on the resolution of the coincidence measurement system (Shen et al. 2015). The highest precision for quantum ranging will reach the physical measurement limitation defined by the Heisenberg uncertainty principle (Giovannetti et al. 2001).

Based on different treatments of the observation information and dynamic information, methods for satellite orbit determination are divided into the kinematic (Švehla and Rothacher 2003) and the dynamic (Schutz et al. 2013). With the vogue of research on orbit determination for low Earth orbit (LEO) satellites (XianPin 2009), reduced-dynamic orbit determination (Wu et al. 1991; Montenbruck et al. 2005) was established for solving dynamic orbit determination over-reliance on dynamic force models and problems with kinematic orbit determination. Due to the unavailable conditions for quantum ranging experiments between stations and satellites at present, the paper has introduced simulated quantum ranging to the preliminary application of satellite orbit determination. Some factors affecting the accuracy of QOD are discussed by simulation. Orbit determination utilizing pseudo range, carrier phase or satellite laser ranging by dynamic method is referred to as traditional orbit determination (TOD), and that using quantum technology is referred to as QOD. By the comparison of QOD and TOD, it is really expectant for QOD to exert itself in the future.

Principle of second-order correlation ranging with quantum entanglement

A crucial issue for quantum ranging is quantum entanglement, which is a phenomenon wherein the quantum properties of two particles become codependent, with the properties of one being instantaneously affected by measurements conducted on the other. On the separation of the pair of particles remaining in a state of quantum uncertainty, even by a huge distance, and on measuring one particle's spin the other particle's spin will automatically resolve itself in the other direction. This effect occurs instantaneously, apparently breaching the velocity of light and the rules of relativity (Wong 2014). Owing to the limitation of technology today, the realization of quantum ranging between stations and satellites needs photons as the targets for measurement, and requires laser pulses as the transmission medium. And different to the principle of SLR (Combrinck 2010), quantum ranging is based on the principle of second-order correlation ranging with quantum entanglement. Hong et al. (1987) proposed the measurement accuracy for arriving at time differences at the femtosecond level and for path differences at the micron level by measuring entangled photons transmitting along two different paths according to the theory of second-order quantum interference (see Fig. 1).

To begin with, the laser generates a continuous beam of pumping light. Then the beam is split into two beams of perpendicular polarization entangled light by B. Next, the beams transmit, respectively, along two different paths of L_1 and L_2 to the single photon detectors $E1$ which can transform the beams from optical signals into electrical signals, and the adjustable optical retarder D is emplaced at a point of L_2. The time of the signals arriving at $E1$ will be recorded and used for generating two binary discrete

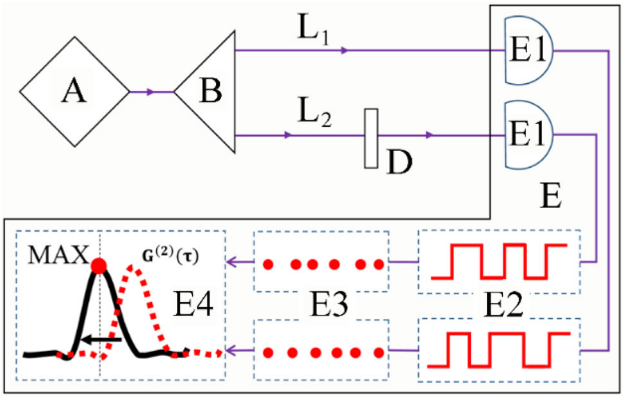


Fig. 1 Diagrammatic view for the principle of second-order correlation ranging with quantum entanglement. Where A is a laser, B is a beam splitter, D is an adjustable optical retarder, and E is a high-speed acquisition part for entangled photons

sequences by coincidence counting. Finally, the sequences are transformed by positive Fourier with convolution operation in the frequency domain and superposition operation in the time domain. Different samples of the second-order correlation distribution function $G^{(2)}(t_1 - t_2)$ corresponding to entangled photons will be obtained with different time delays by adjusting D. The second-order correlation distribution function is expressed by the following formula

$$G^2(t_1 - t_2) = \int \phi(\omega_0/2 + \omega, \omega_0/2 - \omega) \mathrm{e}^{-i\omega(t_1 - t_2)} \mathrm{d}\omega, \tag{1}$$

where the described function $G^{(2)}(t_1 - t_2)$ is a Fourier transform of weight function $\phi(\omega_1, \omega_2)$ with respect to the photon frequency ω. The quantity ω_0 is the frequency of light from the laser A. The non-periodic curve of the function $G^{(2)}(t_1 - t_2)$ will be formed by fitting the samples. The maximum will appear at the only peak of the curve when the transmission time of beams t_1 is equal to t_2.

The geometric thickness of the well calibrated D is d, and n is the effective refractive index. The delay Δt_D of the beam passing through D can be altered by adjusting d and n. The peak value of the second-order correlation distribution function $G^{(2)}(t_1 - t_2)$ will appear until t_1 is equal to t_2. At this time, the value Δt_D and the path difference ΔL between L_1 and L_2 are, respectively, denoted as

$$\Delta t_D = (n - 1)d/c \tag{2}$$

$$\Delta L = L_1 - L_2 = c \cdot \Delta t_D = (n - 1)d, \tag{3}$$

where c is the speed of light in vacuum.

Ranging data processing

As shown in Fig. 2, the laser, beam splitter, adjustable optical retarder and high-speed acquisition and processing part constitute the device of E, which is the ranging measurement based on second-order correlation ranging with quantum entanglement. The position vectors of the beam splitter and the single photon detector are $\vec{r}_0$ and $\vec{r}_1$. The known position vectors of the baseline stations $R1$ and $R2$ are $\vec{\boldsymbol{R}}_1$ and $\vec{\boldsymbol{R}}_2$. Vector $\vec{r}_\mathrm{s}$ is the position of the satellite.

The process of measurement begins with two light beams of entanglement generated by E and then transmitted to $R1$ and $R2$. The satellite reflects the two beams of laser pulses back to $R1$ and $R2$, and then, $L1$ and $L2$ are, respectively, received by the single photon detectors of E for further processing. Last, we can get the difference of two transmission path delay Δt_D by continuously adjusting the optical retarder D when the peak value of the distribution function appears.

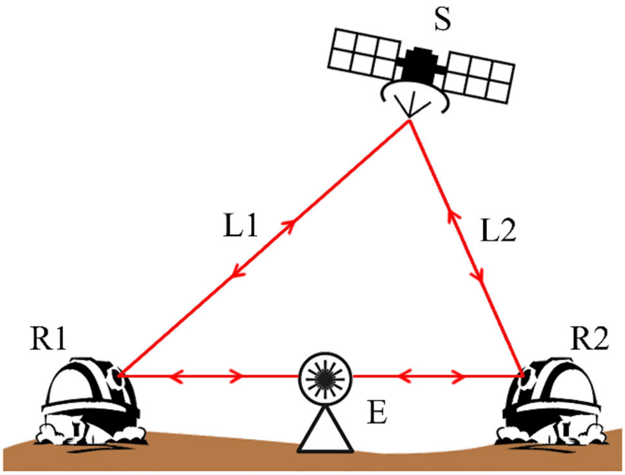


Fig. 2 Schematic diagram of satellite orbit determination based on quantum ranging. Where E is the device of ranging measurement based on second-order correlation ranging with quantum entanglement. Stations *R1* and *R2* establish the baseline. S is the satellite

The paths L_1 and L_2 can be described as

$$L_1 = 2\left|\vec{\boldsymbol{r}}_s - \vec{\boldsymbol{R}}_1\right| + \left|\vec{\boldsymbol{R}}_1 - \vec{\boldsymbol{r}}_0\right| + \left|\vec{\boldsymbol{R}}_1 - \vec{\boldsymbol{r}}_1\right| \tag{4}$$

$$L_2 = 2\left|\vec{\boldsymbol{r}}_s - \vec{\boldsymbol{R}}_2\right| + \left|\vec{\boldsymbol{R}}_2 - \vec{\boldsymbol{r}}_0\right| + \left|\vec{\boldsymbol{R}}_2 - \vec{\boldsymbol{r}}_1\right|. \tag{5}$$

Two parts of systematic errors involved for L_1 and L_2 are hardware errors from measurement terminals and systematic errors generated in the process of laser pulses transmitting between stations and satellites. The hardware errors for L_1 and L_2 are marked as $\Delta_{B,1s}$ and $\Delta_{B,2s}$. Errors in the transmission process are set as $\Delta\rho_{\mathrm{TR},1s}$, $\Delta\rho_{\mathrm{TR},2s}$, $\Delta\rho_{\mathrm{RL,1s}}$ and $\Delta\rho_{\mathrm{RL,2s}}$. The subscript TR denotes atmospheric refraction, and the RL denotes general relativity. The paths between L_1 and L_2 are differentiated by the subscripts of $1S$ and $2S$. So the expressions can be further available as

$$L_1 = 2\left|\vec{\boldsymbol{r}}_s - \vec{\boldsymbol{R}}_1\right| + \left|\vec{\boldsymbol{R}}_1 - \vec{\boldsymbol{r}}_0\right| + \left|\vec{\boldsymbol{R}}_1 - \vec{\boldsymbol{r}}_1\right| + \Delta_{B,1s} + 2\Delta\rho_{\mathrm{TR},1s} + 2\Delta\rho_{\mathrm{RL,1s}} \tag{6}$$

$$L_2 = 2\left|\vec{\boldsymbol{r}}_s - \vec{\boldsymbol{R}}_2\right| + \left|\vec{\boldsymbol{R}}_2 - \vec{\boldsymbol{r}}_0\right| + \left|\vec{\boldsymbol{R}}_2 - \vec{\boldsymbol{r}}_1\right| + \Delta_{B,2s} + 2\Delta\rho_{\mathrm{TR},2s} + 2\Delta\rho_{\mathrm{RL,2s}}. \tag{7}$$

Supposed that the E is emplaced at the mid-point of the baseline composed of $R1$ and $R2$. The key is to make the laser and single photon detectors at the mid-point so as to simplify the calculation. In this case, there is

$$\left|\vec{\boldsymbol{R}}_1 - \vec{\boldsymbol{r}}_0\right| + \left|\vec{\boldsymbol{R}}_1 - \vec{\boldsymbol{r}}_1\right| = \left|\vec{\boldsymbol{R}}_2 - \vec{\boldsymbol{r}}_0\right| + \left|\vec{\boldsymbol{R}}_2 - \vec{\boldsymbol{r}}_1\right| \tag{8}$$

Taking Eqs. (6), (7), (8) and (3) together, the difference between the distances of the satellite to $R1$ and $R2$ is given as

$$\left|\vec{\boldsymbol{r}}_s-\vec{\boldsymbol{R}}_1\right|-\left|\vec{\boldsymbol{r}}_s-\vec{\boldsymbol{R}}_2\right|=(n-1)d/2+\left(\Delta_{B,2s}-\Delta_{B,1s}\right)/2 + \left(\Delta\rho_{\mathrm{TR},2s}-\Delta\rho_{\mathrm{TR},1s}\right) + \left(\Delta\rho_{\mathrm{RL},2s}-\Delta\rho_{\mathrm{RL},1s}\right), \tag{9}$$

where the known observation ρ is the distance difference, then

$$\rho=(L_1-L_2)/2=(n-1)d/2 \tag{10}$$

$$\rho=\left|\vec{\boldsymbol{r}}_s-\vec{\boldsymbol{R}}_1\right|-\left|\vec{\boldsymbol{r}}_s-\vec{\boldsymbol{R}}_2\right|+\Delta_{B,1s}/2+\Delta\rho_{\mathrm{TR},1s} + \Delta\rho_{\mathrm{RL},1s}-\Delta_{B,2s}/2-\Delta\rho_{\mathrm{TR},2s}-\Delta\rho_{\mathrm{RL},2s}. \tag{11}$$

Giving values to the vectors as: $\vec{\boldsymbol{r}}_s=(x_s,y_s,z_s)^{\mathrm{T}}$, $\vec{\boldsymbol{R}}_1=(x_1,y_1,z_1)^{\mathrm{TR}}$ and $\vec{\boldsymbol{R}}_2=(x_2,y_2,z_2)^{\mathrm{T}}$. The approximations are: $\vec{\boldsymbol{r}}_s^{\,*}=\left(x_s^*,y_s^*,z_s^*\right)^{\mathrm{T}}$, $\vec{\boldsymbol{R}}_1^{\,*}=\left(x_1^*,y_1^*,z_1^*\right)^{\mathrm{T}}$ and $\vec{\boldsymbol{R}}_2^{\,*}=\left(x_2^*,y_2^*,z_2^*\right)^{\mathrm{T}}$ separate, respectively.

Due to the relationship in Eq. (11), we can get a linearised formula by Taylor series as

$$\begin{aligned} V&=\boldsymbol{A}_{1s}\cdot\begin{bmatrix}\delta x_s\\ \delta y_s\\ \delta z_s\end{bmatrix}-\boldsymbol{A}_{2s}\cdot\begin{bmatrix}\delta x_s\\ \delta y_s\\ \delta z_s\end{bmatrix}-l\\ &=(\boldsymbol{A}_{1s}\text{—}\boldsymbol{A}_{2s})\begin{bmatrix}\delta x_s\\ \delta y_s\\ \delta z_s\end{bmatrix}-l=(\boldsymbol{A}_{1s}\text{—}\boldsymbol{A}_{2s})\boldsymbol{\delta r}_s-l.\end{aligned} \tag{12}$$

In the formula, $\boldsymbol{A}_{1s}=\begin{pmatrix}\frac{\Delta x_{1s}}{\rho_{1s}^*} & \frac{\Delta y_{1s}}{\rho_{1s}^*} & \frac{\Delta z_{1s}}{\rho_{1s}^*}\end{pmatrix}$, $\boldsymbol{A}_{2s}=\begin{pmatrix}\frac{\Delta x_{2s}}{\rho_{2s}^*} & \frac{\Delta y_{2s}}{\rho_{2s}^*} & \frac{\Delta z_{2s}}{\rho_{2s}^*}\end{pmatrix}$,$\boldsymbol{\delta r}_s=\begin{pmatrix}\delta x_s & \delta y_s & \delta z_s\end{pmatrix}^{\mathrm{T}}$, where $\Delta x,\Delta y$ and Δz are the differences between the satellite and the station corresponding to the three dimensional coordinate components themselves, and δx_s, δy_s, δz_s are the corrections to the satellite position. The values of $\rho_{1s}^*$and ρ_{2s}^* are the distance calculated by the approximate coordinates of the satellite and the station. The variable l is derived as

$$l=\rho-\left(\rho_{1s}^*-\rho_{2s}^*\right)=(n-1)d/2-\left(\rho_{1s}^*-\rho_{2s}^*\right). \tag{13}$$

The positional corrections for a satellite in given epoch can be presented with the state transition matrix $\boldsymbol{\Psi}(t,t_0)$ as

$$\boldsymbol{\delta r}_s=\boldsymbol{\Psi}(t,t_0)\cdot\boldsymbol{\delta X}, \tag{14}$$

where $\boldsymbol{\delta X}$ is the correction vector for the values of initial position, velocity and solar radiation pressure parameters. The state transition matrix $\boldsymbol{\Psi}(t,t_0)$ can be obtained by integrating over variational equation, which comes by taking the partial of satellite orbital motion equation with respect to the initial state vector according to the principle of dynamic orbit determination.

Taking Eqs. (12) and (14) into consideration, observation equations with parameters at the initial orbit to be estimated can be

$$V=\left(\boldsymbol{A}_{1s}\boldsymbol{\Psi}(t,t_0)\quad\text{—}\boldsymbol{A}_{2s}\boldsymbol{\Psi}(t,t_0)\right)\cdot\boldsymbol{\delta X}-l. \tag{15}$$

Here, satellite orbits can be resolved according to the least square adjustment. Because in the process of quantum ranging, observations are directly calculated by the parameters geometric thickness d and effective refractive index n of the optical retarder. And the clock information is not involved with the observation equation. Consequently, clock errors are not resolved in the solution process.

Simulation and precision analysis

It is considered that the research achievements from TOD for systematic errors with satellite terminals are also applicable to QOD. Therefore, systematic error models of satellite terminals for QOD are directly modified by the TOD models relevantly. As early as 1987, the L. Mandel research group obtained the accuracy of 16 μm for second-order correlation measurements with entangled photons using the Hong–Ou–Mandel (HOM) interferometer (Hong et al. 1987). But there is currently no available equipment for quantum ranging between a station and a satellite. So it is assumed that the hardware errors $\Delta_{B,1s}$ and $\Delta_{B,1s}$ are disposed as corrected, and then random errors of micron scale are added to simulated observations. Errors from atmospheric refraction and general relativity in the transmission process are to be modified with models of Mendes et al. (2002) and Iorio (2001).

The arc-length of observations simulated is 1 day. First, the position information of the satellite corresponding to the time needed is obtained by fitting the precise ephemerides from Helmholtz-center Potsdam-German Research Centre for Geosciences (GFZ) (Uhlemann et al. 2015). Then, a distance is calculated by the station's known position after coordinate transformation with the position of the satellite. The initial observations simulated will be obtained after the difference operation for distances with errors from simulating atmospheric refraction and general relativity in pairs. Eventually, the final observations are generated by initial observations mixed into random errors within positive or negative 5 μm.

The experiment takes observation data from 8 tracking stations of the International GNSS Monitoring and Assessment System (IGMAS) (Jiao et al. 2011) and 4 stations of the Multi-GNSS Experiment (MGEX) (Montenbruck et al. 2013) on June 10, 2016. Moreover, baselines for quantum ranging need to be established. But if any two stations are allowed to be paired, the number of baselines will be 66. Hence, four rules are set up to control the number of baselines. The first rule is that only two stations with distance within 2500 km can be combined spontaneously. The second rule is that a station in a baseline with the distance over 2500 km can only be attached to one baseline. It is not allowed to appear in other baselines. The third rule is that a short baseline has the priority to be established. The last rule is that the station of FTNA can take part in the establishment of two baselines to avoid a station being left alone. Thus, the number of baselines is 14 in total (see Fig. 3).

The tests of orbit solutions are the four groups listed in Table 1. Test-1 of TOD is resolved by measured data. The other three tests are of QOD by simulation. Among the rest, the solution of Test-2 is resolved with both atmospheric refraction and general relativity corrected. The solution of Test-3 is resolved without modification for atmospheric refraction. The solution of Test-4 is resolved without general relativity corrected.

On account of the constellation of BDS with Geostationary Orbit (GEO), Inclined Geosynchronous Orbit (IGSO) and Medium Orbit (MEO), satellites of C01 (GEO), C06 (IGSO) and C11 (MEO) are selected to be tested and analyzed. The results resolved are compared with the precise ephemerides. Then, values of root mean square (RMS) are counted, and statistical analyses are done with RMS values in along-track, cross-track, radial directions and three-dimensions (3D).

Precision evaluation of QOD

By comparison, the solution of QOD with the precise ephemerides, the values of 3D RMS for C01 is at the decimetre level, and the values for both C06 and C11 are at the centimetre level. The averaged 3D RMS for the solution of Test-1 with Test-2 are listed in Table 2, where shows that the accuracy of QOD for C01 (GEO), C06 (IGSO) and C11 (MEO) surpass the TOD about 20, 30 and 10 times, respectively.

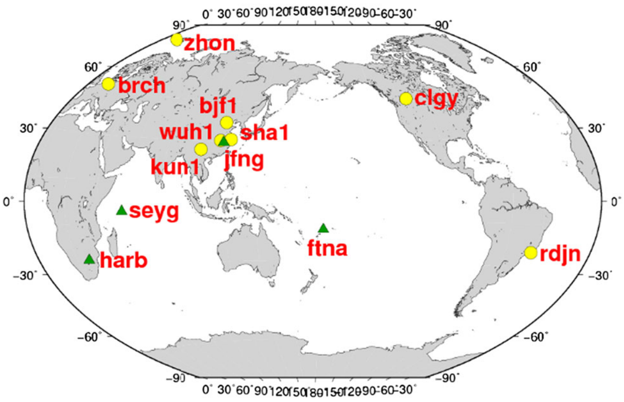


Fig. 3 Distribution of selected IGMAS (yellow circle) and MGEX (green triangle) stations used

As for the poor changes of geometry structure for GEO satellites with respect to stations and the imperfect solar radiation pressure model, the accuracy of C01 becomes the worst. Limited by the short observing arc-length, and the quantity and distribution of stations, the TOD accuracy of C06 and C11 barely reach the decimetre level. The results of QOD are obviously better than the TOD. The advantage which can be clearly picked out from QOD to TOD is the high accuracy of observations. In other words, the effect of improving the accuracy for observations may be equivalent to extending the arc-length. It is also speculated that the sensitivity and correlation among the observed data will be weakened with the accuracy improved. On the other hand, because some correction models, like the solar radiation pressure model, are not refined enough, the case of QOD accuracy increasing without limitation does not take place.

Atmospheric refraction with QOD

In the process of quantum ranging, the paths of laser pulses transmitted between stations and satellites are not straight lines but are affected by the atmospheric refraction. Based on the atmospheric refraction model proposed by Mendes et al. (2002), station BJF1 and satellite C11 are taken as examples for discussion and analysis. The atmospheric refraction corrections at each epoch of the day between BJF1 and satellites are collected,and values are in the range of 2.450–22.216 m. As shown in Fig. 4, because the relative positions between stations and GEO satellites change little as time elapses, the corrections change smoothly compared with the IGSO and MEO, and the mean value is 4.440 m. Conversely, for the IGSO and MEO satellites, corresponding to the great and continuous positional changes between stations and satellites, the corrections fluctuate substantially. But the averaged correction values for GEO, IGSO and MEO satellites are approximate. The mean values for IGSO and MEO satellites are 4.123 and 4.513 m, respectively.

After the collection of atmospheric refraction corrections at each epoch of the day between the C11 and stations,the values are in the range of 2.454–25.665 m, and with the population mean of 5.224 m. The difference between averaged corrections for each station and the population mean is within 0.5 m. As illustrated in Fig. 5, corrections for different stations in some epochs are approximate. For the further study of atmospheric refraction varying between C11 and other stations, the corrections from epoch 1915–2038 are collected corresponding to stations of WUH1, BJF1, KUN1, SHA1, ZHON, JFNG and FTNA which can observe satellite C11 at the same time. Taking the corrections of WUH1 as the standard,

Table 1 Groups and models for orbit determination

Group	TEST-1	TEST-2	TEST-3	TEST-4
Orbit determination mode	TOD	QOD		
Observations	Pseudo range, carrier phase	Simulated observations of quantum ranging		
Atmospheric refraction		Correction applied (Mendes et al. 2002)	No correcting	Correction applied (Mendes et al. 2002)
General relativity	Correction applied (McCarthy and Petit 2003)	Correction applied (Iorio 2001)		No correcting
Sampling rate	30 s			
Tropospheric delay	Saastamoinen (Boehm et al. 2006)			
Geopotential	EGM96 model (12 × 12) (Lemoine et al. 1997)			
M-body gravity	Sun, Moon and all planets (JPL DE405)			
Tide	Solid Earth tide, pole tide, ocean tide; IERS Conventions 2003 (McCarthy and Petit 2003)			
Solar radiation pressure	Reduced BERN five parameters with no initial value (Springer et al. 1999)			
Estimator	Standard least square			

Table 2 Comparison of averaged 3D RMS for Test-1 and Test-2

	Averaged 3D RMS [mm]		
	C01	C06	C11
Test-1	2340	900	951
Test-2	113	30	75

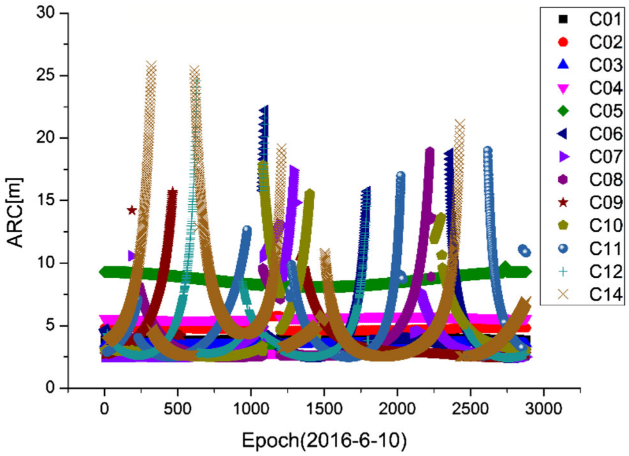


Fig. 4 Atmospheric refraction corrections (ARC) between BJF1 and satellites

deviations are obtained by making differences between the standard and other corrections of stations. Then the absolute values of deviations are drawn in Fig. 6 corresponding to the epochs.

The sequence of stations sorted by the distance from WUH1 to the other six stations is JFNG, SHA1, KUN1, BJF1 and ZHON. As shown in Fig. 6, the deviations between corrections of WUH1 and the other six stations have a certain relationship with the distance of the two stations. The closer the distance between two stations, the smaller and smoother difference of corrections will be. When the distance is far, the difference will be uncertain. But to what extent the difference weakened is related to the distance, elevation and the weather of the stations around, etc (see Fig. 7, Table 3).

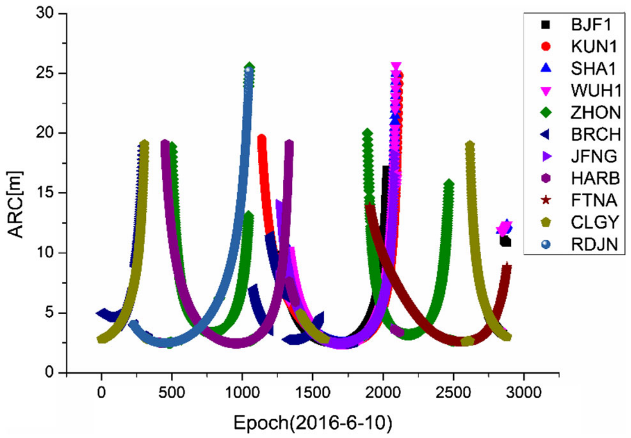


Fig. 5 Atmospheric refraction corrections (ARC) between C11 and stations

Comparing the results of Test-3 with Test-2, the influence of atmospheric refraction on QOD is clear. The results with atmospheric refraction corrected have improved greatly in the along-track, cross-track and radial directions. The averaged 3D RMS for C01, C06 and C11 from Test-3 increased to 4106, 2233 and 2051 mm, and the orbit accuracy decreases by 36, 74 and 27 times, respectively, in

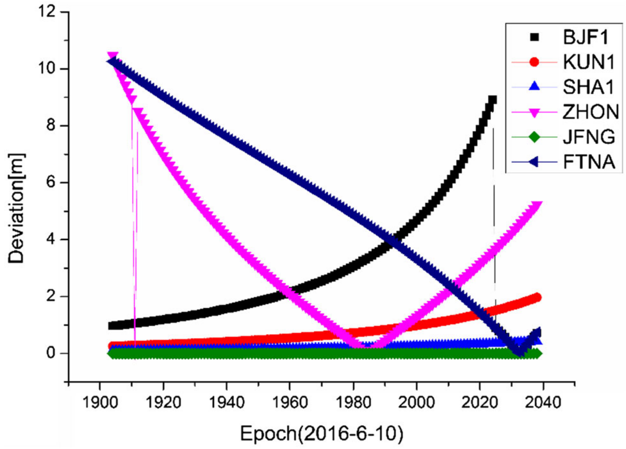


Fig. 6 Absolute Values of deviations by making difference between atmospheric refraction corrections of WUH1 and corrections of the other stations from epoch 1915–2038 of the day

contrast to the satellites of Test-2. Therefore, the corrections for atmospheric refraction are indispensable to QOD in most cases.

General relativity with QOD

The gravitational field will bend the transmission path of laser pulses according to the general theory of relativity. Therefore, it becomes essential to discuss and analyze the influence of general relativity on QOD. Referring to the general relativity model presented by Iorio (2001), station BJF1 and satellite C11 are taken as examples for discussion and analysis. The general relativity corrections at each epoch of the day between the BJF1 and satellites are collected.

As shown in Fig. 8, the general relativity corrections of GEO satellites fluctuate a little, while the corrections for IGSO and MEO satellites fluctuate greatly with changes in the relative position. The values of corrections are within ± 4 m. The averaged value of corrections for IGSO satellites is the largest and about 1.629 m. Then, followed by MEO satellites, the mean is 0.878 m. The averaged value for GEO satellites is the smallest, which is 0.176 m.

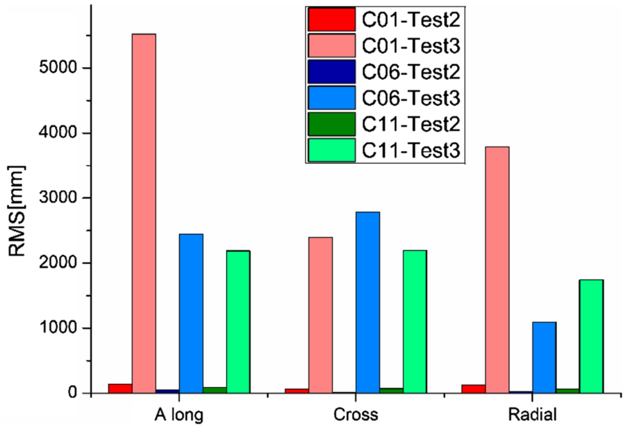


Fig. 7 RMS of orbits compared with Test-2 and Test-3 in along-track, cross-track, and radial directions

Table 3 Comparison of averaged 3D RMS statistics for Test-2 and Test-3

	3D mean RMS [mm]		
	C01	C06	C11
Test-2	113	30	75
Test-3	4106	2233	2051

General relativity corrections at each epoch of the day between satellite C11 and the ground stations are collected and drawn in Fig. 9. What is shown is that the corrections of different stations coincide on the same curve. For further analysis, the mean corrections from different stations corresponding to the epoch is calculated. Then, deviations are obtained by taking differences between the mean and the corrections of stations at each epoch.

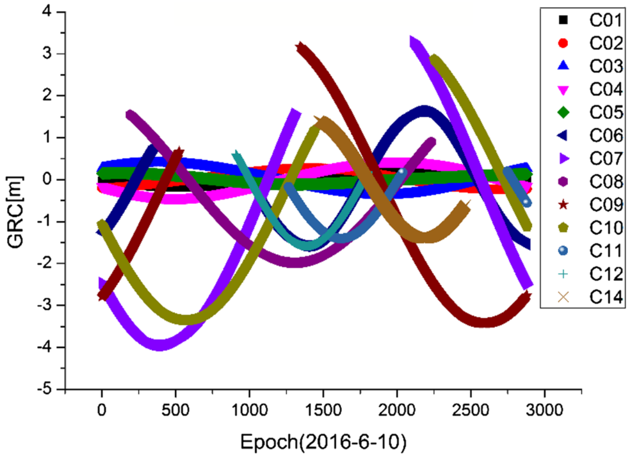


Fig. 8 General relativity corrections (GRC) between BJF1 and satellites

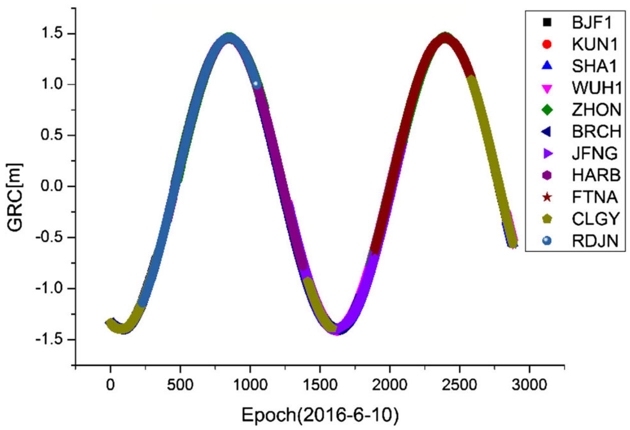


Fig. 9 General relativity corrections (GRC) between C11 and stations

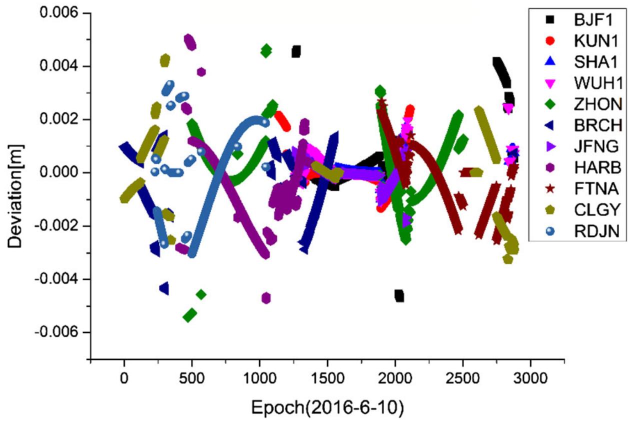


Fig. 10 Values of deviations by making difference between the mean of general relativity corrections of an epoch from different stations and corrections of the other stations corresponding to the epoch

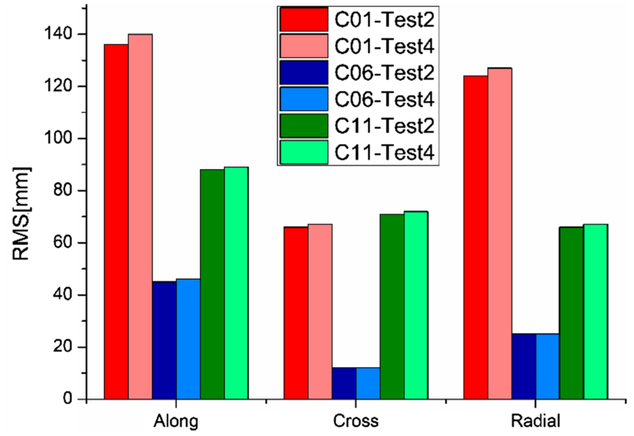


Fig. 11 RMS of orbits compared with Test-2 and Test-4 in along-track, cross-track, and radial directions

Assuming the mean of corrections at each epoch as the standard to be referred, the deviations are within 0.006 m. Thus, it is available to suppose that the difference of corrections between a satellite and stations at the same epoch is at the micron scale. Furthermore, it is equivalent to say that the influence of general relativity on observations of quantum ranging is about millimeter level (see Figs. 10, 11).

Comparing the results of Test-4 with Test-2, the differences of RMS in the along-track, cross-track and radial directions are not obvious. The possible reason proposed is that the influence of systematic errors with satellite terminals on orbit determination goes far beyond the general relativity to QOD. Therefore, the influence of general relativity on QOD can be ignored in some cases until effectual improvements for the systematic errors appear.

Conclusions

On the basis of presenting the principle of quantum ranging, observations of quantum ranging are introduced to satellite orbit determination. And factors affecting QOD are preliminarily discussed and analyzed by comparisons of tests with QOD and TOD. In contrast with TOD, QOD has the advantages over higher orbit precision, easier models for correction and faster speed in the convergence. Meanwhile, it is also limited to the current technical approaches, and the imperfection of high directivity from laser pulse is inherited by quantum ranging. Besides, systematic errors with satellite terminals will become the biggest obstacle for improving the precision of QOD.

Acknowledgments This study was supported by National Natural Science Foundation of China (Grant No. 41674042). Professor Jinghua Qu, Doctor Hanbing Peng and Guofeng Ji gave some valuable proposals. All are greatly acknowledged.

Compliance with ethical standards

Competing interests The authors have declared that no competing interests exist.

References

Boehm J, Niell A et al (2006) Global Mapping Function (GMF): a new empirical mapping function based on numerical weather model data. Geophys Res Lett 33(7):1–4

Combrinck L (2010) Satellite laser ranging. In sciences of Geodesy-I. Springer, Berlin, pp 301–338

D'Ariano GM, Lo PP et al (2001) Using entanglement improves the precision of quantum measurements. Phys Rev Lett 87(27 Pt 1):270404

Degnan JJ (1993) Millimeter accuracy satellite laser ranging: a review, American Geophysical Union

Giovannetti V, Lloyd S et al (2001) Quantum-enhanced positioning and clock synchronization. Nature 412(6845):417

Giovannetti V, Lloyd S et al (2002) Positioning and clock synchronization through entanglement. Phys Rev A 65(2):130–132

Glauber RJ (1997) The quantum theory of optical coherence. Phys Rev 130(6):2529–2539

Hong CK, Ou ZY et al (1987) Measurement of subpicosecond time intervals between two photons by interference. Phys Rev Lett 59(18):2044

Iorio L (2001) Satellite laser ranging and general relativity. Gen Relativ Gravit 43(12):3243–3245

Jiao W, Ding Q et al (2011) Monitoring and assessment of GNSS open services. Sci Sin 64(S1):S19–S29

Lemoine FG, Smith DE et al (1997) The development of the NASA GSFC and NIMA joint geopotential model. Springer, Berlin

McCarthy DD and Petit G (2003) IERS Technical Note No. 32. IERS Conventions: 33-56

Mendes VB, Prates G et al (2002) Improved mapping functions for atmospheric refraction correction in SLR. Geophys Res Lett 29(10):51–53

Montenbruck O, van Helleputte T et al (2005) Reduced dynamic orbit determination using GPS code and carrier measurements. Aerosp Sci Technol 9(3):261–271

Montenbruck O, Steigenberger P et al (2013) IGS-MGEX: preparing the ground for multi-constellation GNSS science. Espace 9(1):42–49

Ozawa M (2001) Position measuring interactions and the Heisenberg uncertainty principle. Phys Lett A 299(1):1–7

Pearlman MR, Degnan JJ et al (2002) The international laser ranging service. Adv Space Res 30(2):135–143

Schutz BE, Tapley BD et al (2013) Dynamic orbit determination using GPS measurements from TOPEX/POSEIDON. Geophys Res Lett 21(19):2179–2182

Shen Y, Xu L et al (2015) Relative orbit determination for satellite formation flying based on quantum ranging. Adv Space Res 56(4):680–692

Springer TA, Beutler G et al (1999) A new solar radiation pressure model for GPS satellites. GPS Solut 2(3):50–62

Uhlemann M, Gendt G et al (2015) GFZ Global Multi-GNSS network and data processing results. Springer, Berlin

Valencia A, Scarcelli G et al (2004) Distant clock synchronization using entangled photon pairs. Appl Phys Lett 85(13):2655–2657

Švehla D, Rothacher M (2003) Kinematic and reduced-dynamic precise orbit determination of low earth orbiters. Adv Geosci 1(1):47–56

Wong B (2014) Wuantum entanglement. Quantum Inform 81(2):865–942

Wu SC, Yunck TP et al (1991) Reduced-dynamic technique for precise orbit determination of low earth satellites. J Guid Control Dyn 14(1):2143–2153

XianPin Q (2009) Research on Precision Orbit Determination Theory and Method of Low Earth Orbiter Based on GPS Technique, Zhengzhou Institute of Surveying and Mapping. D

Xiao JJ, Fang C et al (2013) Distance ranging based on quantum entanglement. Chin Phys Lett 30(10):100