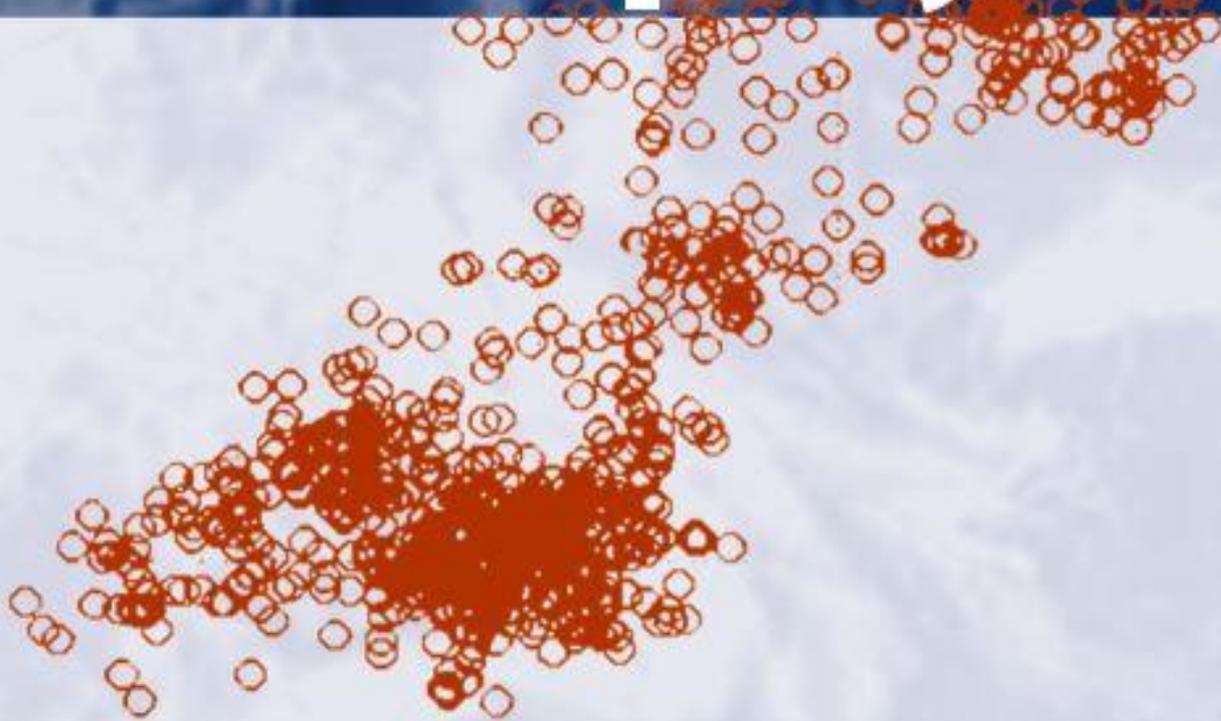


Volume 68 • Number 1 • February 2020

Acta Geophysica



PAN
POLISH ACADEMY OF SCIENCES



Institute of Geophysics
Polish Academy of Sciences



Springer



Seismic time–frequency spectrum analysis based on local polynomial Fourier transform

Pengfei Qi¹ · Yanchun Wang¹

Received: 12 July 2019 / Accepted: 4 October 2019 / Published online: 13 November 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

Time–frequency analysis technology is widely used in non-stationary seismic data analysis. The energy concentration of the spectrum depends on the consistency of the kernel function of the time–frequency analysis method and the instantaneous frequency variation of the signals. The conventional time–frequency analysis methods usually require that the local instantaneous frequency of the signals remains unchanged or linearly changed. So it is difficult to accurately characterize the instantaneous frequency nonlinear variation of the non-stationary signal. The local polynomial Fourier transform (LPFT) method can effectively describe the instantaneous frequency variation by local high-order polynomial fitting and obtain the results with high spectral and energy concentration. The numerical simulations and field seismic data applications show that the time–frequency spectrum results obtained by LPFT can reflect the instantaneous frequency variation characteristics of the seismic data, while ensuring the concentration of time–frequency energy.

Keywords Non-stationary seismic data · Time–frequency analysis · Local polynomial Fourier transform · Energy concentration

Introduction

Numerous studies have indicated that seismic waves vary continuously with increasing propagation distance due to the influence of geometric diffusion, absorption attenuation and fluid, resulting in the non-stationary seismic data (e.g., Rene et al. 1986; Wang 2006; Wang et al. 2012; Yuan et al. 2017). Time–frequency analysis technique is an effective approach of non-stationary signals characterization. By transforming the one-dimensional signal in the time domain into the time–frequency domain, the relationship between the frequency components over time can be characterized (Puryear et al. 2012; Liu et al. 2016; Yuan et al. 2019a).

Traditional time–frequency analysis methods include short-time Fourier transform (STFT) (Durak and Arıkan 2003), continuous wavelet transform (CWT) (Sinha et al. 2009), S transform (ST) (Wu and Castagna 2017), matching pursuit (MP) (Mallat and Zhang 1993), etc. STFT realizes the characterization of the time–frequency

relationship of non-stationary signals by calculating the Fourier transform of the truncated signal in the time window. However, its time window function is fixed, and the time–frequency resolution is the same regardless of the low-frequency or the high-frequency components. So STFT belongs to the single-resolution analysis method (Zhong and Huang 2010). To conquer the shortcomings of single-resolution STFT, CWT and ST make use of variable time windows instead of fixed ones. The time window is adjusted automatically by the methods. Wider time windows are used for the low-frequency components, while narrower time windows are used for the high-frequency components, thereby realizing multi-resolution analysis of the signal (Phinyomark et al. 2011; Li et al. 2016). As a different approach, the MP method adopts the time–frequency atomic dictionary instead of the time window function. The waveform and width of the atom are defined based on the instantaneous frequency (IF), instantaneous phase and envelope of the local signal. Then the multi-resolution analysis of the signal is realized by Wigner–Ville distribution (Wang 2010). So far, these time–frequency analysis methods have been widely applied in seismic exploration, such as high resolution processing (Smith et al. 2008; Radad et al. 2015), denoising (Parolai 2009;

✉ Yanchun Wang
wangyc@cugb.edu.cn

¹ School of Geophysics and Information Technology, China University of Geosciences, Beijing 100083, China

Quadfeul and Aliouane 2014), structural interpretation (Wang et al. 2019), reservoir prediction (Partyka et al. 1999; Naseer and Asim 2017) and hydrocarbon detection (Sun et al. 2002; Castagna et al. 2003; Yuan et al. 2019b).

In fact, the traditional time–frequency analysis method implies the hypothesis of quasi-stationary in application (Kadambe and Boudreaux-Bartels 1992) that the statistical features of non-stationary signal does not change in the local time domain. Therefore, these methods are essentially zero-order fitting of the time–frequency features. That is, based on the orthogonal rectangular time–frequency grid, the line segments parallel to the time axis or the frequency axis are used to approximate the time–frequency characteristics of the signal in the local time–frequency plane (Yang et al. 2014). Take STFT for example, its basis function is sine or cosine function, the instantaneous frequency in the time window remains unchanged. It may lead to poor concentration of time–frequency energy and difficult to accurately delineate the local time–frequency characteristics of non-stationary signals.

To weaken the prerequisite of quasi-stationary, Mann and Haykin (1995) proposed a linear chirplet transform (LCT) method that allows the time–frequency grid to be tilted. The characterization of time–frequency features of non-stationary signals was improved by using a line segment with a certain slope in the local time–frequency plane to approximate the time–frequency characteristics of signals. However, since the slope parameter of the chirplet in the LCT is fixed, the time–frequency energy spectrum with higher concentration can be obtained only when the slope value is the same as the local time–frequency slope of the signal. In order to adapt to different time–frequency tilt variations, Yu and Zhou (2016) generalized this method and proposed the general linear chirplet transform (GLCT) method. The key of GLCT is to scan the slope values of the chirplet at local time–frequency points to obtain time–frequency energy with different concentrations and to select the slope value corresponding to the highest energy concentration as the time–frequency slope of the current time–frequency point. Therefore, the LCT and GLCT methods are essentially the first-order fitting of the local time–frequency characteristics of the signal. Accurate analysis results can be achieved especially when the local time–frequency characteristics of the non-stationary signal conform to linear variation.

Actually the local time–frequency characteristics of non-stationary signals tend to vary nonlinearly. Therefore, Katkovich (1998) proposed the local polynomial Fourier transform (LPFT) method, which allows the time–frequency grids to generate the local curvature. High-order characterization of local time–frequency variations of signals is achieved by polynomial fitting. LPFT can accurately characterize the IF variation of signals with high time–frequency energy concentration (Li et al. 2011).

In this paper, firstly, the LPFT method is analyzed on the basis of the analytical signal expression of Ville (1948). The calculation process of the coefficients in polynomial demodulation operator is given as well. Then, through the synthetic data, we compare LPFT methods with the traditional time–frequency analysis methods, such as STFT, CWT, ST and MP. Finally, LPFT is applied to the real seismic data, which verifies the validity of the method.

Methods

According to Ville (1948) theory, an analytical signal $s(t)$ can be expressed as

$$s(t) = A(t) \exp \left[i \int f(t) dt \right], \quad (1)$$

where t represents time, $A(t)$ represents instantaneous amplitude, $f(t)$ represents instantaneous frequency (IF), i represents the imaginary unit, and the integral result represents instantaneous phase.

The STFT of the signal $s(t)$ can be expressed as (Cohen 1995)

$$\begin{aligned} S(\tau, \omega) &= \int_{-\infty}^{\infty} w(t - \tau) s(t) \exp(-i\omega t) dt \\ &= \int_{-\infty}^{\infty} w(t - \tau) A(t) \exp(-i\omega t) \exp \left[i \int f(t) dt \right] dt \end{aligned} \quad (2)$$

where $S(\tau, \omega)$ represents the time–frequency spectrum at time τ and frequency ω , and $w(t - \tau)$ represents the time window.

According to the Taylor formula, the IF $f(t)$ can be Taylor expanded in the time window $w(t - \tau)$ as

$$f(t) = f(\tau) + \sum_{k=1}^n \frac{1}{k!} f^{(k)}(\tau) (t - \tau)^k + R_n(t), \quad (3)$$

where $!$ represents the factorial, $f^{(k)}(\tau)$ represents the k th derivative of $f(t)$ at time τ , $R_n(t)$ represents the Lagrangian remainder, and n represents the order of the Taylor expansion. Ignore $R_n(t)$ and substitute Eq. (3) into (2), we have

$$\begin{aligned} S(\tau, \omega) &= \int_{-\infty}^{\infty} w(t - \tau) A(t) \exp(-i\omega t) \exp \\ &\quad \times \left\{ i \left[f(\tau) t + \sum_{k=1}^n \frac{1}{k! \cdot (k+1)} f^{(k)}(\tau) (t - \tau)^{k+1} \right] \right\} dt. \end{aligned} \quad (4)$$

Therefore, the amplitude at the time–frequency point $(\tau, f(\tau))$ can be expressed as

$$\begin{aligned}
 |S(\tau, f(\tau))| &= \left| \int_{-\infty}^{\infty} w(t - \tau) \exp(-if(\tau)t)A(t) \exp \right. \\
 &\quad \times \left. \left\{ i \left[f(\tau)t + \sum_{k=1}^n \frac{1}{k! \cdot (k+1)} f^{(k)}(\tau)(t - \tau)^{k+1} \right] \right\} dt \right| \\
 &= \left| \int_{-\infty}^{\infty} w(t - \tau)A(t) \exp \right. \\
 &\quad \times \left. \left\{ i \left[\sum_{k=1}^n \frac{1}{k! \cdot (k+1)} f^{(k)}(\tau)(t - \tau)^{k+1} \right] \right\} dt \right| \\
 &\leq \left| \int_{-\infty}^{\infty} w(t - \tau)A(t)dt \right| \tag{5}
 \end{aligned}$$

Due to the presence of the modulation component in Eq. (5)

$$\exp \left\{ i \left[\sum_{k=1}^n \frac{1}{k! \cdot (k+1)} f^{(k)}(\tau)(t - \tau)^{k+1} \right] \right\}, \tag{6}$$

the time–frequency amplitude at $(\tau, f(\tau))$ is reduced, which makes the amplitude of STFT unable to focus on the $f(t)$. To achieve that the amplitude maxima are obtained at $(\tau, f(\tau))$ so that the time–frequency spectrum obtains a higher time–frequency energy concentration, it is necessary to eliminate the exponential term in Eq. (6). Therefore, a polynomial demodulation operator is introduced into STFT,

$$\exp \left\{ -i \left[\sum_{k=1}^n \frac{1}{k! \cdot (k+1)} c_k(\tau)(t - \tau)^{k+1} \right] \right\}, \tag{7}$$

where c_k is the coefficient of the polynomial demodulation operator.

Thus, Eq. (2) becomes

$$\begin{aligned}
 S(\tau, \omega) &= \int_{-\infty}^{\infty} w(t - \tau)s(t) \exp(-i\omega t) \\
 &\quad \times \exp \left\{ -i \left[\sum_{k=1}^n \frac{1}{k! \cdot (k+1)} c_k(\tau)(t - \tau)^{k+1} \right] \right\} dt, \tag{8}
 \end{aligned}$$

and

$$\begin{aligned}
 S(\tau, f(\tau)) &= \int_{-\infty}^{\infty} w(t - \tau)s(t) \exp(-if(\tau)t) \exp \\
 &\quad \times \left\{ -i \left[\sum_{k=1}^n \frac{1}{k! \cdot (k+1)} c_k(\tau)(t - \tau)^{k+1} \right] \right\} dt \\
 &= \int_{-\infty}^{\infty} w(t - \tau)A(t) \exp \\
 &\quad \times \left\{ i \left[\sum_{k=1}^n \frac{1}{k! \cdot (k+1)} (c_k(\tau) - f^{(k)}(\tau))(t - \tau)^{k+1} \right] \right\} dt \tag{9}
 \end{aligned}$$

According to Eq. (5), when $c_k(\tau) = f^{(k)}(\tau)$ ($k = 1, 2, \dots, n$), $S(\tau, f(\tau))$ reaches the maximum value and the time–frequency energy is the most concentrated.

Equation (8) is the expression of n th-order LPFT. When $n = 1$ and the c_1 is fixed, the LPFT reduces to LCT. When $n = 0$, the LPFT becomes STFT.

Taking signal

$$s(t) = \sin [2\pi \cdot 55 \cdot t + 2\pi \cdot \sin(25 \cdot t)] \tag{10}$$

as an example, the spectrum of each order LPFT is calculated by using Eq. (8). According to the Ville (1948) formula, the IF of signal (10) can be calculated as

$$f(t) = 55 + 25 \cdot \cos(25 \cdot t). \tag{11}$$

Since Eq. (11) is a continuous derivable function, its derivative expressions can be calculated. For the convenience of expression, only the first to third derivatives are calculated as follows

$$\begin{aligned}
 f^{(1)}(t) &= -25^2 \cdot \sin(25 \cdot t) \\
 f^{(2)}(t) &= -25^3 \cdot \cos(25 \cdot t). \\
 f^{(3)}(t) &= 25^4 \cdot \sin(25 \cdot t) \tag{12}
 \end{aligned}$$

The high-concentration time–frequency energy solution of the signal $s(t)$ can be obtained through substituting $f^{(k)}(\tau)$ ($k = 1, 2, 3$) into Eq. (12) by $c_k(\tau)$ ($k = 1, 2, 3$) in Eq. (8), respectively.

Figure 1 shows the signal $s(t)$ (Eq. 10), the IF (Eq. 11) and the results of zeroth–third-order LPFT, which are represented by LPFT k ($k = 0, 1, 2, 3$). In general, the peak position of the time–spectrum energy of each order LPFT, also known as the time–frequency ridges (Gribonval 2001), all concentrate near the IF. However, the energy variation of LPFT0 is too large, which is greatly deviated from the real energy of the signal. Compared to LPFT0, the consistency of the time–frequency energy of other order LPFTs is greatly improved. In Fig. 2, time–frequency ridges and corresponding amplitudes of LPFTs are extracted for IF detection (Terrien et al. 2008). It can be seen in Fig. 2b, the IF errors of LPFT2 and LPFT3 are smaller than LPFT0 and LPFT1. For time–frequency ridge amplitude (Fig. 2c), the variation range of LPFT0 is too large, while the variation range of other LPFTs is small. LPFT3 takes the smallest one, which is consistent with the true amplitude variation trend of the signal.

In order to analyze the concentration of time–frequency energy near the IF of each order LPFT, the index of spectrum concentration (SC) is defined,

$$\begin{aligned}
 [f_{\min}, f_{\max}] &= \arg_f \left\{ \text{TFA}(f, t_j) = \frac{\sqrt{2}}{2} \max(\text{TFA}(f, t_j)) \right\}, \\
 \text{SC}(t_j) &= \max(\text{TFA}(f, t_j)) / (f_{\max} - f_{\min}) \tag{13}
 \end{aligned}$$

where TFA represents time–frequency amplitude, f_{\min} and f_{\max} are the lower and upper limits of the frequency analysis, respectively. The first expression in Eq. (13) indicates that

Fig. 1 The zeroth–third-order LPFT with known instantaneous frequencies and derivatives. **a** The signal (10), **b** IF, **c** LPFT0, **d** LPFT1, **e** LPFT2 and **f** LPFT3

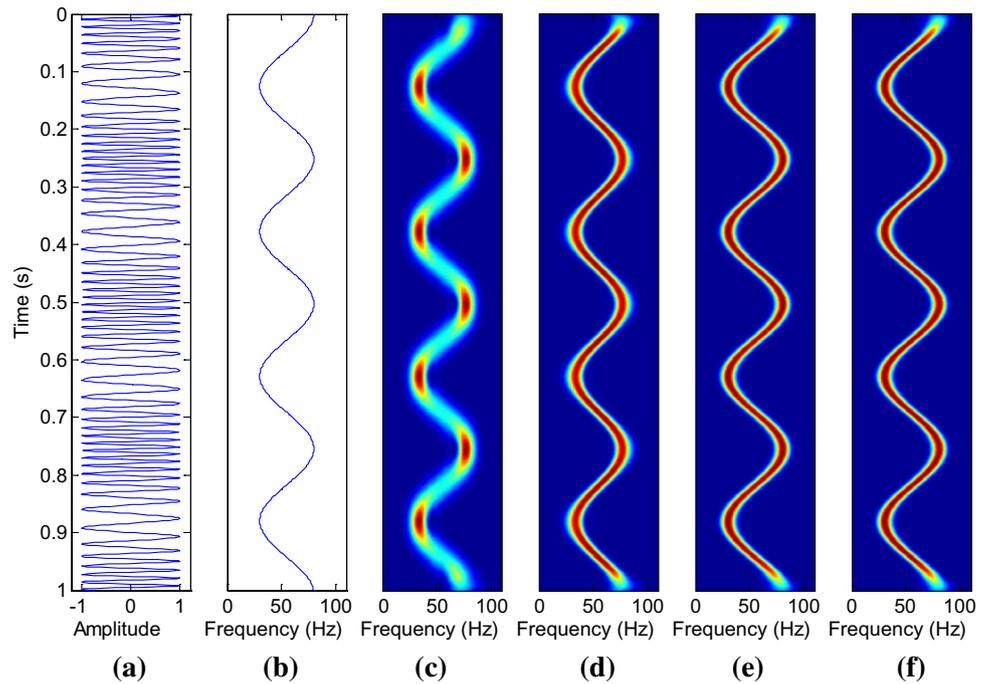
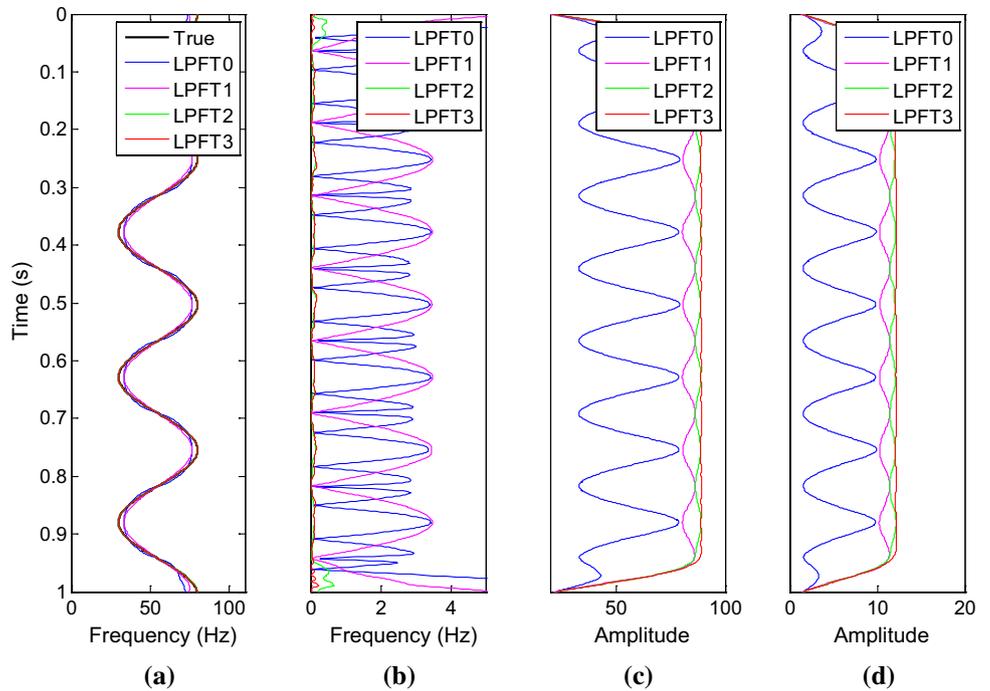


Fig. 2 IF detection and spectral concentration analysis. **a** The true and detection values of IF, **b** IF detection error, **c** time–frequency amplitude at the IF position and **d** spectral concentration of each order LPFT



the values of f_{\min} and f_{\max} are the frequencies corresponding to $\sqrt{2}/2$ of the maxima of TFA.

Based on Eq. (13), when the SC reaches the maximum value, the kernel function of LPFT

$$w(t - \tau) \exp(-i\omega t) \exp \left\{ -i \left[\sum_{k=1}^n \frac{1}{k! \cdot (k+1)} c_k(\tau) (t - \tau)^{k+1} \right] \right\} \tag{14}$$

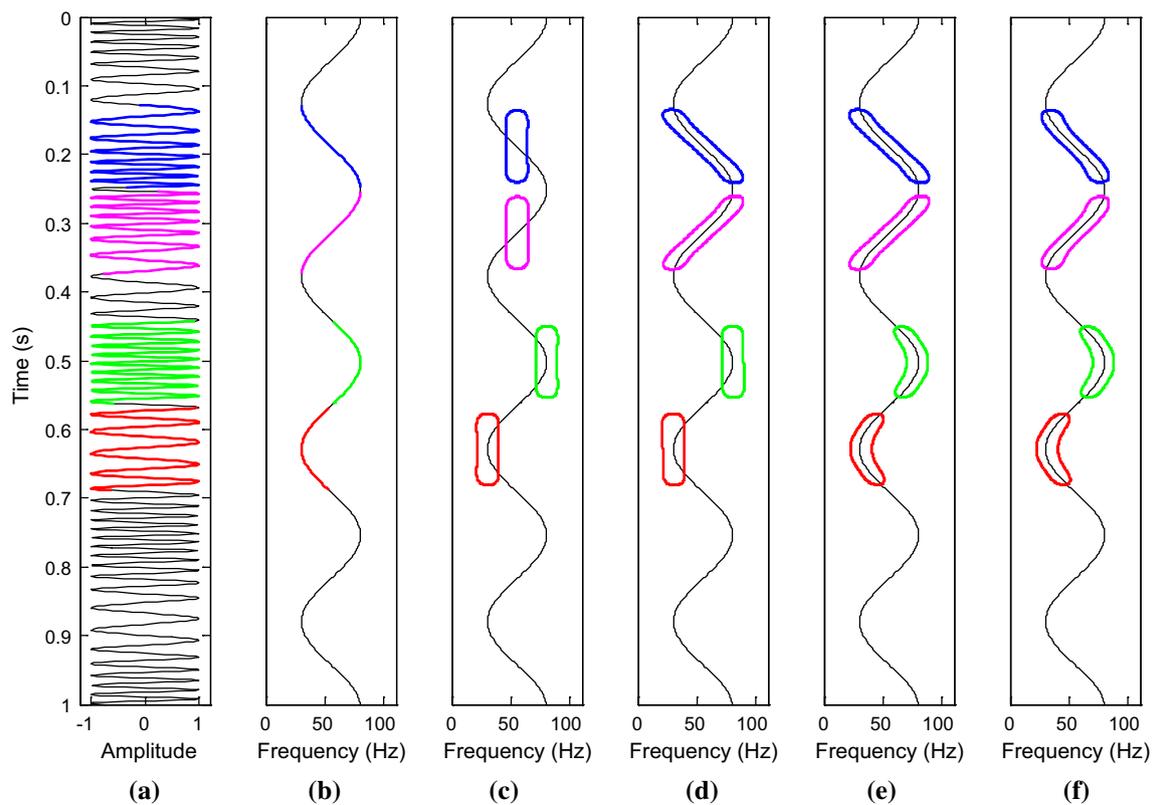
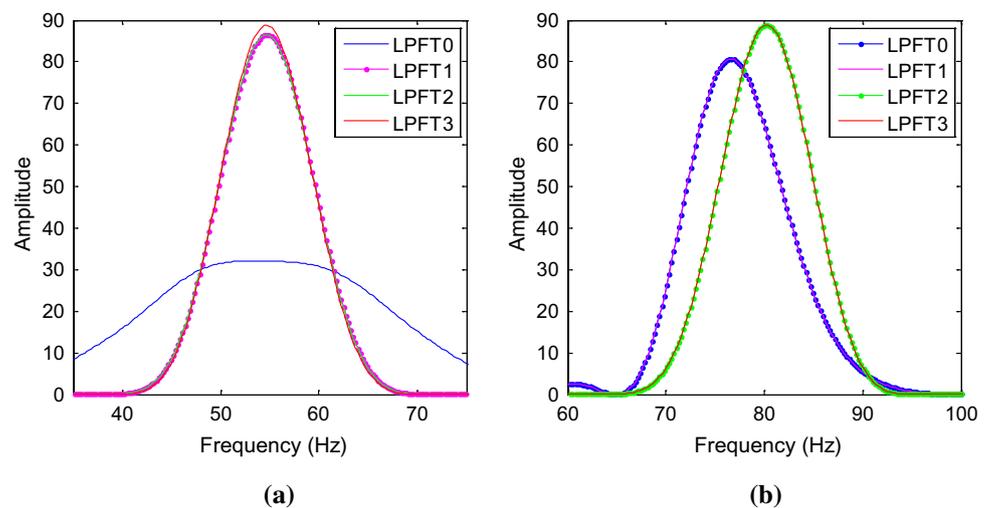


Fig. 3 The correspondence between the spectral shape of the LTFT kernel function and the IF of the signal. **a** Signal, **b** IF, **c** LPFT0, **d** LPFT1, **e** LPFT2 and **f** LPFT3

Fig. 4 Spectral amplitude at the typical time **a** 0.188 s and **b** 0.502 s in the time–frequency spectrum of signal (10)



and the frequency variation characteristics of the signal tend to be consistent, and the concentration of the spectrum is the highest. Figure 2d shows the SC of each order LPFT. As the order increases, the spectral concentration of the LPFT is enhanced correspondingly.

For further analysis, the time–frequency spectral contours of the integral of Eq. (14) are calculated targeting the signal

(10) with 0.188 s, 0.314 s, 0.502 s, and 0.629 s as the time window centers respectively (Fig. 3). Meanwhile, the spectral amplitudes at 0.188 s and 0.502 s (Fig. 4) are extracted to analyze the properties of the LPFT kernel function.

1. As shown in Fig. 3c, the time–frequency shape of the LPFT0's kernel function is independent of the IF, which

is always parallel to the time or frequency axis. Therefore, when the time–frequency shape of kernel function is more consistent with IF (0.502 s and 0.629 s), the spectral concentration is higher and the time–frequency amplitude is larger (blue dotted line in Fig. 4b). When the time–frequency shape of kernel function is less consistent with IF (0.188 s and 0.314 s), the spectral concentration is lower and the time–frequency amplitude is smaller (blue line in Fig. 4a).

- For LPFT1 (Fig. 2d), the exponential coefficient of the kernel function is a linear function, so the time–frequency shape can be rotated by a certain angle along the time axis or the frequency axis. Thus for 0.188 s and 0.314 s, the time–frequency spectra of the kernel function are tilted along the direction of the IF. Compared with LPFT0, the spectral concentration and time–frequency amplitude are improved (pink dotted line in Fig. 4a). For 0.502 s and 0.629 s, the time–frequency shapes of the kernel function are similar to that of LPFT0, so the spectral concentration and time–frequency amplitude are close to LPFT0 (pink line in Fig. 4b).
- For LPFT2 and LPFT3, The exponential coefficients of the kernel function are quadratic and cubic, respectively, so the time–frequency shape can change nonlinearly. For 0.188 s and 0.314 s, the time–frequency shape of LPFT3 is closest to the IF, so its spectral concentration is the highest with the largest time–frequency amplitude (red line in Fig. 4a). The time–frequency shape of LPFT2 is similar to that of LPFT1, so its time–frequency amplitude is also close to that of LPFT1 (green line in Fig. 4a). For 0.502 s and 0.629 s, the time–frequency shapes of LPFT2 and LPFT3 are similar. They all bend with the change of IF, and high degrees of coincidence are obtained. Therefore, the time–frequency amplitudes have reached maximum values, meanwhile the spectrum concentration is the highest (green dotted line and red line in Fig. 4b).

Therefore, under the condition that the coefficients of each order of the polynomial demodulation operator are known in Eq. (7), LPFT can obtain the time–frequency analysis results with a high spectral concentration. For the real signal, it is difficult to obtain the derivatives directly since the IF is unknown. The calculation procedure of n th-order LPFT is designed following the parameter estimation recursive algorithm of Yang et al. 2014:

- Since LPFT0 does not contain the polynomial demodulation operator, the initial time–frequency spectrum is calculated based on LPFT0 firstly.
- Detect the time–frequency ridge to obtain the IF $f(t)$.

- For each time position τ , the derivatives $f^{(k)}(\tau)$ ($k = 1, 2, \dots, n$) are obtained by the least squares method within the time window $w(t - \tau)$ according to Eq. (3).

Suppose the length of the time window is $2m + 1$, Eq. (3) is written in the matrix form

$$\mathbf{A}_{2m \times n} \mathbf{x}_{n \times 1} = \mathbf{b}_{2m \times 1}, \quad (15)$$

where

$$\mathbf{A} = \begin{bmatrix} (t_{(-m)} - \tau) & \frac{1}{2!}(t_{(-m)} - \tau)^2 & \frac{1}{3!}(t_{(-m)} - \tau)^3 & \dots & \frac{1}{n!}(t_{(-m)} - \tau)^n \\ \dots & \dots & \dots & \dots & \dots \\ (t_{(-1)} - \tau) & \frac{1}{2!}(t_{(-1)} - \tau)^2 & \frac{1}{3!}(t_{(-1)} - \tau)^3 & \dots & \frac{1}{n!}(t_{(-1)} - \tau)^n \\ (t_{(1)} - \tau) & \frac{1}{2!}(t_{(1)} - \tau)^2 & \frac{1}{3!}(t_{(1)} - \tau)^3 & \dots & \frac{1}{n!}(t_{(1)} - \tau)^n \\ \dots & \dots & \dots & \dots & \dots \\ (t_{(m)} - \tau) & \frac{1}{2!}(t_{(m)} - \tau)^2 & \frac{1}{3!}(t_{(m)} - \tau)^3 & \dots & \frac{1}{n!}(t_{(m)} - \tau)^n \end{bmatrix}$$

$$\mathbf{x} = [f^{(1)}(\tau) \ f^{(2)}(\tau) \ f^{(3)}(\tau) \ \dots \ f^{(n)}(\tau)]^T,$$

$$\mathbf{b} = [f(t_{(-m)}) \ \dots \ f(t_{(-1)}) \ f(t_{(1)}) \ \dots \ f(t_{(m)})]^T - f(\tau).$$

The solution of Eq. (15) is

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b}, \quad (16)$$

where λ is the damping factor, and \mathbf{I} is unit matrix.

- Let $c_k(\tau) = f^{(k)}(\tau)$ ($k = 1, 2, \dots, n$), calculate the LPFT n of the signal based on Eq. (8).
- Return to step 2, the IF $f(t)$ is calculated cyclically. The termination condition lies in reaching the maximum iterations or the error of two adjacent IF is less than a preset threshold.
- Output the LPFT n satisfying the iteration termination condition.

Based on the flow mentioned above, the first–third-order LPFT time–frequency spectra of the signal (10) are computed and compared with the results of LCT and GLCT (Fig. 5). For convenience of comparison, two time–frequency tilt angles of $\pi/8$ (Fig. 5a) and $-\pi/8$ (Fig. 5b) are chosen for LCT, respectively. As the time–frequency slope in each LCT remains unchanged, the amplitude performance is either high or low, showing an intensive inconsistency and deviation from the real time–frequency amplitude of the signal. By changing the time–frequency slope in each local time–frequency point, GLCT improves the accuracy of time–frequency amplitude (Fig. 5c). Figure 5d–f is the time–frequency spectra of the first–third-order LPFT, respectively. As the order increases, the consistency and accuracy of the time–frequency amplitude step forward.

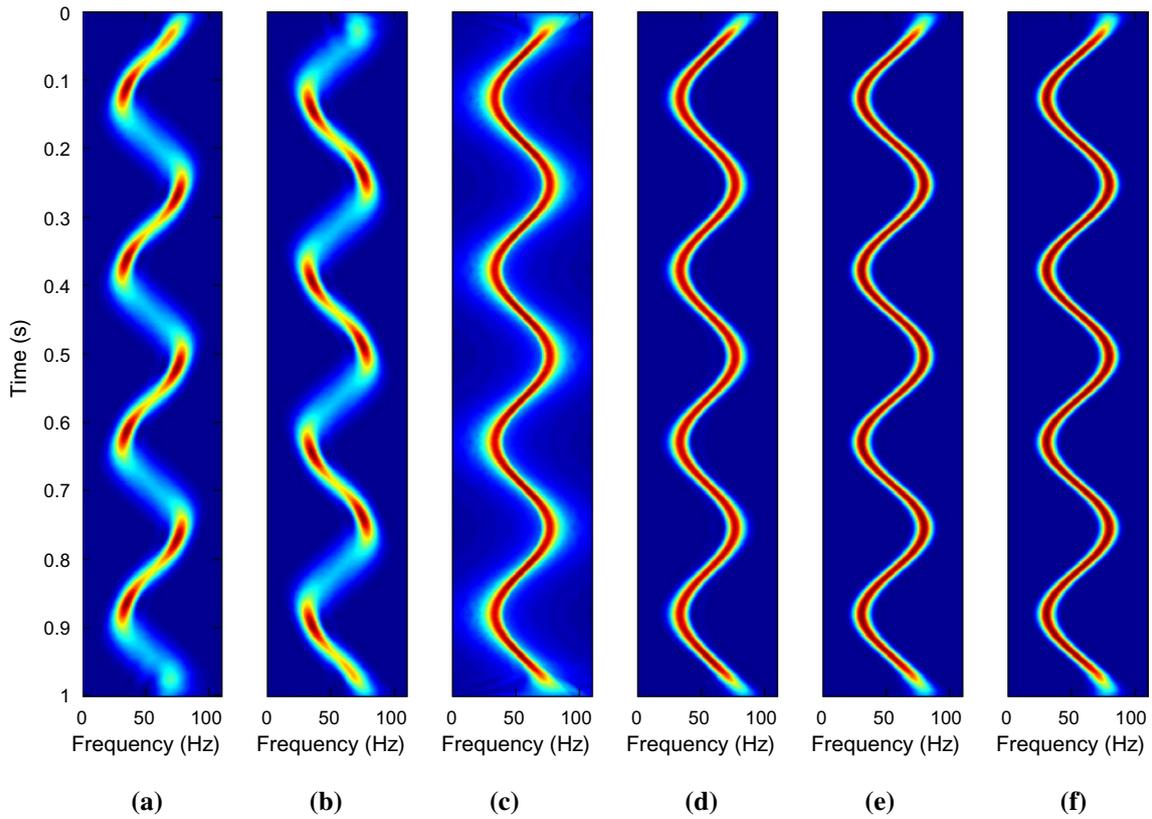


Fig. 5 Time–frequency spectra of LCT with time–frequency tilt angle of **a** $\pi/8$ and **b** $-\pi/8$, **c** GLCT, **d** LPFT1, **e** LPFT2 and **f** LPFT3

Fig. 6 IF detection and spectral concentration analysis. **a** The true and detection values of IF, **b** IF detection error, **c** time–frequency amplitude at the IF position, and **d** spectral concentration of GLCT and first–third-order LPFT

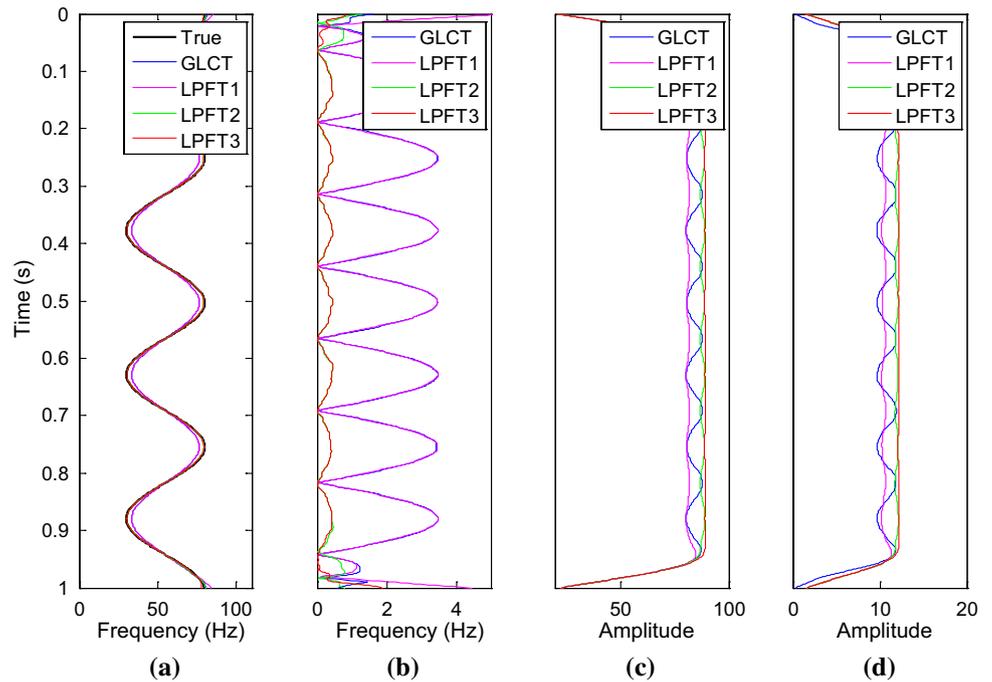


Fig. 7 Spectral amplitude at the typical time **a** 0.188 s and **b** 0.502 s in the time–frequency spectrum of signal (10)

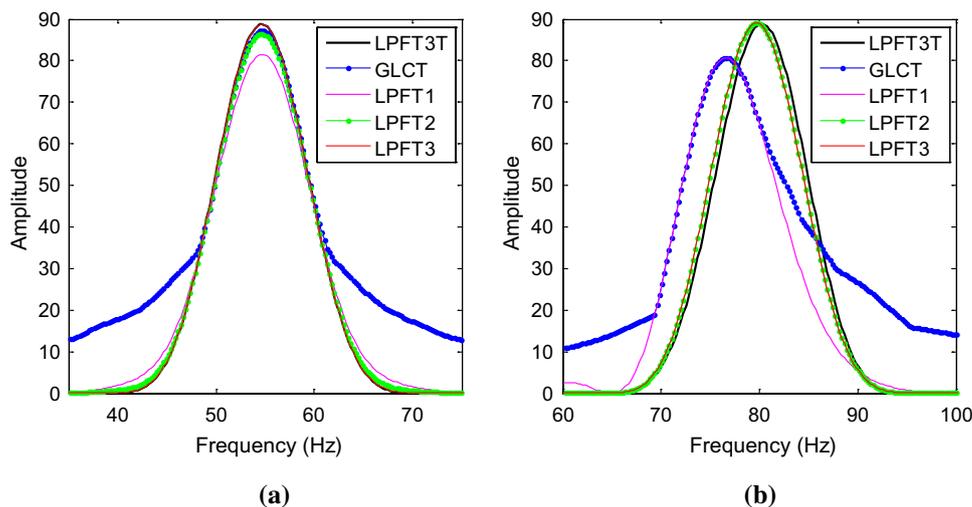


Figure 6 shows the results of IF detection and spectral concentration analysis about the four methods of GLCT, LPFT1, LPFT2 and LPFT3 in Fig. 5. It can be seen from the IF detection errors (Fig. 6b), since both GLCT and LPFT1 belong to the first-order fitting of the local time–frequency characteristics of the signal, the error values have the same magnitude. LPFT2 and LPFT3 reduce the IF errors by the high-order fitting of the local time–frequency characteristics. Comparative analysis of time–frequency spectra (Fig. 6c) shows LPFT3 has the largest amplitude. Furthermore, the SC of the four time–frequency spectra are calculated based on Eq. (13). Figure 6d shows that LPFT3 has the highest spectral concentration.

For further analysis, the spectral amplitudes of GLCT and first–third-order LPFT at time 0.188 s and 0.502 s are extracted (Fig. 7). Compared with the third-order LPFT based on the true IF derivatives (Fig. 1) which is represented by LPFT3T here, it can be seen that the result of GLCT deviates the farthest from LPFT3T. Compared with GLCT, the accuracy of each order LPFT is improved. The higher the order, the more obvious the improvement of the accuracy.

Examples

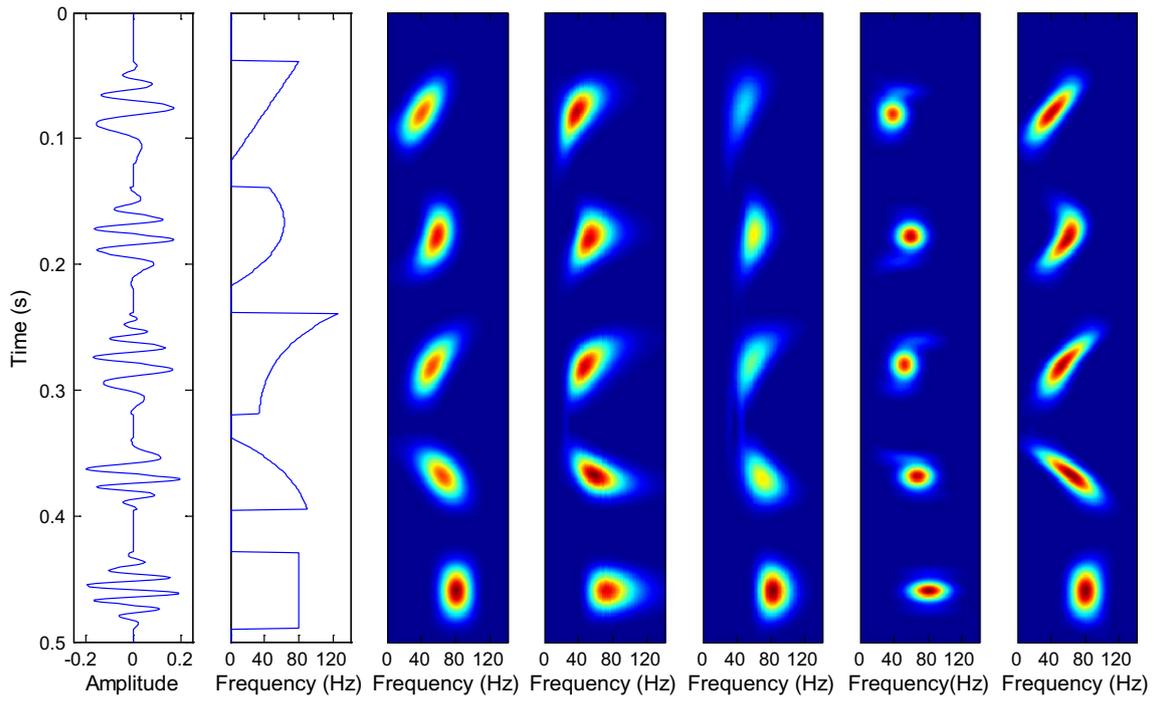
The synthetic traces are generated by using chirplet and Ricker wavelet with different IF characteristics, respectively. The LPFT of the traces are obtained. Considering the variation of magnitude in the IF, only the third order is calculated. It is also compared with the traditional time–frequency analysis methods in seismic exploration, which are STFT, CWT, ST, and the MP method based on the Ricker atomic dictionary (Liu and Marfurt 2007).

Figure 8a shows the synthetic trace with chirplet, IF and its time–frequency spectrum by STFT, CWT, ST, MP and LPFT, respectively. The five chirplets are generated by the formula

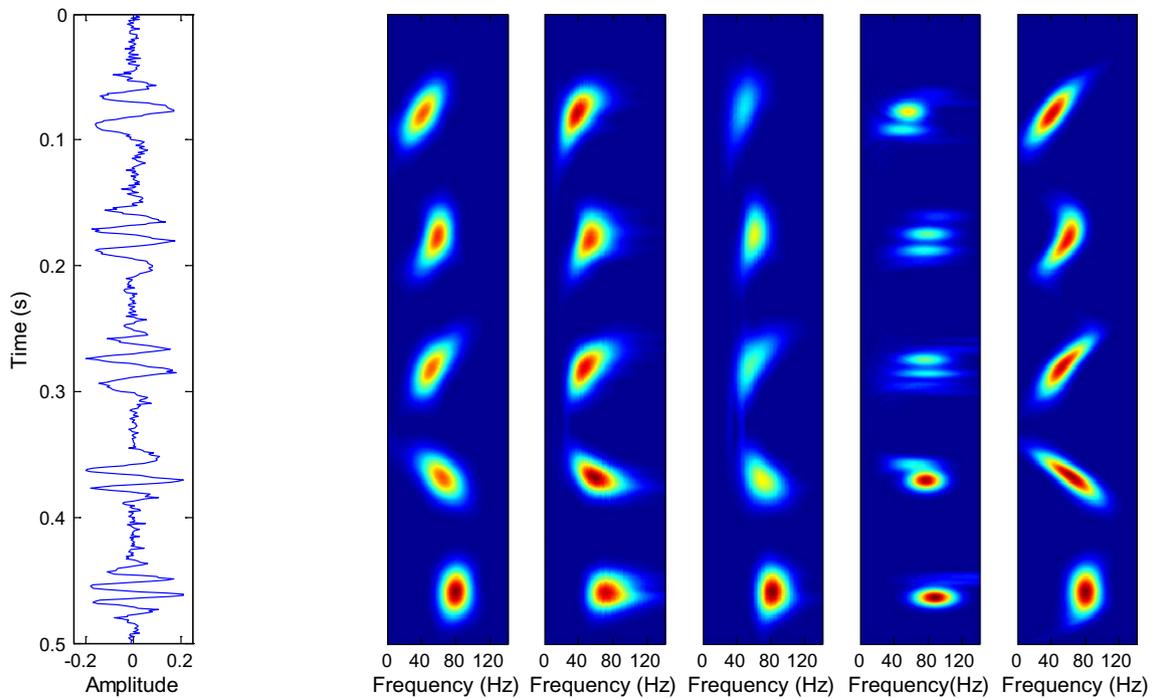
$$w(t) = \cos \{ 2\pi [\omega_1(t - \tau_1) + \omega_2(t - \tau_2)^2 + \omega_3(t - \tau_3)^3 + \omega_4(t - \tau_4)^4] \} \quad (17)$$

multiplied with a Gaussian window. The parameters of each chirplet are shown in Table 1.

Among the five chirplets, the IF of the first one changes linearly; the 2nd, 3rd and 4th are nonlinear; and the 5th is constant. For STFT, CWT and ST, the time–frequency morphology reflects the IF variation of each wavelet to a certain extent, but the spectral magnitude shows a large difference, especially for ST. This is due to the quasi-stationary assumptions of these methods. When the IF does not change in the time window, such as the fifth chirplet, the local signal conforms to the stationary characteristic. The perfect fitting of the signal can be obtained based on the orthogonal time–frequency grid, so the time–frequency shape and amplitude of the signal are well preserved. However, when the IF changes in the time window, such as the first–fourth chirplets, the local signal does not conform to the stationary features. The orthogonal time–frequency grid can only achieve an approximate fitting of the IF, which inevitably undermines the time–frequency variation and thus damages the time–frequency amplitude. For the MP method, the signal-based instantaneous autocorrelation transform can well preserve the time–frequency energy and have a higher time–frequency resolution. However, its time–frequency shape is difficult to reflect the variation characteristics of the IF. This is because the frequency-related parameters in the atomic dictionary only reflect the central frequency of the local signal and fail to reflect the frequency variation characteristics in the local time period. For LPFT, the polynomial demodulation operator can fit the IF effectively. Thus, the time–frequency shape can accurately reflect the IF variation characteristics, and the time–frequency energy of the signal can also be well preserved.



(a)



(b)

Fig. 8 Synthetic trace with chirplet and time–frequency spectra analysis. **a** Time–frequency spectra of noise-free data, from left to right, they are synthetic trace, IF, and the spectra by STFT, CWT, ST, MP

and LPFT, respectively. **b, c** Time–frequency spectra with 5% and 10% noise, respectively, where the subgraphs have the same meaning as **a**

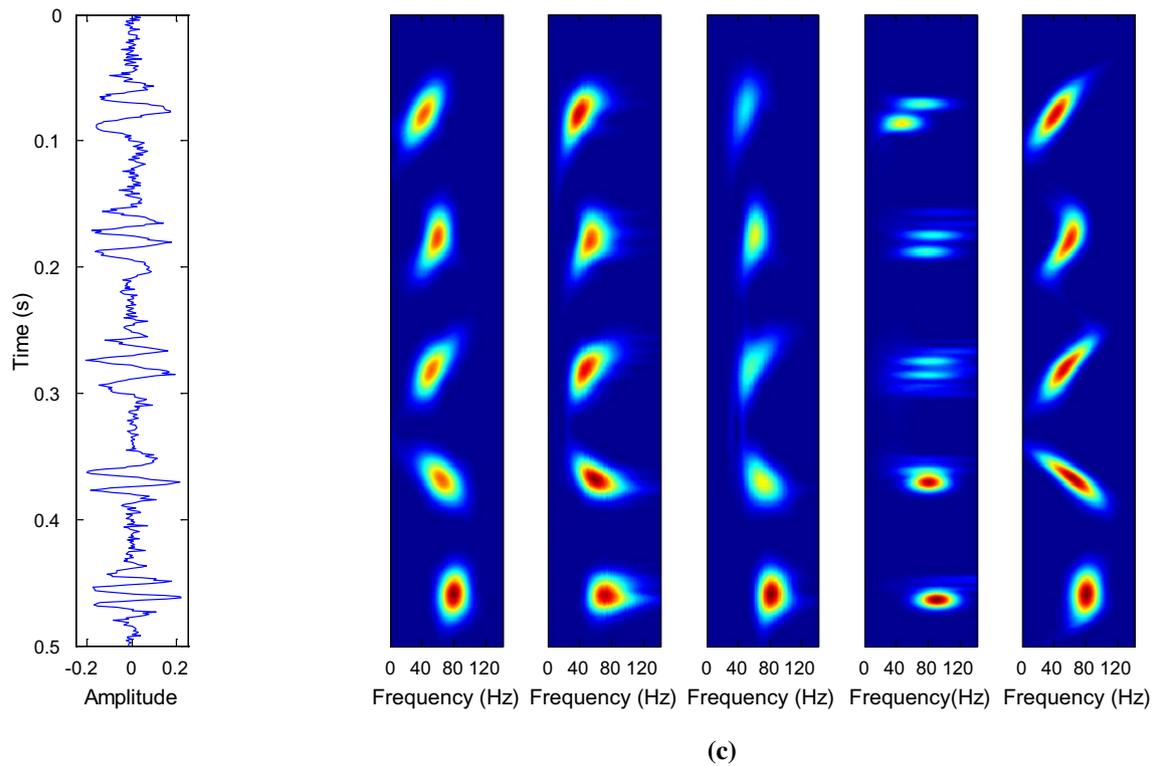


Fig. 8 (continued)

Table 1 The parameters of each chirplet in Fig. 8a

Number	ω_1 (Hz)	τ_1 (s)	ω_2 (Hz)	τ_2 (s)	ω_3 (Hz)	τ_3 (s)	ω_4 (Hz)	τ_4 (s)
1	30	0	500	0.05	0	0	0	0
2	50	0	300	0.055	8000	0.04	0	0
3	30	0	100	0.1	1500	0.1	30,000	0.1
4	100	0.028	100	0.1	1500	0.1	30,000	0.1
5	80	0.02	0	0	0	0	0	0

Figure 8b, c shows the time–frequency spectrum of different methods after adding, respectively, 5% and 10% random noise to the synthetic signal (Fig. 8a). It can be seen from the results that these methods all have anti-noise ability to a certain extent. Among the five methods, the deteriorated result of MP is mainly caused by the imperfect match between the atomic dictionary and signal.

Similar to Figs. 8a and 9a is a synthetic trace associated with the Ricker wavelet. With the increase of time, the wavelet is gradually changed from symmetrical to asymmetrical, and the IF gradually appears tailing, which corresponds to the dispersion in the process of seismic wave propagation (Wang 2004). The time–frequency energy characteristics of each wavelet are well preserved in five time–frequency

spectrum results. However, compared with STFT, CWT, ST and MP, the time–frequency morphology of LPFT reflects the IF variation characteristics of each wavelet more accurately. Therefore, LPFT has the advantage of retaining both time–frequency morphological features and energy features. The anti-noise capability of these methods shows up in the results of noisy synthetic trace (Fig. 9b, c). In addition, compared to Fig. 8b, c, the anti-noise ability of the MP method has been improved due to the perfect match of the time–frequency dictionary and the signal.

To further verify the ability of the proposed method, the interlayer velocity model is designed in Fig. 10. The model is divided into three groups from top to bottom, and the thickness of the interbed layer is gradually thinned. Synthetic trace

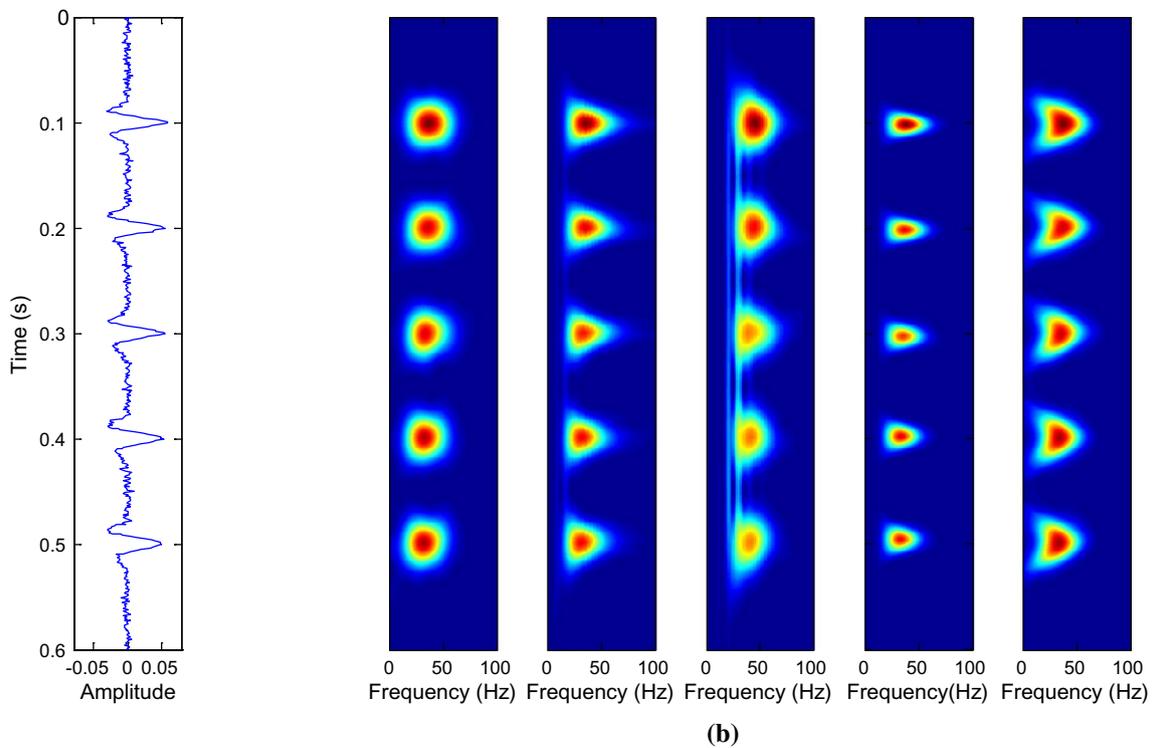
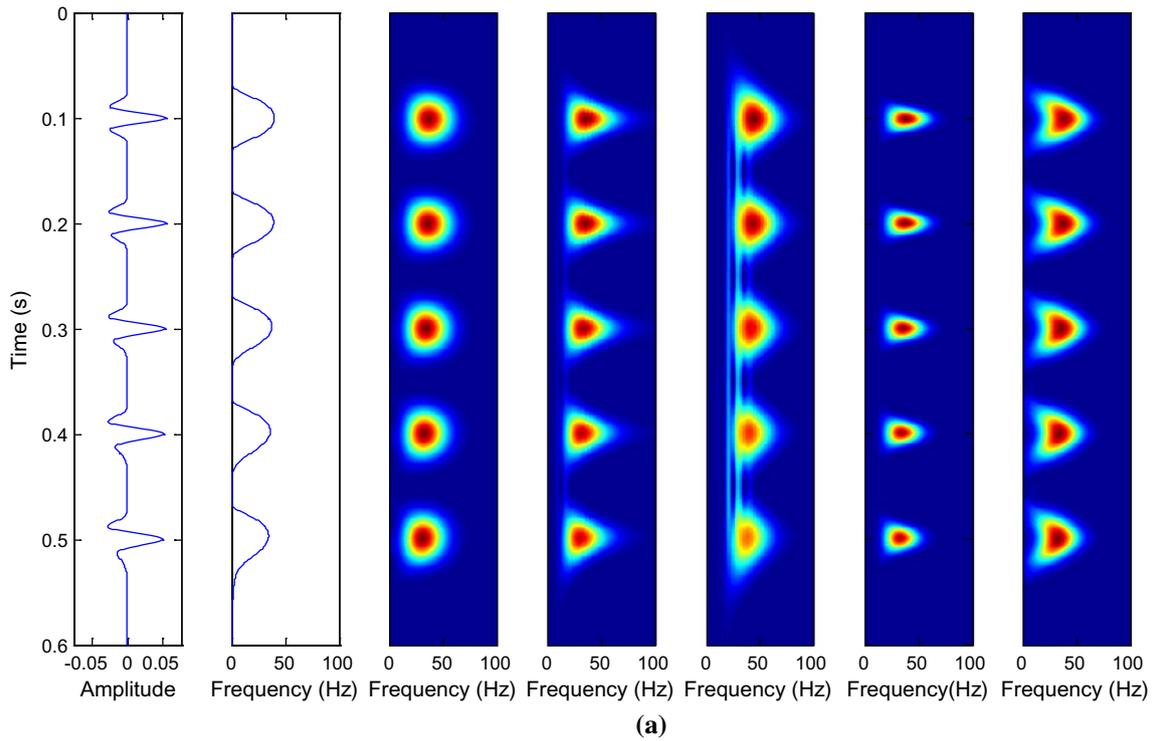


Fig. 9 Synthetic trace with Ricker wavelet and time–frequency spectra analysis. **a** Time–frequency spectra of noise-free data, from left to right, they are synthetic trace, IF, and the spectra by STFT, CWT,

ST, MP and LPFT, respectively. **b, c** Time–frequency spectra with 5% and 10% noise, respectively, where the subgraphs have the same meaning as **a**

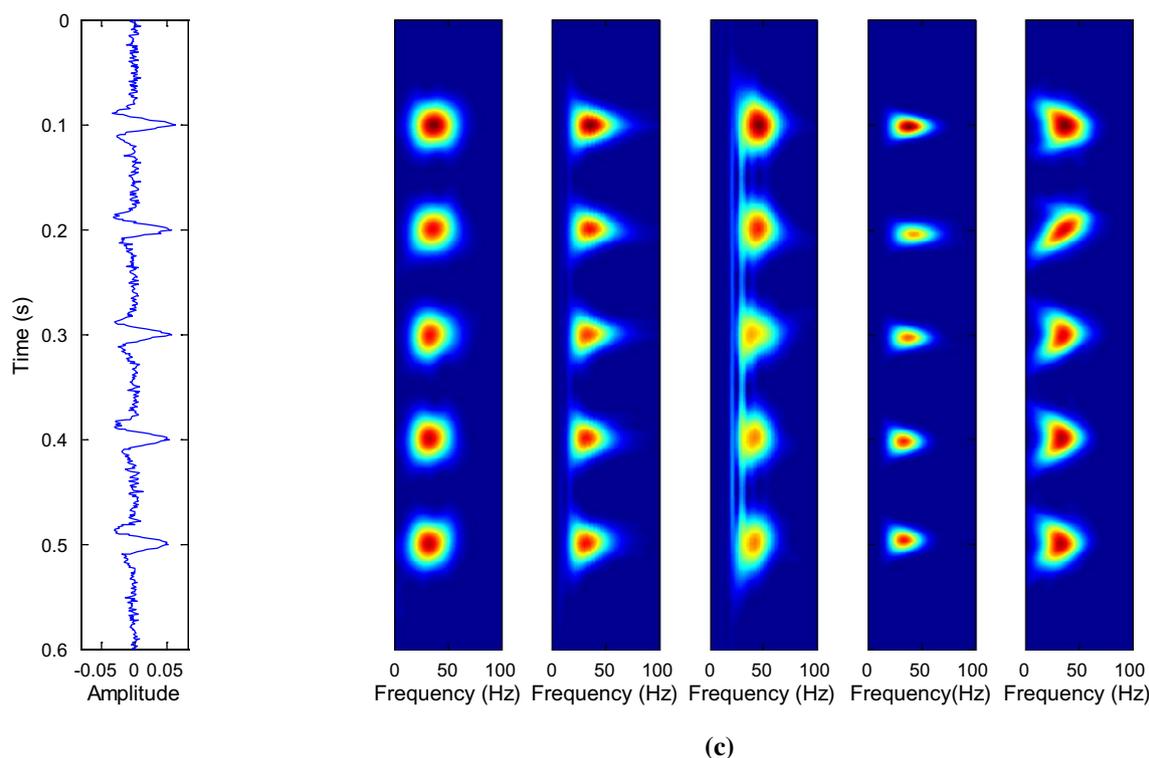


Fig. 9 (continued)

and IF curve are obtained based on the zero-phase Ricker wavelet (Fig. 11). For the first group of the interlayers, the IF curve changes slightly. STFT, CWT, MP and LPFT maintain the IF variation characteristics and energy relationship. For the second group of the interlayers, the IF shows wavy characteristics. STFT, CWT, ST and MP are all unable to describe the change, while only LPFT3 can retain the IF variation feature well. For the third group, the IF exhibits an arc-like feature. STFT, CWT, ST, and LPFT can all describe this time–frequency morphology, and the morphological variation of the LPFT spectrum is closest to the IF.

Since LPFT is a high-order form of STFT, the computational cost of LPFT is greater than that of STFT. For a more comprehensive comparison, the time consumption of different TFA methods in the three examples (Figs. 8, 9, 10) was recorded. The CPU of the tested computer is Intel Xeon 5 2.8 GHz, and the RAM is 16 GB. It can be seen from Table 2 that, among the five methods, STFT and ST take less time, CWT takes moderate time, and MP and LPFT take more time.

Applications

Figure 12 shows a two-dimensional post-stack seismic profile that has undergone the conventional processing such as static correction, denoising, deconvolution, migration and

stacking. The seismic reflection characteristics of typical fluvial facies strata are displayed between 2.9 and 3.1 s, that is, short axial strong reflection in the overall weak reflection background. The weak reflection mainly reflects the mudstone background, while the strong reflection is usually the response of channel sandstone. The three events marked in the elliptic region in Fig. 12 have already been proved to be the response of three thin channel sandstones stacked vertically. The planar distribution characteristics of the three channel sandstones are shown in Fig. 13. The representative seismic traces of the channel sandstones (trace number 95) and the mudstone background (trace number 170) are extracted, and the corresponding time–frequency spectra are calculated.

Figure 14 shows the time–frequency spectra of the seismic trace at the channel sandstones. On the whole, the channel sand bodies between 2.9 and 3.1 s have strong time–frequency energy. By comparison, note that CWT and ST cannot distinguish the response of the three sand bodies well. Although STFT distinguishes the three sand bodies, typical low-frequency oscillations occur in the time–frequency spectrum due to the short-time window (Castagna et al. 2003). MP shows a high resolution, and the response of the three sand bodies are all labeled as independent energy

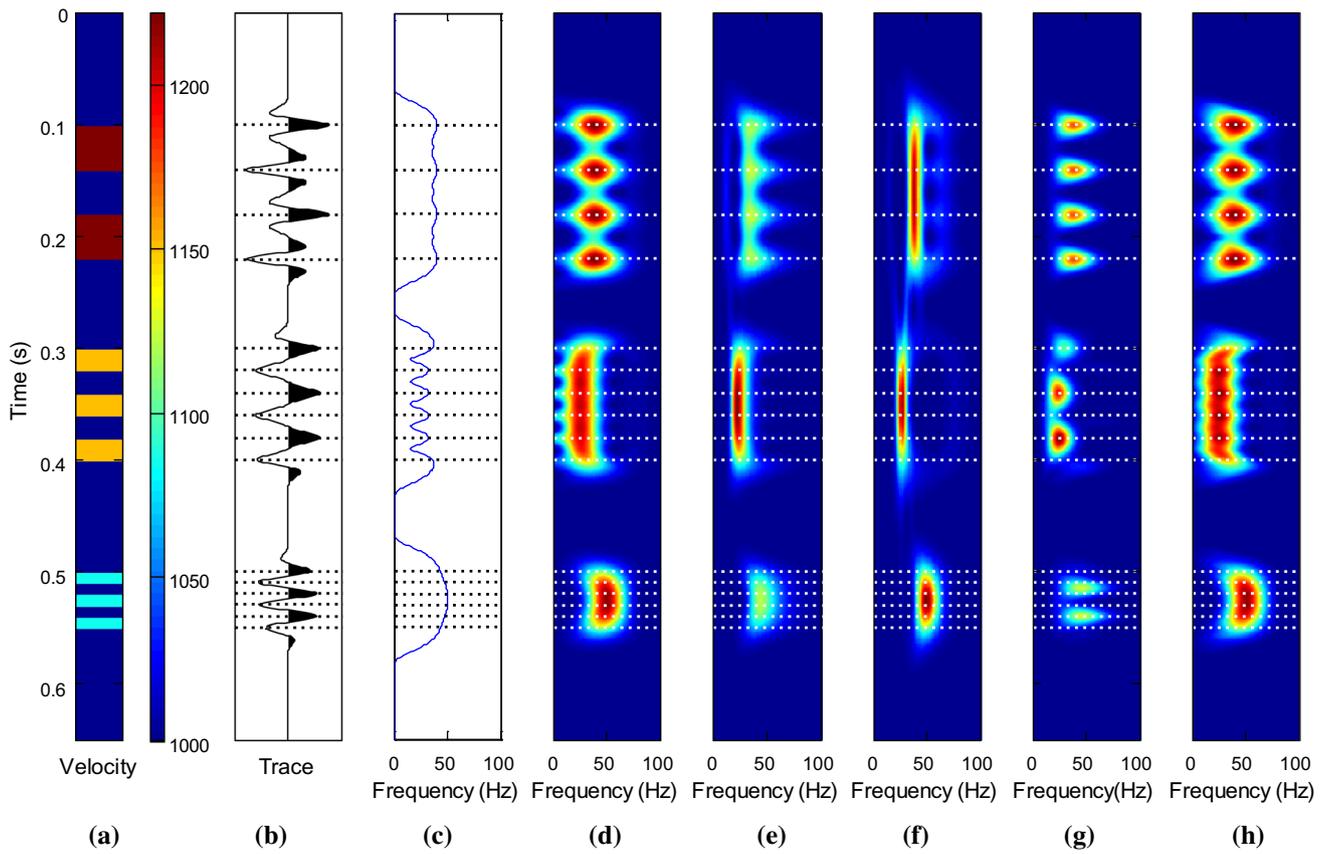


Fig. 10 Time–frequency spectra analysis of seismic trace with interbedded model. From left to right, they are **a** velocity model, **b** synthetic trace, **c** IF and time–frequency spectrum by **d** STFT, **e** CWT, **f** ST, **g** MP and **h** LPFT, respectively

Fig. 11 Ricker wavelet **(a)** and the amplitude spectrum **(b)**

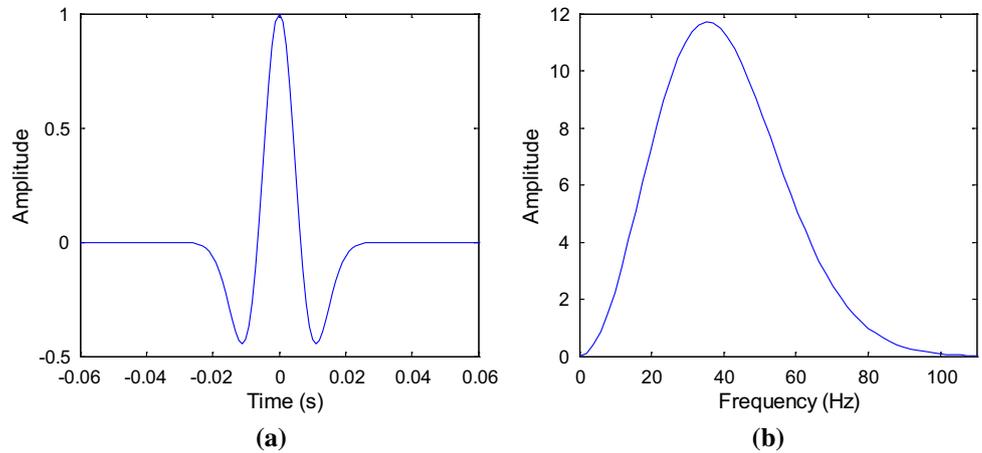


Table 2 The cost time of different TFA methods on the three traces in Figs. 8, 9 and 10

Synthetic trace	STFT (s)	CWT (s)	ST (s)	MP (s)	LPFT (s)
Chirplet	0.12	0.25	0.09	0.32	0.39
Ricker	0.14	0.29	0.11	0.41	0.48
Interbed	0.17	0.42	0.15	0.58	0.64

clusters. However, the shapes are similar, the local frequency variation characteristics of each waveform cannot be delineated. LPFT not only discerns the three sand bodies, but also depicts the local frequency variation characteristics.

Figure 15 shows the time–frequency spectra of the seismic trace at the mudstone background. Since the seismic reflection between the target layers is weak, the corresponding

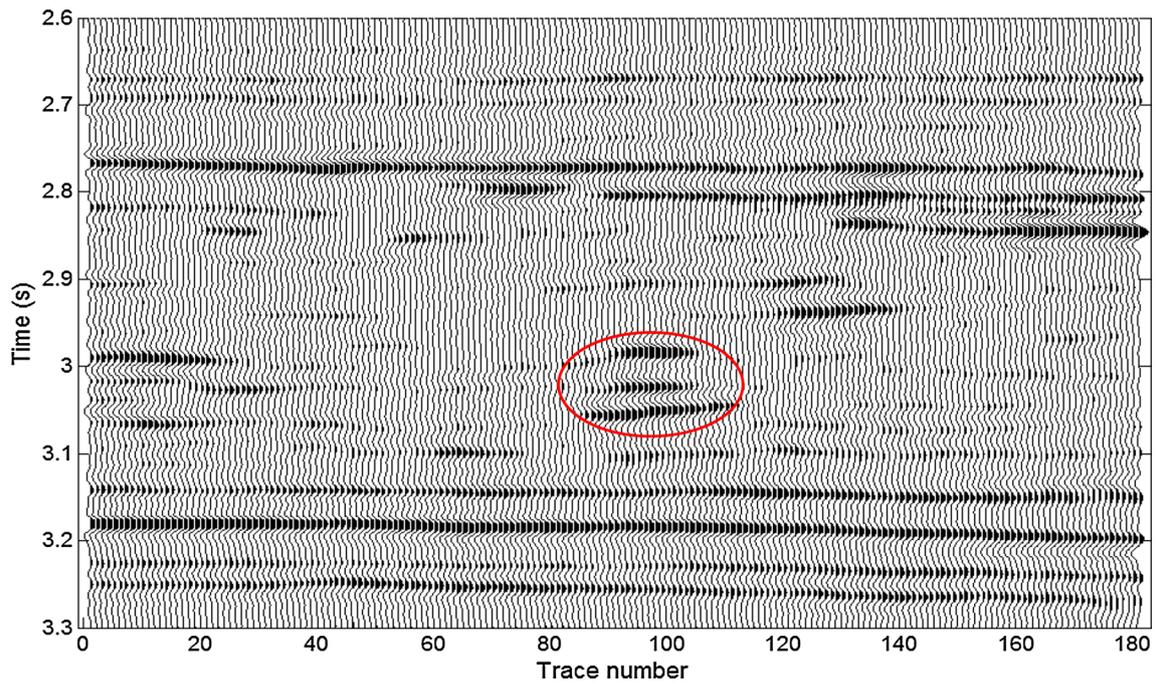
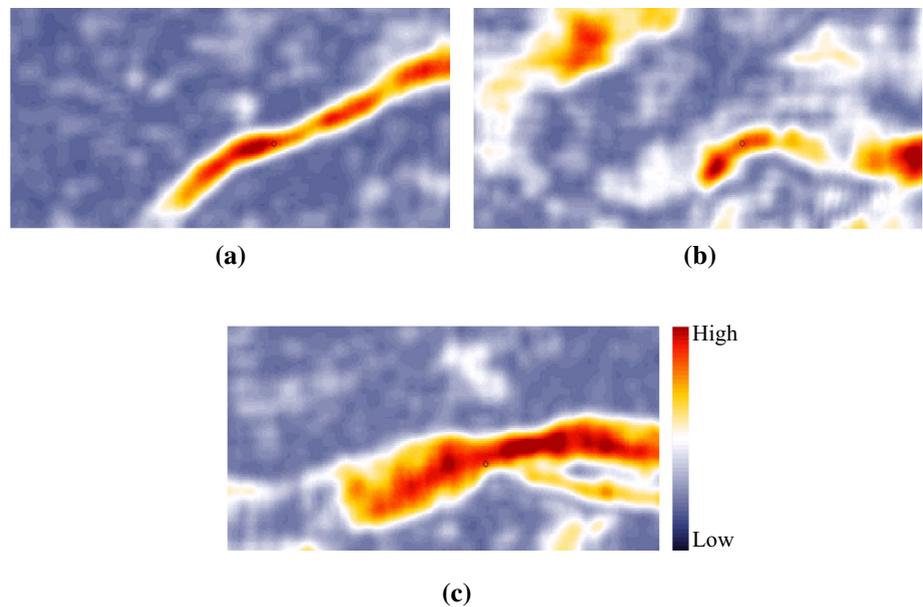


Fig. 12 Real seismic data. The events marked by the elliptic are the response of three channel sandstones

Fig. 13 Planar distribution characteristics of the three channel sandstones in Fig. 12 based on amplitude attribute. The position of the black circle corresponds to the 95th trace in Fig. 12



time–frequency energy is also feeble. However, outside the target layer, from the time–frequency spectrum of strongly reflected waveform at 2.8 s, it can be seen that MP and LPFT

have a high resolution. For the time–frequency spectrum of strongly reflected waveform at 3.2 s, LPFT well retains the IF variation characteristics of seismic waves.

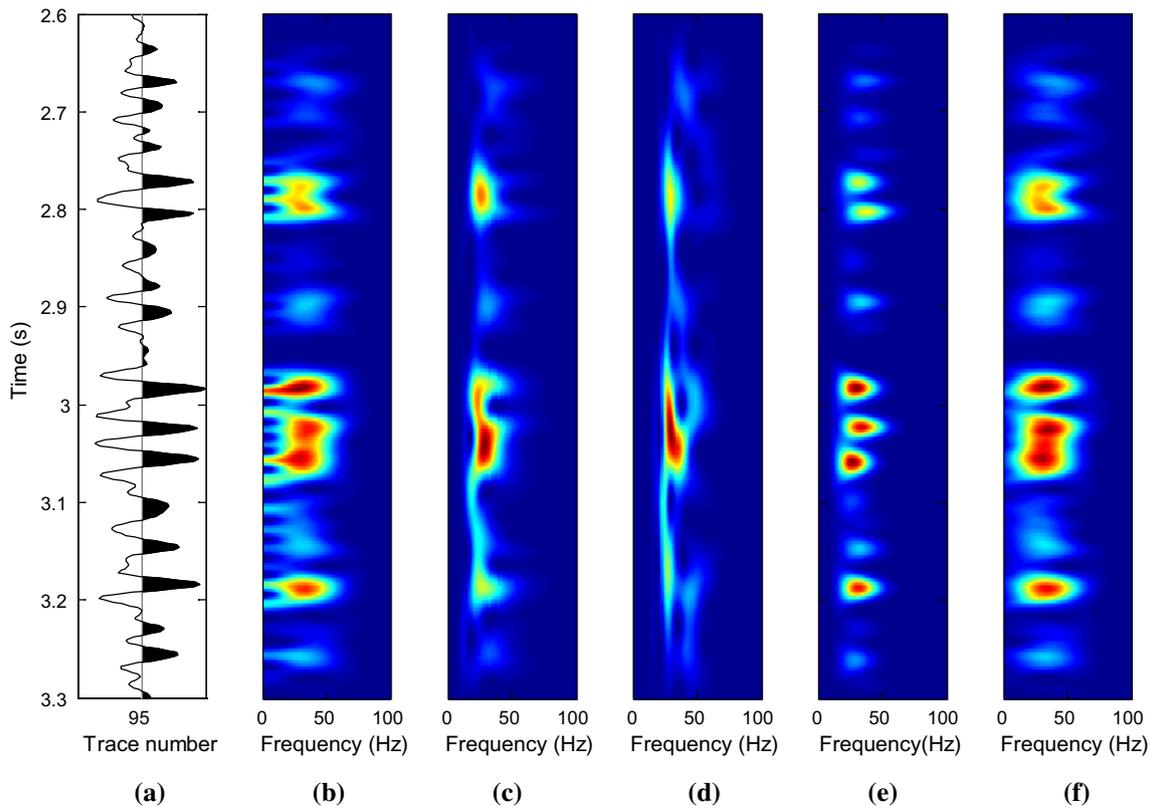


Fig. 14 Time–frequency spectrum analysis of seismic trace at the location of channel sandstones. From left to right, they are **a** seismic trace, and time–frequency spectrum by **b** STFT, **c** CWT, **d** ST, **e** MP and **f** LPFT, respectively

Conclusions

The LPFT method based on the Ville’s analytical signal shows that LPFT is a higher order form of STFT and LCT. Through the polynomial demodulation operator, LPFT can achieve high-order fitting of local time–frequency features of the signal through accurate determination of the demodulation operator coefficients. The calculation process of LPFT given in this paper can solve the demodulation operator coefficients, and then obtain the time–frequency spectrum results with high energy concentration.

For non-stationary seismic traces, LPFT not only exhibits high time–frequency resolution, but also

effectively preserves the instantaneous frequency and amplitude variation characteristics of local seismic signals. These properties can be utilized for target detection with frequency and amplitude anomalies, for example, in gas-bearing reservoirs with dispersion and attenuation characteristics. LPFT can be applied to prestack gathers, with which the prestack prediction of gas-bearing reservoirs can be realized by using the varied characteristics of frequency with offset. In addition, LPFT can be combined with spectral decomposition technology to realize the post-stack prediction of gas-bearing reservoirs related to low-frequency shadow phenomenon. These potential research will be carried out in the future.

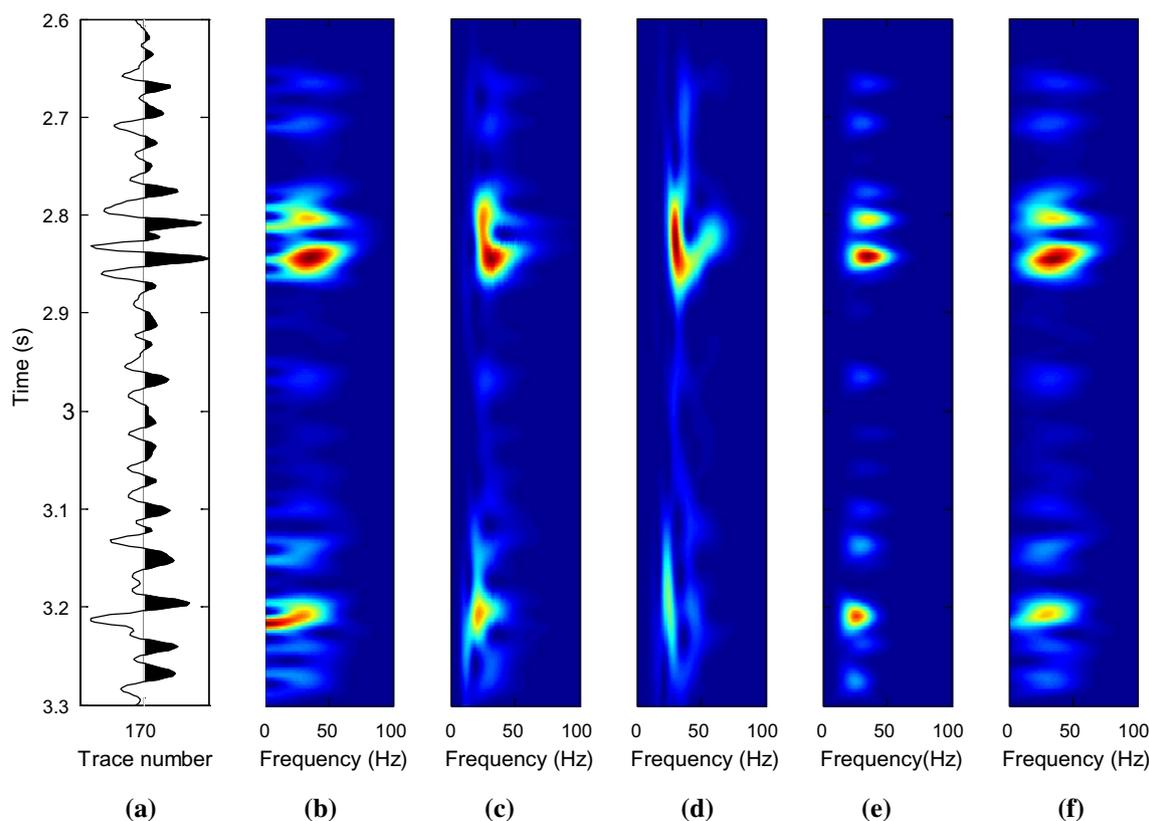


Fig. 15 Time–frequency spectrum analysis of seismic trace at the location of the mudstone background. From left to right, they are **a** seismic trace, and time–frequency spectrum by **b** STFT, **c** CWT, **d** ST, **e** MP and **f** LPFT, respectively

Acknowledgements This work was financially supported by the National Science and Technology Major Project of China (2016ZX05003-003) and Inversion Methods Study of Sedimentary Model Constrained Lithofacies and Fluid Identification (P18075-6).

References

- Castagna JP, Sun SJ, Siegfried RW (2003) Instantaneous spectral analysis: detection of low-frequency shadows associated with hydrocarbons. *Lead Edge* 22(2):120–127
- Cohen L (1995) *Time–frequency analysis*. Prentice Hall, Upper Saddle River
- Durak L, Arıkan O (2003) Short-time Fourier transform: two fundamental properties and an optimal implementation. *IEEE Trans Signal Process* 51(5):1231–1242
- Gribonval R (2001) Fast matching pursuit with a multiscale dictionary of Gaussian chirps. *IEEE Trans Signal Process* 49(5):994–1001
- Kadambe S, Boudreaux-Bartels GF (1992) A comparison of the existence of ‘cross terms’ in the Wigner distribution and the squared magnitude of the wavelet transform and the short-time Fourier transform. *IEEE Trans Signal Process* 40(10):2498–2517
- Katkovnik V (1998) Discrete-time local polynomial approximation of the instantaneous frequency. *IEEE Trans Signal Process* 46(10):2626–2637
- Li X, Bi G, Stankovic S, Zoubir AM (2011) Local polynomial Fourier transform: a review on recent developments and applications. *Signal Process* 91(6):1370–1393
- Li D, Castagna JP, Goloshubin G (2016) Investigation of generalized S-transform analysis windows for time–frequency analysis of seismic reflection data. *Geophysics* 81(3):V235–V247
- Liu J, Marfurt KJ (2007) Instantaneous spectral attributes to detect channels. *Geophysics* 72(2):P23–P31
- Liu NH, Gao JH, Jiang XD, Zhang ZS, Wang Q (2016) Seismic time–frequency analysis via STFT-based concentration of frequency and time. *IEEE Geosci Remote Sens Lett* 14(1):127–131
- Mallat SG, Zhang Z (1993) Matching pursuits with time–frequency dictionaries. *IEEE Trans Signal Process* 41(12):3397–3415
- Mann S, Haykin S (1995) The chirplet transform: physical considerations. *IEEE Trans Signal Process* 43(11):2745–2761
- Naseer MT, Asim S (2017) Detection of cretaceous incised-valley shale for resource play, Miano gas field, SW Pakistan: spectral decomposition using continuous wavelet transform. *J Asian Earth Sci* 147:358–377
- Quadfeul SA, Aliouane L (2014) Random seismic noise attenuation data using the discrete and the continuous wavelet transforms. *Arab J Geosci* 7(7):2531–2537
- Parolai S (2009) Denoising of seismograms using the S transform. *Bull Seismol Soc Am* 99(1):226–234
- Partyka G, Gridley J, Lopez J (1999) Interpretational applications of spectral decomposition in reservoir characterization. *Lead Edge* 18(3):353–360

- Phinyomark A, Limsakul C, Phukpattaranont P (2011) Application of wavelet analysis in EMG feature extraction for pattern classification. *Meas Sci Rev* 11(2):45–52
- Puryear CI, Portniaguine ON, Cobos CM, Castagna JP (2012) Constrained least-squares spectral analysis: application to seismic data. *Geophysics* 77(5):V143–V167
- Radad M, Gholami A, Siahkoochi HR (2015) S-transform with maximum energy concentration: application to non-stationary seismic deconvolution. *J Appl Geophys* 100(118):155–166
- Rene RM, Fitter JL, Forsyth PM, Kim KY, Murray DJ, Walters JK, Westerman JD (1986) Multicomponent seismic studies using complex trace analysis. *Geophysics* 51(6):1235–1251
- Sinha S, Routh P, Anno P (2009) Instantaneous spectral attributes using scales in continuous-wavelet transform. *Geophysics* 74(2):WA137–WA142
- Smith M, Perry G, Stein J, Bertrand A, Yu G (2008) Extending seismic bandwidth using the continuous wavelet transform. *First Break* 26(6):97–102
- Sun S, Castagna JP, Siegfried RW (2002) Examples of wavelet transform time–frequency analysis in direct hydrocarbon detection. In: SEG technical program expanded abstracts, pp 457–460
- Terrien J, Marque C, Germain G (2008) Ridge extraction from the time–frequency representation (TFR) of signals based on an image processing approach: application to the analysis of uterine electromyogram AR TFR. *IEEE Trans Biomed Eng* 55(5):1496–1503
- Ville J (1948) Theorie et application de la notion de signal analytique. *Cables Transm* 93:61–74
- Wang YH (2004) Q analysis on reflection seismic data. *Geophys Res Lett* 31:L17606
- Wang YH (2006) Inverse Q -filter for seismic resolution enhancement. *Geophysics* 71(3):V51–V60
- Wang YH (2010) Multichannel matching pursuit for seismic trace decomposition. *Geophysics* 75(4):V61–V66
- Wang LL, Gao JH, Zhao W, Jiang XD (2012) Enhancing resolution of nonstationary seismic data by molecular-gabor transform. *Geophysics* 78(1):V31–V41
- Wang TY, Yuan SY, Gao JH, Li SJ, Wang SX (2019) Multispectral phase-based geosteering coherence attributes for deep stratigraphic feature characterization. *IEEE Geosci Remote Sens Lett* 16(8):1309–1313
- Wu L, Castagna JP (2017) S-transform and Fourier transform frequency spectra of broadband seismic signals. *Geophysics* 82(5):O71–O81
- Yang Y, Peng ZK, Dong XJ, Zhang WM, Meng G (2014) General parameterized time–frequency transform. *IEEE Trans Signal Process* 62(11):2751–2764
- Yu G, Zhou YQ (2016) General linear chirplet transform. *Mech Syst Signal Process* 70:958–973
- Yuan SY, Wang SX, Ma M, Ji YZ, Deng L (2017) Sparse Bayesian learning-based time-variant deconvolution. *IEEE Trans Geosci Remote Sens* 55(11):6182–6194
- Yuan SY, Ji YZ, Shi PD, Zeng J, Gao JH, Wang SX (2019a) Sparse Bayesian learning-based seismic high-resolution time–frequency analysis. *IEEE Geosci Remote Sens Lett* 16(4):623–627
- Yuan SY, Liu Y, Zhang Z, Luo CM, Wang SX (2019b) Prestack stochastic frequency-dependent velocity inversion with rock-physics constraints and statistical associated hydrocarbon attributes. *IEEE Geosci Remote Sens Lett* 16(1):140–144
- Zhong JG, Huang Y (2010) Time–frequency representation based on an adaptive short-time Fourier transform. *IEEE Trans Signal Process* 58(10):5118–5128



Data integration for earthquake disaster using real-world data

Chuanzhao Tian^{1,2} · Guoqing Li¹

Received: 6 January 2019 / Accepted: 25 October 2019 / Published online: 6 November 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

The purpose of entity resolution (ER) is to identify records that refer to the same real-world entity from different sources. Most traditional ER studies identify records based on string-based data, so the ER problem relies mostly on string comparison techniques. There is little research on numeric-based data. Traditional ER approaches are widely used in many domains, such as papers, gene sequencing and restaurants, but they have not been used in an earthquake disaster. In this paper, earthquake disaster event information that was collected from different websites is denoted with numeric data. To solve the problem of ER in numeric data, we use the following methods to conduct experiments. First, we treat numbers as strings and use string-based approaches. Second, we use the Euclidean distance to measure the difference between two records. Third, we combine the above two strategies and use a comprehensive approach to measure the distance between the two records. We experimentally evaluate our methods on real datasets that represent earthquake disaster event information. The experimental results show that a comprehensive approach can achieve high performance.

Keywords Data integration · Earthquake disaster · Numeric data · Entity resolution

Introduction

Recently, as increasing commercial data, government agencies data and scientific research data have become available, and data science, especially big data science, has become increasingly more important. Data all over the world have reached PB levels, and the techniques that allow efficient mining and analyzing of data have attracted increasing interest. However, before data are used, data usability is ensured first, which also makes the problem of data management as an important concern. However, there are no unique identifiers that tell us which records from one database correspond to those in other databases. The primary problem is matching records that relate to the same entities in the real world from many databases, which has been of rising importance in many domain applications. This problem is called entity resolution (Ayat et al. 2014; Elmagarmid et al. 2006), record linkage, deduplication and record matching.

Entity resolution has been commonly used for improving data quality and integrity (Ayat et al. 2012; Magnani and Montesi 2010) to reuse already existing data sources for new studies and to reduce the cost of data acquisition (Winkler 2004; Christen 2011). Clark (2004) and Kelman et al. (2010) used linked data to collect information that can improve health policies, information that is traditionally collected with expensive surveys and time-consuming methods.

Record linkage is used not only in the health sector but also in statistical agencies to link census data for future analysis (Winkler 2006). Ayat et al. (2014) solved the entity resolution problem in the probabilistic data domain for finding astronomical objects in astrophysics data and suspect detection in anticriminal police databases. Li et al. (2015) used a rule-based method for entity resolution that achieved high performance.

One important area of record linkage is approximate string matching (Monge 2000). Many scholars have studied string matching, and it has been one of the most studied domains in information science (Miller et al. 1980; Chang and Lampe 1992; Knuth et al. 1977; Zhu et al. 2018; Du and Chang 1994; Galil and Giancarlo 1988; Levenshtein 1966; Boyer and Moore 1977; Peterson 1980; Lee et al. 2014). In (Levenshtein 1966), the main method was based on the edit distance. Edit distance is the minimum number of operations

✉ Guoqing Li
ligq@radi.ac.cn

¹ Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China

² University of Chinese Academy of Sciences, Beijing 100049, China

on individual characters required to transform one string into another string (Miller et al. 1980; Peterson 1980). In Miller et al. (1980), two different problems were considered by the authors, one under the definition of equivalence and a second using similarity. In Gomaa and Fahmy (2013), the authors presented many types of string matching similarity measures.

Currently, the methods for entity resolution can be roughly divided into two systems. One is a pairwise method, which performs the pairwise matching of records and then determines whether the two records are similar according to the similarity measure of the record pair. The second method is based on clustering. This method clusters records by defining distance metrics that measure the degree of similarity between records. Records that are grouped together in the same class are considered similar.

The entity resolution method based on pairwise comparison is a simple solution. Its basic idea is to calculate the similarity of two records. If the similarity exceeds the predefined similarity threshold, the two records are considered similar; otherwise, the two records are not considered similar. Therefore, the key technology based on record-to-comparison is to define the attribute's comparison function so that the returned value reflects the degree of similarity between the values of the two tuples on the attribute.

On the similarity measure function of the attribute, there are many similarity comparison functions of the string: edit distance (Khan et al. 2011; Monge 2000), affine gap distance (Waterman et al. 1976), Smith–Waterman distance (Pinheiro and Sun 1998; Baeza-Yates and Gonnet 1992), Q-gram distance (Sutinen and Tarhio 1995; Ukkonen 1992), Jaro distance metric (Jaro 1980) and so on (Navarro 2001).

The classification of the entity resolution and whether the record is the same entity is divided into categories; different entities are divided into different categories. There are many kinds of entity identification and classification technologies. According to the classification process, the samples are trained and can be divided into unsupervised classification and supervised classification (Sun et al. 2016).

The supervision method relies on the user to provide accurate training data to train the rules or data models to identify the test data based on the classification algorithm (Köpcke et al. 2010). Training data need to be marked by experts in the field, but the algorithm's user is not always a domain expert, which leads to difficulties in obtaining training data and therefore results in limitations in the supervisory approach. It is not easy to obtain accurate training data, and training data take a lot of time. The focus of this paper is the unsupervised entity identification method. Traditional entity recognition determines two entities by measuring a fixed threshold and whether the data object (object for short) matches; however, this method has been proven to be unable to produce high-accuracy results (Hassanzadeh et al. 2009).

The problem of entity resolution has been studied by many researchers. However, most fields of research are paper matching, mail markets, biological genes and other string types. Although there are many similarity comparison functions for strings, the similarity comparison methods for numeric data are still primitive. It is typical to use numbers as strings or simple range queries to find similar values. Koudas et al. (2004) considered the type and distribution of numeric data in the discovery of redundant tuples and extended the concept of cosine similarity to measure numeric data. Disaster data also have the problem of entity resolution. Through the analysis of disaster events, taking earthquake disasters as an example, the attribute information of events is mostly numeric data. There are few studies on the entity identification of numeric data; even if there are studies on numeric data, most of them are of definite value, such as zip codes and phone numbers.

Since the expression of earthquake disaster events comes from different websites around the world, the language used will be different, and the description of place names may be different. Therefore, it is very difficult to use natural language to identify disaster events. Therefore, earthquake disaster events can be expressed as a set of entities of numeric data for entity identification. This paper is based on the above, using numeric data as the research object. Through the identification of disaster events, it solves the problem of whether the data obtained by different websites in the actual situation correspond to the same event.

The rest of this paper is organized as follows. In Sect. 2, we briefly describe the earthquake events and the problems. Then, Sect. 3 presents techniques and programs for matching records. Section 4 describes the data and tests the experiment, and Sect. 5 discusses the experimental results. Finally, Sect. 6 concludes the paper and discusses interesting directions for future work.

Problem statement

The data of earthquake disaster events all over the world were collected from professional websites, such as USGS (United States Geographical Survey) Earthquake Hazards Program and CENC (China Earthquake Network Center). In this paper, an earthquake event is described by its attributes, such as the time of the earthquake; the location of the epicenter, which contains latitude and longitude; and the magnitude of the earthquake. Table 1 shows eight records with magnitudes of approximately 5.8 identified by r_{ij} . The records with ids r_{11} , r_{12} and r_{13} refer to the same earthquake event denoted as e_1 ; the records with r_{21} , r_{22} and r_{23} refer to e_2 ; and the records with r_{31} and r_{32} refer to e_3 . Thus, identifying e_1 , e_2 and e_3 using the information given in Table 1 is a task for entity resolution. Because of the information

Table 1 Different records from different sources of earthquake events (may be the same events)

Entity	Id	Magnitude	Latitude	Longitude	Time
e_1	r_{11}	5.8	53.23	65.32	2017-10-28 19:11:03
	r_{12}	5.7	53.24	65.33	2017-10-28T19:11:02.760Z
	r_{13}	5.8	53.22	65.34	2017-10-29 03:11:03
e_2	r_{21}	5.8	78.63	12.36	2017-10-28T16:16:06.930Z
	r_{22}	5.7	78.64	12.36	2017-10-28 16:16:14
	r_{23}	5.9	78.65	12.364	2017-10-29 00:16:14
e_3	r_{31}	5.6	23.654	32.724	2017-10-28T16:13:54.390Z
	r_{32}	5.9	23.656	32.726	2017-10-28 16:13:50

released by different agencies, the values for the earthquake will be different. This also increases the difficulty for entity resolution since these records have a similar magnitude but different latitude, longitude and time in different records.

Taking Table 1 as an example, the values of the attributes are almost all numeric data, such as longitude, latitude and magnitude. However, for the same earthquake event, these records all have different values of attributes.

We outline a few important attribute-level conflicts below:

1. When a relatively large earthquake occurs, many small aftershocks occur in a very short period. The time and place of these aftershocks are very similar. Of course, in the same period, the earthquake will also occur in various parts of the world. It is very difficult for the time attribute to determine whether it is the same earthquake event. However, it does not affect the use of similarity in time to divide it into one that may be the same blocking module.
2. As mentioned earlier in this article, these data were collected from different agencies. For the same earthquake event, the value of the earthquake-related attributes issued may be different due to different agencies. For example, in the Wenchuan earthquake, the USGS reported an 8.1-magnitude earthquake, and the CNEC reported an 8.0-magnitude earthquake. There will be a certain amount of error.
3. For attribute values of an earthquake event, the standards used by different agencies are also different. For example, for the time attribute, some use UTC time, and some use Beijing time.

Methods

In this section, we will describe the five methods that were mentioned above. They are Euclidean distance, edit distance, Jaccard distance, cosine similarity and comprehensive approach. First, we will explain the process of the ER model.

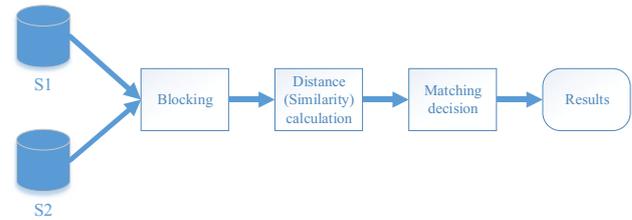


Fig. 1 The framework for entity resolution includes three main modules: blocking, distance (similarity) calculation and module matching decision

The ER model used in this paper is shown in Fig. 1. It mainly includes three modules: blocking modules, distance (similarity) calculation modules and matching decision modules.

- **Blocking module**

If two datasets, A and B, are to be directly matched, each record from A and B will be compared, resulting in a large number of $n_A \times n_B$ between two records (n_A, n_B denoting the number of records in the database). However, the expense of comparing the records' field values creates a bottleneck in a record linkage or deduplication system (Steorts et al. 2014; Christen and Goiser 2007).

To reduce the large number of potential comparisons, the entity resolution model uses blocking technology. Traditional blocking has been used by I. P. Fellegi et al. in record linkage since the 1960s (Fellegi and Sunter 1969). All records that have the same blocking key values (BKVs) are inserted into the same block, and then, only the records within the same block are compared with each other. Each record is inserted into one block only. This can reduce the number of useless object comparisons and calculations. In this paper, blocks can be made based on the time of the earthquake to reduce the number of record comparisons.

- **Distance (similarity) calculation module**

Using the distance (similarity) function to compare two candidate records, the smaller the distance (the greater the

degree of similarity) is, the more likely it is that two record objects correspond to the same entity; otherwise, the more likely it is that they are not the same entity. Currently, there are many text similarity functions, and the attributes of data tuples in this paper are mostly numeric data. How to design an appropriate similarity function will have a great influence on the recognition result.

- Matching decision module

The module calls the judgment function to analyze the similarity of the record pair to determine whether the record pair matches. The judgment function may be a threshold value. For a value greater than or equal to the threshold value, it indicates a match; otherwise, it does not match. The judgment function may also be obtained by training some functions and corresponding parameters and then using these functions and parameters to determine whether the new data match. Different methods use different thresholds, such as edit distance similar thresholds may be 2, 3, 4, 5, etc. The edit distance of two entities is greater than threshold. The two compared entities may be two different entities. For Euclidean distances, the threshold may be 0.95. Then, greater than 0.95 is judged to be the same entity.

The research content of this paper is numeric data, including various measured data, latitude, longitude, time and magnitude. The data are measured by each agency itself, even for the same event. Because the published values are different, the previous entity identification technology will be difficult to identify. This is also one of the challenges.

Therefore, in this paper, we will use the following two strategies to experiment:

- (1) Treat numbers as strings using string-based data methods, such as edit distance.
- (2) Numbers are measured according to their specific semantics to get distance metrics, such as Euclidean distance. We will use the comprehensive approach, which contains both edit distance and Euclidean distance.

Euclidean distance

The Euclidean distance between points a and b is the length of the line segment connecting them (ab).

In Cartesian coordinates, if $a = (a_1, a_2, \dots, a_n)$ and $b = (b_1, b_2, \dots, b_n)$ are two points in Euclidean n -space, then the distance (d) from a to b is given by the Pythagorean formula:

$$d(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}.$$

Edit distance

The edit distance between two strings $S1$ and $S2$ is the minimum number of operations on individual characters required to transform string $S1$ into $S2$. There are three types of edit operations:

Insert a character into the string; if $S1 = ab$, $S2 = acb$, then insert the character 'c' into $S1$, so $S1 \rightarrow S2$

Substitute one character with a different character for character changes from abd to acd ($b \rightarrow c$).

Delete a character from the string, changing acd to ad by deleting character 'c'.

Each above edit operation has a cost of 1. This version of edit distance is also known as the Levenshtein distance (Gomaa and Fahmy 2013). Ristad and Yianilos (1998) use a method to automatically determine the costs from a set of equivalent words written in different ways. Editing distance metrics is very effective for capturing typographical errors but is generally not valid for other types of mismatches.

Jaccard distance

The Jaccard distance, which measures dissimilarity between two sets, A and B , is complementary to the Jaccard coefficient and is obtained by subtracting the Jaccard coefficient from 1, or, equivalently, by dividing the difference of the sizes of the union and the intersection of two sets by the size of the union:

$$d_J(A, B) = 1 - J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

where $J(A, B)$ denotes the Jaccard similarity coefficient:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

Cosine similarity

Cosine similarity is a measure of similarity between two nonzero vectors of an inner product space that measures the cosine of the angle between them:

$$\text{Cos}(A, B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

where A_i and B_i are components of vectors A and B , respectively.

Comprehensive approach

Definition 1 (*Pairwise-entity resolution*) Given two data sources A and B of records and a match function F , pairwise-entity resolution is the process of identifying a set of record

pairs $(a,b) \in A \times B$ such that $\{(a,b) \mid a \in A, b \in B, F(a,b) \leq \tau\}$; τ is a threshold value, and thus, we can say that a and b are the same real-world entity. Depending on the domain knowledge, F can be different. If the attribute of the records is a string, we can use edit distance or Smith–Waterman distance. However, if the attribute is numeric, we can employ Euclidean distance or Manhattan distance. In other words, F contains a set of distance functions. $F = \{f_1, \dots, f_n\}$.

$$F(a, b) = \sum W_{f_i} \times f_i(a, b)$$

where f_i is a distance function in F and W_{f_i} is a specific weight indicating the importance of $f_i(a, b)$.

Since records contain multiple attributes, each attribute contributes differently to the record similarity. Therefore, it is very important to consider the weight of the attribute contribution to the similarity, which can significantly improve the accuracy of the record's similarity calculation. In this paper, the weight value is determined according to the knowledge and experience of domain experts. In this method, the weight values of latitude–longitude, time and magnitude are 0.5, 0.3 and 0.2, respectively.

Experimental evaluation

To evaluate the effectiveness of the different methods we mentioned in the previous sections, we use real-world datasets by Java. These experiments are performed on a computer having the following specifications: 2.66 GHz Intel Quad-Core Processor W3520, 4 GB of RAM and 120 GB hard disk. The operating system used is Windows 10.

Test datasets

All records used to test these methods were gathered from websites. These data were real-world data. The first real-world resources were generated by the USGS Earthquake Hazards Program, which provides information about real-time earthquakes, an online catalog search of its archives, earthquake maps and statistics. The second resources were generated by CENC, which collects monitoring data, processing and services. The third resources were generated by GEOFON, which is the largest earthquake data center in Europe supported by Germany (Fan 2016). Table 2 shows the basic information about our experimental data. As shown in Table 3, we can see the number of true matches. We use USGS, CENC and GEOFON to denote the data from websites of USGS, CENC and GEOFON, respectively. We use D_i to denote the comparison between resources USGS and GEOFON, USGS and CENC, CENC and GEOFON, respectively.

Table 2 The basic information of different data sources used in this paper. In column three, the number of records of each data source is listed

Year	Source name	Number of records
2017	USGS	125,188
	GEOFON	4708
	CENC	806
2016	USGS	119,686
	GEOFON	5218
	CENC	902
2015	USGS	120,597
	GEOFON	4353
	CENC	943

Table 3 The number of true matched and the short name for different datasets used in the evaluation

Year	Dataset	Dataset name	Number of true matched
2017	D ₁	CENC-GEOFON	215
	D ₂	USGS-CENC	246
	D ₃	USGS-GEOFON	4406
2016	D ₄	CENC-GEOFON	304
	D ₅	USGS-CENC	511
	D ₆	USGS-GEOFON	4875
2015	D ₇	CENC-GEOFON	266
	D ₈	USGS-CENC	487
	D ₉	USGS-GEOFON	4239

Experimental setting

Before the records were matched, as the data came from different agencies, there were some problems. One of them was the unit of the attribute. For example, for the depth, some records used the unit of km, and some used m. In the time fields, some records used the time of UTC + 8, and some used UTC time.

By analyzing the data, we find that in fact most of the data in the sources are not relevant and do not need to be compared. Therefore, the related data records are allocated in one block, separated from the unrelated data records. Blocking techniques can generally be divided into two categories: one is the partitioning technique where there is no intersection between the blocks; the other is the partitioning technique where there is intersection between the blocks.

The blocking model mentioned in the previous section can reduce the large number of potential comparisons. In (Christen 2011), some blocking methods were presented, such as traditional blocking, sorted neighborhood and Q-gram. In this paper, we employed two blocking categories.

Both use time of the records attributes. In this paper, first method, we employed the traditional blocking method to create blocking key values (BKVs) by using time, one of record attributes. In this method, there is no intersection between two blocks. Secondly, we used time-slid-widow (Time-SW) method so that there would be some intersection records between two blocks.

Quality and complexity measures

To evaluate the accuracy of the similarity entity resolution, we used well-known measures, namely precision (pr), recall (re) and F1 score. We used n_m to denote the total number of matched record pairs and n_A and n_B to denote the size of datasets A and B . Then, we used tp (true positive) and fp (false positive) to denote the numbers of true matched, false matched and record pairs generated by our methods. We used precision (pr) to denote the percentage of similar records among the records that have a similarity score above the threshold θ . The pr, re and F1 may be expressed as:

$$\text{pr} = \frac{\text{tp}}{\text{tp} + \text{fp}}$$

$$\text{re} = \frac{\text{tp}}{n_m}$$

$$\text{F1} = 2 * \frac{\text{pr} * \text{re}}{\text{pr} + \text{re}}$$

In this paper, we used five methods to resolve the problem of the real-world earthquake events data entity resolution. The first is edit distance, in which we make all attributes into a string and then obtain the edit distance of two records. The second is Euclidean distance the same as cosine similarity, in which we treat the attributes as a three-dimensional vector to obtain the Euclidean distance or cosine similarity between two records. The third is Jaccard distance, and the fourth is cosine similarity. The fifth method is a comprehensive approach, which uses a different distance function for different attributes. For the attribute time, we can exploit the string edit distance. For the location attribute, we can make use of Euclidean distance. For the magnitude attribute, we use the difference based on the subtraction of two records.

Results and discussion

We report on the accuracy of our methods with pr, re and F1. The pr and re are shown in Fig. 2. The F1 is shown in Fig. 3. For D_3 , the CoA method achieved the pr value above 0.99, the re value above 0.94 and even F1 value above 0.96. These values were almost the highest of these

methods. For all datasets, the value of Jaccard and cosine was very low by using BKVs technology, and however, they became very high by using Time-SW. This means that the appropriate blocking technique can avoid unnecessary false and improve the effect.

It was observed that the CoA method showed higher accuracy compared to the other methods because the different attributes had different weights for the distance measure of the two records. In other words, different attributes had different importance for the entity resolution of earthquake events.

Surprisingly, the accuracy of EdD was also higher. That is, treating numbers as a string for entity resolution can also achieve higher accuracy.

In the experiment, we found that to obtain a higher value, the values of pr and F1 sometimes become very low. This happens when D_2 uses cosine and Jaccard method. The value of re was more than 0.6, but the values of both pr and F1 were very low. Overall, the values obtained by the cosine and Jaccard method were all very low. This also shows that in this experiment, using cosine similarity and Jaccard distance for numeric data directly in record matching did not achieve a very good effect.

Table 4 shows that without blocking the number of comparisons would be super huge. So how to reduce the number of comparisons is always very important thing for ER.

We compare the values of the F1 scores of different methods. These experiments show that CoA produces good results with both blocking methods. No matter which kind of blocking technology is used in EdD and CoA method, there will be similar effects. Jaccard distance does not perform very well, no matter what kind of blocking technology. It can also be said that Jaccard distance is not suitable for numerical data ER in our experiments. Through our experiments, we can see that Time-SW has better effect. The result of this is because of the successful blocking, reducing many unnecessary candidate records. Form Fig. 3 D_1 , D_2 , D_3 , D_6 , D_9 , these show that the EuD method also archives a better performance.

Conclusions

When an earthquake disaster occurs, the same earthquake information can be described or reported by multiple data sources. To manage these data more efficiently, for the first time, the application of entity resolution technology was used in earthquake disasters for data integration. This paper studied three years of earthquake events from three sources by using five methods. The results indicated the following conclusions.

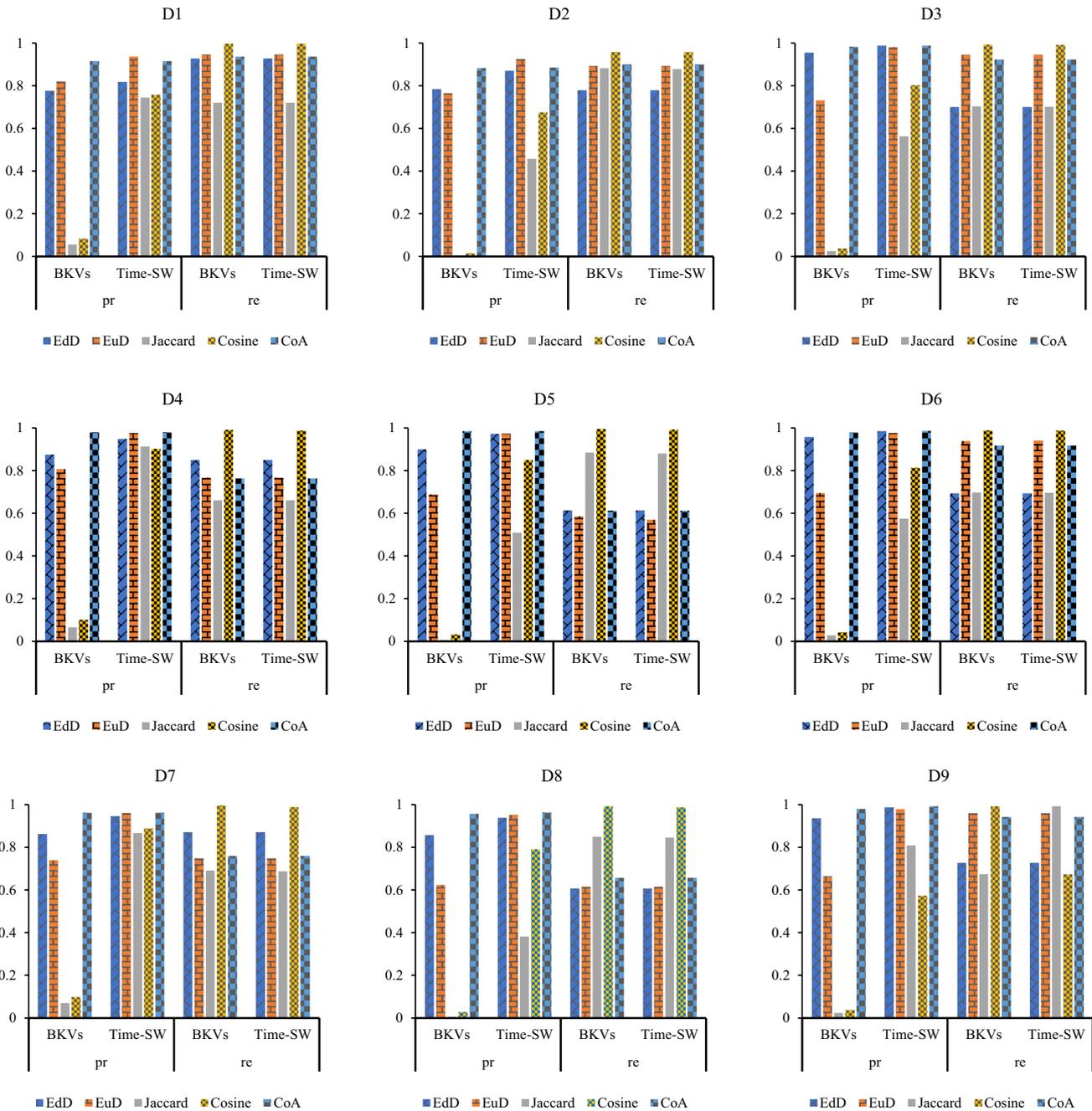


Fig. 2 The different results of precision and recall for different datasets by different methods, respectively

- (1) For these datasets, the CoA method achieved the highest performance. In other words, different attributes had different importance for record identification.
- (2) Directly using the Jaccard distance and cosine similarity for numerical data did not result in high performance.
- (3) Treating numbers as a string using the EdD method also achieved better performance. Numeric-based data and

- string-based data must have something in common to facilitate record identification.
- (4) Accurate blocking technology can improve efficiency and reduce comparison times.

Due to the different information published by different departments or agencies, whether it is historical disaster data management or real-time disaster data management,

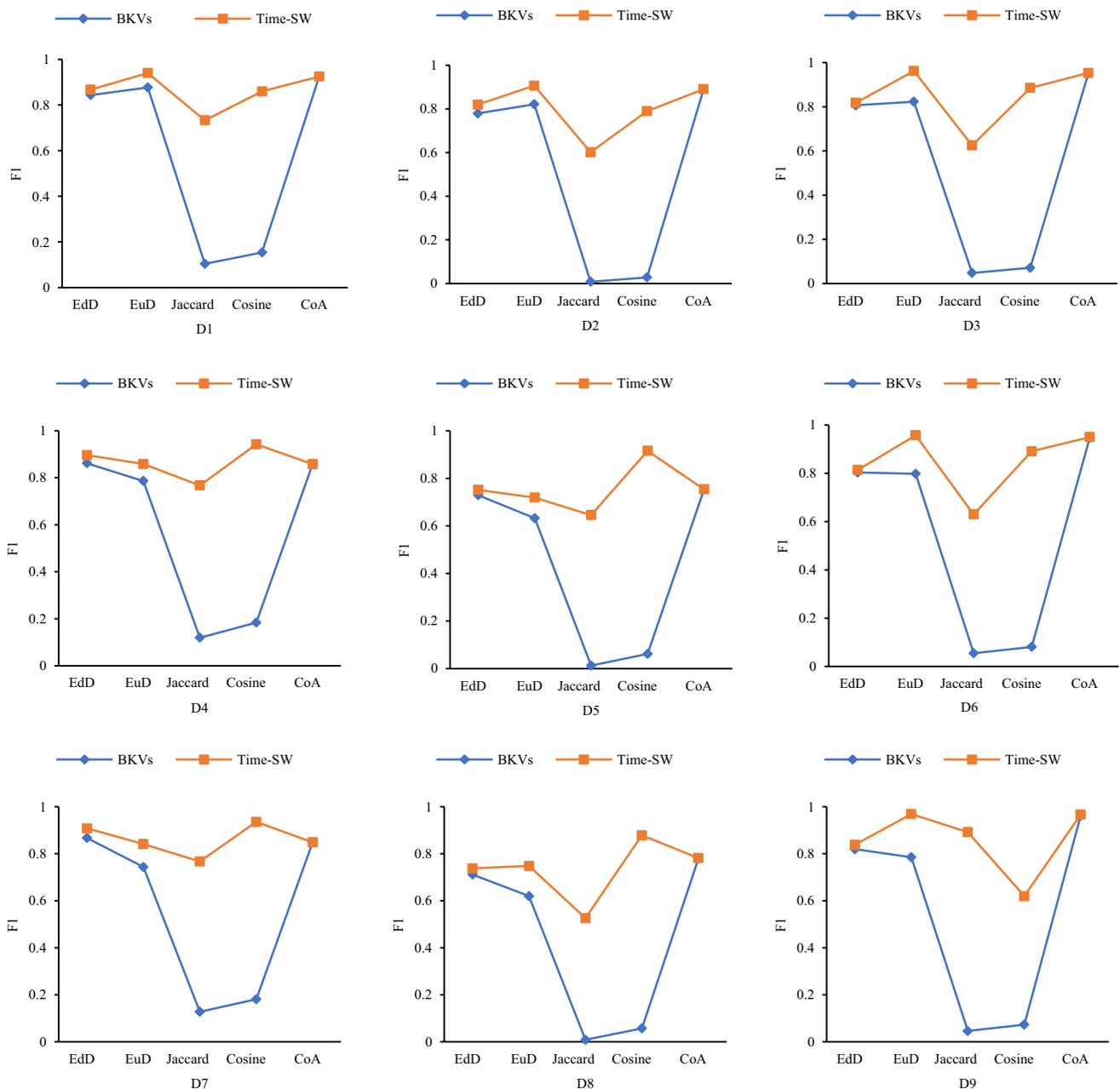


Fig. 3 F1 scores for different methods of different datasets. In general, F1 of Time-SW is higher than BKVs based on the same method

it will be helpful to integrate these data as much as possible to make more accurate decisions for disaster reduction (Geller 2007). By identifying disaster events from different data sources, disaster data can be better managed, including remote sensing data, economic data and loss data, to better cope with disasters. Through data management, you can find more and better data research disasters. It is very difficult to understand and find earthquake events from different data sources, and an abstraction and simplification based on these data are needed. Our research aims to make

these data easy for decision makers and the public to use, without having to invest a lot of time and effort.

In this paper, we used five methods for the challenging problem of entity resolution from real-world data that came from earthquake disaster websites. The attributes of earthquakes are almost entirely numeric data. This paper used a framework to solve the entity resolution problem. We studied three years of earthquake disaster events from three sources. However, the data from earthquake disaster events are massive and constantly updated. With the improvement

Table 4 The number of comparisons for each dataset by using different methods

Dataset	BKVs	Time-SW	Without blocking
2017			
D ₁	11,020	301	3794648
D ₂	283,705	1155	100901528
D ₃	1,634,608	8898	589385104
2016			
D ₄	13,601	349	4625456
D ₅	299,133	1399	107956772
D ₆	1,707,869	9582	624521548
2015			
D ₇	12992	313	4104879
D ₈	313106	1348	113722971
D ₉	1469716	8270	524958741

in data acquisition ability over time, it will become more and more difficult to manage the massive data from earthquake disaster events, so that it can be foreseen that more efficient ER technology will emerge, as well as new methods and models that will be proposed in the future.

Acknowledgements The authors thank the anonymous referees for their valuable comments and suggestions, which improved the technical content and the presentation of the article. This research was supported by the National Key Research and Development Program of China (2016YFB0501504).

References

- Ayat N, Afsarmanesh H, Akbarinia R, Valduriez P (2012) An uncertain data integration system. In: On the Move to meaningful internet systems: Otm
- Ayat N, Akbarinia R, Afsarmanesh H, Valduriez P (2014) Entity resolution for probabilistic data. *Inf Sci* 277:492–511
- Baeza-Yates R, Gonnet GH (1992) A new approach to text searching. *Commun ACM* 35(10):74–82
- Boyer RS, Moore JS (1977) A fast string searching algorithm. *Commun ACM* 20(10):762–772
- Chang WI, Lampe J (1992) Theoretical and empirical comparisons of approximate string matching algorithms. In: Combinatorial pattern matching, third annual symposium, CPM 92, Tucson, Arizona, USA, April 29–May 1, 1992, Proceedings. Springer
- Christen P (2011) A survey of indexing techniques for scalable record linkage and deduplication. *IEEE Trans Knowl Data Eng* 24(9):1537–1555
- Christen P, Goiser K (2007) Quality and complexity measures for data linkage and deduplication. *Complexity* 43:127–151
- Clark DE (2004) Practical introduction to record linkage for injury research. *Injury Prev* 10(3):186–191
- Du MW, Chang SC (1994) An approach to designing very fast approximate string matching algorithms. *IEEE Trans Knowl Data Eng* 6(4):620–633
- Elmagarmid AK, Ipeirotis PG, Verykios VS (2006) Duplicate record detection: a survey. *IEEE Trans Knowl Data Eng* 19(1):1–16
- Fan X (2016) GEOFON data center. *Recent Dev World Seismol* 452(8):33–41
- Fellegi IP, Sunter AB (1969) A theory for record linkage. *J Am Stat Assoc* 64(328):1183–1210
- Galil Z, Giancarlo R (1988) Data structures and algorithms for approximate string matching. *J Complex* 4(1):33–72
- Geller RJ (2007) Earthquake prediction: a critical review. *Geophys J Int* 131(3):425–450
- Gomaa WH, Fahmy AA (2013) A survey of text similarity approaches. *Int J Comput Appl* 68(13):13–18
- Hassanzadeh O, Chiang F, Lee HC, Miller RJ (2009) Framework for evaluating clustering algorithms in duplicate detection. *Proc VLDB Endow* 2(1):1282–1293
- Jaro MA (1980) UNIMATCH, a record linkage system: users manual. Bureau of the Census
- Kelman CW, Bass AJ, Holman CDJ (2010) Research use of linked health data — a best practice protocol. *Aust N Z J Publ Health* 26(2):251–255
- Khan B, Rauf A, Shah SH, Khusro S (2011) Identification and removal of duplicated records. *World Appl Sci J* 13(5):1178–1184
- Knuth DE, Morris JH Jr, Pratt VR (1977) Fast pattern matching in strings. *SIAM J Comput* 6(2):323–350
- Köpcke H, Thor A, Rahm E (2010) Evaluation of entity resolution approaches on real-world match problems. *Proc VLDB Endow* 3(1–2):484–493
- Koudas N, Marathe A, Srivastava D (2004) Flexible string matching against large databases in practice. In: Thirtieth international conference on very large data bases
- Lee S, Lee J, Hwang SW (2014) Efficient entity matching using materialized lists. *Inf Sci* 261:170–184
- Levenshtein VI (1966) Binary codes capable of correcting deletions, insertions, and reversals. In: *Soviet physics doklady*, vol 10, No 8, pp 707–710
- Li L, Li J, Gao H (2015) Rule-based method for entity resolution. *IEEE Trans Knowl Data Eng* 27(1):250–263
- Magnani M, Montesi D (2010) A survey on uncertainty management in data integration. *J Data Inf Qual* 2(1):1–33
- Miller FP, Vandome AF, MCBrewster J (1980) Approximate string matching. *ACM Comput Surv* 12(4):381–402
- Monge AE (2000) Matching algorithms within a duplicate detection system. *IEEE Data Eng Bull* 23(4):14–20
- Navarro G (2001) A guided tour to approximate string matching. *ACM Comput Surv* 33(1):31–88
- Peterson JL (1980) Computer programs for detecting and correcting spelling errors. *Commun ACM* 23(12):676–687
- Pinheiro JC, Sun DX (1998). Methods for linking and mining massive heterogeneous databases. In: Proceedings of the fourth international conference on knowledge discovery and data mining, August. AAAI Press, pp 309–313
- Ristad ES, Yianilos PN (1998) Learning string-edit distance. *IEEE Trans Pattern Anal Mach Intell* 20(5):522–532
- Steorts RC, Ventura SL, Sadinle M, Fienberg SE (2014) A comparison of blocking methods for record linkage. In: International conference on privacy in statistical databases. Springer, Cham
- Sun CC, Shen DR, Kou Y, Nie TZ, Yu G (2016) Entity resolution oriented clustering algorithm. *J Softw* 27(9):2303–2319 (**in Chinese**)
- Sutinen E, Tarhio J (1995) On using q-gram locations in approximate string matching. In: Algorithms-esa 95, third European symposium, Corfu, Greece, September. DBLP
- Ukkonen E (1992) Approximate string-matching with q-grams and maximal matches. *Theor Comput Sci* 92(1):191–211
- Waterman MS, Smith TF, Beyer WA (1976) Some biological sequence metrics. *Adv Math* 20(3):367–387

- Winkler WE (2004) Methods for evaluating and creating data quality. *Inf Syst* 29(7):531–550
- Winkler WE (2006) Overview of record linkage and current research directions. In: Bureau of the Census
- Zhu B, Suo M, Chen Y, Zhang Z, Li S (2018) Mixed H ∞ and passivity control for a class of stochastic nonlinear sampled-data systems. *J Frankl Inst* 355(7):3310–3329



Full waveform inversion based on a local traveltimes correction and zero-mean cross-correlation-based misfit function

Shiqi Dong¹ · Ligu Han¹ · Yong Hu¹ · Yuchen Yin¹

Received: 18 July 2019 / Accepted: 20 November 2019 / Published online: 4 December 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

Full waveform inversion (FWI) suffers from the cycle skipping problem, because the observed data usually lack low-frequency components or due to errors in the wavelet estimation. In addition, the strong low-frequency non-zero-mean noise can have a large impact on FWI results. Thus, we propose a local waveform traveltimes correction scheme to solve the situations when the observed data lack low-frequency components or when the estimation for the wavelet is incorrect. We use a sliding time window, which is used to decrease the traveltimes differences between the calculated and observed data to increase the cross-correlation between them. Besides, we propose a zero-mean normalized cross-correlation misfit function to reduce the interference of the low-frequency non-zero-mean noise. Therefore, we propose new approaches to improve FWI results whether the observed data lack low-frequency components or the observed data are contaminated by the non-zero-mean low-frequency noise. Numerical examples on Marmousi model show the feasibility of a FWI based on the zero-mean normalized cross-correlation misfit function and a FWI based on the local traveltimes correction method.

Keywords FWI · Local traveltimes correction · Zero-mean normalized cross-correlation

Introduction

Full waveform inversion (FWI) uses all of the information from prestack seismic data to update the subsurface parameters (e.g., velocity, density, and anisotropic parameters) under the constraints of a misfit function. FWI was first implemented in the time domain (Lailly 1983; Tarantola 1984) based on the norm-2 misfit function by minimizing the difference between the observed and calculated data. Later, FWI was implemented in the frequency domain (Pratt et al. 1998), and only a few discrete frequencies were needed to update the physical property parameters of subsurface medium accurately. However, FWI suffers from the cycle skipping problem (Virieux and Operto 2009), which leads FWI into a local minimum (Alkhalifah 2016). A good initial model and sufficient low-frequency components in the observed data can help FWI avoid cycle skipping (Bunks et al. 1995). However, in practice, the ultra-low frequencies are hardly recorded or are contaminated with the

low-frequency noise. In addition, generating a good initial model is also difficult because of the complicated structure of the subsurface. Moreover, the inaccurate estimation for wavelet will result in the calculated data being different from the observed data in both the amplitude and phase information, which will also lead to a failure in waveform matching. To improve the matching between waveforms and solve the cycle skipping problem, many researchers have done a lot of work and their achievements can be mainly divided into three groups.

In the first group, artificial low frequencies, which are somewhat similar in phase to the low-frequency components of the intact observed data, are introduced into FWI. Bozdağ et al. (2011) proposed the instantaneous envelope misfit function to reduce the nonlinearity of inversion and demonstrated that the envelope-based misfit function has direct relationship with the long-wavelength components of velocity variation. Chi et al. (2014) utilized the differences between the envelopes of both observed and calculated data as a misfit function to provide the long-wavelength components of the subsurface velocity for the standard FWI. Wu et al. (2014) proved the physical meaning of the envelope-based method using the modulation signal model. Hu et al. (2017) improved the envelope-based theory and proposed a

✉ Ligu Han
hanliguo@jlu.edu.cn

¹ College of Geo-Exploration Science and Technology, Jilin University, Changchun, China

waveform mode decomposition approach, which reduces the dominant frequency of the adjoint source and obtain richer low-frequency information. Alkhalifah and Choi (2014) utilized Laplace damping to generate artificial low frequencies. Hu (2014) proposed a beat-tone approach, in which low frequencies can be produced artificially by the subtraction of slightly different frequency wavefields. Zhang et al. (2017) proposed that the low-frequency components can be reconstructed by the convolution of artificially designed low-frequency wavelets and the reflected impulse response of the subsurface medium obtained by sparse blind deconvolution. Lian et al. (2018) proposed using an improved wavefield decomposition method in the time–space domain to enhance low-wavenumber components of FWI. Yuan et al. (2019) combined FWI, which can provide long-wavelength components of the P-wave velocity by using low frequencies and long-offset direct/diving/refracted waves, and poststack impedance inversion to develop an FWI-driven impedance inversion, which has the ability to solve the overlap between the FWI-based converted prior impedance model and poststack data, and it can thereby yield a broadband absolute impedance result. Li et al. (2018) proposed that low-frequency components can be produced by multiplying wavefields between slightly different frequencies with each other, and improved upon the beat-tone method by multiplying wavefields with themselves and use a nonlinearly smoothed operator to extract the low-frequency components in the multiplied wavefields. Choi and Alkhalifah (2018) proposed that an exponentially damped wavefield, which contains abundant low-frequency information, and used the deconvolution-based misfit function to obtain the long-wavelength components of subsurface parameters.

The second group includes methods of providing more accurate initial models for FWI, such as traveltime tomography (Bishop et al. 1985; Justice et al. 1989) and migration velocity analysis (Symes and Carazzone 1991; Biondi and Symes 2004), etc. The traveltime or phase information is not as dramatic as the amplitude information, which is the main reason that causes cycle skipping (Luo and Schuster 1991). By extracting phase information using the logarithm of wavefields, an optimal algorithm minimizes the phase differences between the observed and calculated data to prevent cycle skipping (Shin and Min 2006). However, the phase information is always wrapped. Choi and Alkhalifah (2015) proposed a phase unwrapping method to improve the accuracy of FWI based on phase information.

In the third group, waveform-matching techniques are used for FWI to avoid cycle skipping. Warner and Guasch (2014) used the Wiener filter to increase the similarity between the observed and calculated data. Zhu and Fomel (2016) proposed adaptive matching filtering-based FWI, and Engquist et al. (2016) introduced the optimal transport method to FWI, which protects FWI from falling into a local

minimum. Hu et al. (2018) measured the misfit between the observed and calculated data with respect to phase differences without phase unwrapping, which solved the cycle skipping problem from the perspective of a mathematical equation. Yang et al. (2016) proposed the phase-correlation-shifting-based FWI in the frequency domain to increase the cross-correlation between waveforms. Sun and Alkhalifah (2019) proposed adaptive traveltime inversion, which is formulated by minimizing the least-squares errors of the traveltime shift computed from cross-correlation of a penalty function with a matching filter. In addition to these three groups of methods, some other methods like reflection FWI (Xu et al. 2012; Chi et al. 2015; Guo and Alkhalifah 2017), stochastic-algorithm-based FWI (Datta and Sen 2016; Sajeve et al. 2016), the angle difference identity for cosine, which can accurately recover the large-scale structure with high-frequency data (Wang et al. 2019), etc., can also help FWI avoid cycle skipping and obtain an ideal inverted result.

The estimation of uncertainty in the resulting earth models and parameter trade-offs are also as important as the inversion results. Thus, Liu et al. (2019) investigated the application of the square-root variable metric (SRVM) method, a quasi-Newton optimization algorithm, to FWI in a vector version, which allows us to reconstruct the final inverse Hessian at an affordable storage memory cost.

The norm-2 misfit function is the most commonly used misfit function for FWI. However, it is very sensitive to amplitude information and the influence of noise. Choi and Alkhalifah (2012) proposed a global-correlation misfit function, which helped the multisource FWI avoid amplitude imbalance because of the offset range, and Liu et al. (2016) noted that this misfit function improves noise immunity and decreases the influence of the wavelet estimation error. However, seismic data recorded in the field usually contain the strong non-zero-mean noise, and cross-correlation does not work well for this type of noise. In addition, many researchers have done a lot of work on reducing the influence of the wavelet on the inverted results. Source-independent approaches are an effective method for doing so (Lee and Kim 2003; Zhou and Greenhalgh 2003; Choi and Alkhalifah 2011). Moreover, Pan and Huang (2018) proposed FWI based on differential measurements of long-offset data to reduce the influence of inaccurate source signatures.

In this paper, inspired by Yang et al. (2016), we propose a waveform local traveltime correction approach based on the local cross-correlation to solve the cycle skipping problem in time-domain FWI. Many researchers have shown that using local cross-correlation to calculate the traveltime difference is effective (Hale 2006; Fomel 2007; Mikesell et al. 2015; Zhang et al. 2019). Yang achieved phase-correlation-shifting-based FWI in the frequency domain, in which each trace in the calculated data is time-shifted according to the maximal cross-correlation between each two corresponding traces

in the observed and calculated data. However, the traveltime difference between two traces is not a constant. Therefore, the traveltime difference is different for each time sample. Moreover, in time-domain data, direct waves and early arrivals account for most of the energy, which have a large impact on the cross-correlation-based traveltime calculation, and the waveforms of direct waves and early arrivals are generally well-matched, which causes the time shift to be negligible if we regard the traveltime difference between two traces as a constant. Thus, considering these factors, we use a sliding time window to intercept the observed and calculated data, firstly. Then we calculate the cross-correlation between the data in the window and find the exact time corresponding to its maximal value. Next, we time-shift the calculated data in the time window according to the time, which is obtained in the last step. Then, we move the window along the time axis to make the time shifting of the calculated data appear in the next time window until the window has covered all of the time samples, which means completing the time shifting for all of the calculated data such that the matching between the observed and calculated data can be increased. In addition, since the calculated data are time-shifted, the phase information of the calculated data is closer to that of the observed data, but the amplitude information is incorrect. Therefore, to decrease the influence of the amplitude error in the gradient, we use a normalized global-correlation misfit function (Choi and Alkhalifah 2012). In order to deal with the situation when the observed data contain the strong low-frequency non-zero-mean noise, we propose a zero-mean normalized cross-correlation misfit function. The calculation of the zero-mean (subtracting the average value) helps FWI to decrease its sensitivity to the non-zero-mean low-frequency noise, which is usually recorded in the observed data. Thus, our method can improve the FWI results in situations when the observed data lack of low-frequency components or the inaccurate estimation of the wavelet or the observed data are contaminated with the non-zero-mean low-frequency noise. Finally, we use numerical tests of the Marmousi model to verify the feasibility of our method.

Theory

Review of FWI with norm-2 misfit function

The norm-2 misfit function aims to minimize the difference between the observed and calculated data, and it is expressed as

$$E(v) = \frac{1}{2} \sum_r \int_t [d_{\text{cal}}(r, t, v) - d_{\text{obs}}(r, t)]^2 dt, \quad (1)$$

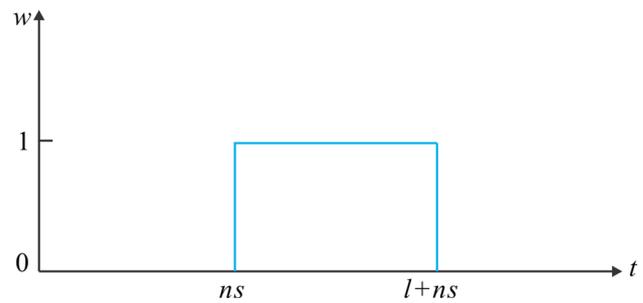


Fig. 1 The rectangular time window

where d_{cal} and d_{obs} represent the calculated and observed data, respectively, r is the receiver position variable, t represents the time variable, and v denotes the P-wave velocity of the subsurface medium. The calculated data can be obtained by a finite-difference scheme based on an acoustic wave equation, which can be expressed in 2D as

$$\frac{\partial^2 u_f}{\partial x^2} + \frac{\partial^2 u_f}{\partial z^2} - \frac{1}{v^2} \frac{\partial^2 u_f}{\partial t^2} = s, \quad (2)$$

where x and z are variables indicating the horizontal distance and the depth, respectively, s is the source, and u_f is the forward-propagated wavefield. The gradient of this misfit function can be expressed as

$$\frac{\partial E(v)}{\partial v} = \sum_r \int_t \left(\frac{\partial d_{\text{cal}}(r, t, v)}{\partial v} \right)^T [d_{\text{cal}}(r, t, v) - d_{\text{obs}}(r, t)] dt, \quad (3)$$

where $\partial d_{\text{cal}}(r, t, v)/\partial v$ is the partial derivative wavefield and T represents the transposition. Calculating the partial derivative wavefield directly is highly time consuming. Therefore, we use the adjoint state method (Tarantola 1984) to obtain the gradient. The adjoint source is expressed as: $d_{\text{cal}}(r, t) - d_{\text{obs}}(r, t)$, and by calculating the zero-lag cross-correlation between the forward-propagated and back-propagated wavefield, we can obtain the gradient:

$$\frac{\partial E(v)}{\partial v} = -\frac{2}{v^3} \sum_x \int_t \frac{\partial^2 u_f}{\partial t^2} u_b dt, \quad (4)$$

where u_b is the adjoint wavefield.

Waveform local traveltime correction

Using an inaccurate initial model for FWI results in a mismatch in the phase or traveltime information between the observed and calculated data. In this case, the calculated and observed data cannot be matched well, which will lead to cycle skipping. To avoid cycle skipping, we use a sliding time window to decrease the traveltime difference between

each two corresponding traces in the observed and calculated data. We use a rectangular time window (Fig. 1):

$$w(t, n) = \begin{cases} 1, & ns \leq t \leq l + ns, \quad n = 0, 1, 2, \dots, N, \\ 0, & \text{else,} \end{cases} \quad (5)$$

where l is the length of the time window, s is the time step of the window shift, n is the number of shifts, and N is the maximum number of shifts. Three steps in total to complete the local traveltimes correction:

Step 1 The time window starts at the beginning of the time axis and calculates the cross-correlation between the observed and calculated data in the window:

$$C_{rw}(\tau) = \frac{1}{l} \sum_{t=ns}^{l+ns} [w(t, n)d_{\text{obs}}(r, t)][w(t, n)d_{\text{cal}}(r, t + \tau, v)], \quad (6)$$

$$1 - l \leq \tau \leq l - 1,$$

where C_{rw} is the cross-correlation of the observed and calculated data in the same time window.

Step 2 Take the τ value at the maximum of C_{rw} to time-shift the calculated data in the window:

$$d_{\text{cal}}^{w*}(r, t, v, n) = d_{\text{cal}}^w(r, t + \tau, v, n), \quad (7)$$

where d_{cal}^w is the calculated data in the time window and d_{cal}^{w*} is the calculated data after the traveltimes correction in the window.

Step 3 Shift the time window along the time axis and time-shift the calculated data in the later windows until the window contains the last time sample of the trace, which implies that the finish of the local traveltimes correction for the whole trace:

$$d_{\text{cal}}^*(r, t, v) = d_{\text{cal}}^{w*}(r, t, v, n), \quad n \quad \text{from } 0 \text{ to } N \text{ in sequence,} \quad (8)$$

where $d_{\text{cal}}^*(r, t, v)$ is the calculated data after the traveltimes correction. Therefore, the traveltimes differences between the observed and calculated data can be reduced.

We use an example to demonstrate the local traveltimes correction method. Figure 2a shows one pair of traces from the observed and calculated data. The positions of ①–⑥ show that the traveltimes differences of the crests and troughs between the calculated and observed data are all larger than half a cycle, which causes cycle skipping. Then we use a rectangular time window to intercept the observed and calculated data from the beginning of the time axis when $l = 200$ ms (Fig. 2b), and Fig. 2c shows the waveforms of the calculated and observed data in the first time window. Then, we calculate the cross-correlation between the waveforms shown in Fig. 2c and obtain the value of τ ($\tau = -34$ ms) when C_{rw} is maximal (Fig. 2d), which means the calculated data in the time window need a time shift of 34 time samples along the negative direction of the time axis to reach

the maximal value of C_{rw} . After time shifting, as shown in Fig. 2e, the time differences in the positions of ①–⑤ are less than half a cycle, and the 34 time samples at the end of the time axis in the calculated data within the time window are set to zero (position ⑥). Similarly, if the calculated data are time-shifted along the positive direction of the time axis, then after time shifting, the time samples at the front end of the time axis will be set to zero. The time samples of the calculated data, which are set to zero, will have tiny contribution to the gradient in the global-correlation-based misfit function (Choi and Alkhalifah 2012). Next, the time window moves with one step length ($s = 100$ ms) along the positive direction of the time axis, and then time-shift the calculated data in the second window (Fig. 2f). Then, we continually loop over all these steps until the time window contains the last time sample in the time axis.

We use an example to better illustrate our method. Figure 3a, b shows the blended-source observed and the calculated data, respectively. The blended source contains 9 shots. Each grid point on the surface is a receiver. A ricker wavelet with a peak frequency of 20 Hz is used as the source (Fig. 4), and the total recording time is 2000 ms. Figure 5a shows the traveltimes differences between the observed and calculated data when $l = 50$ ms and $s = 20$ ms, and it indicates that except for direct waves near the sources, there are significant differences in the traveltimes and the traveltimes differences at different times are not the same. Therefore, regarding the traveltimes difference of each trace between the observed and calculated data as a constant is inaccurate, which is not helpful in reducing the cycle skipping problem. Moreover, Fig. 5b shows that the values of the normalized cross-correlation among the observed and calculated data for many traces are small, which implies that the waveforms of the observed and calculated data do not align well. However, after a traveltimes correction ($l = 50$ ms, $s = 20$ ms), the traveltimes differences between different times of the data are reduced, and only some times of the traveltimes differences are large (Fig. 6a). In addition, the values of the cross-correlation between the observed and calculated data have increased significantly (Fig. 6b). Therefore, this numerical test demonstrates the effectiveness of our method, which is helpful for reducing the traveltimes differences between the observed and calculated data.

Next, we discuss the criterion for an optimal time window length (l) and the time step (s). Firstly, the choice of the time step should be smaller than or equal to the window length in order to ensure that the traveltimes differences of all time samples are corrected. Then, we compare the influence of different window lengths. By comparing Figs. 5a and 7a–e, we demonstrate that a short time window behaves better than a long one in terms of the accuracy of the traveltimes differences between the observed and calculated data. However, if we set l to be too small, then it will be meaningless

Fig. 2 Illustrations of the waveform local traveltime correction process: **a** one pair of traces of observed data and calculated data; **b** intercepting the observed and calculated data with a time window ($l=200$ ms), the dotted line represents the rectangular time window; **c** the waveforms within the first time window; **d** the cross-correlation between the observed and calculated data shown in **c**, and the τ corresponding to its maximum value; **e** the calculated data after local traveltime correction within the first time window and the observed data within the first time window; **f** the calculated data after the first local traveltime correction and the original observed data, then the time window moves one time step along the time axis ($s=100$ ms)

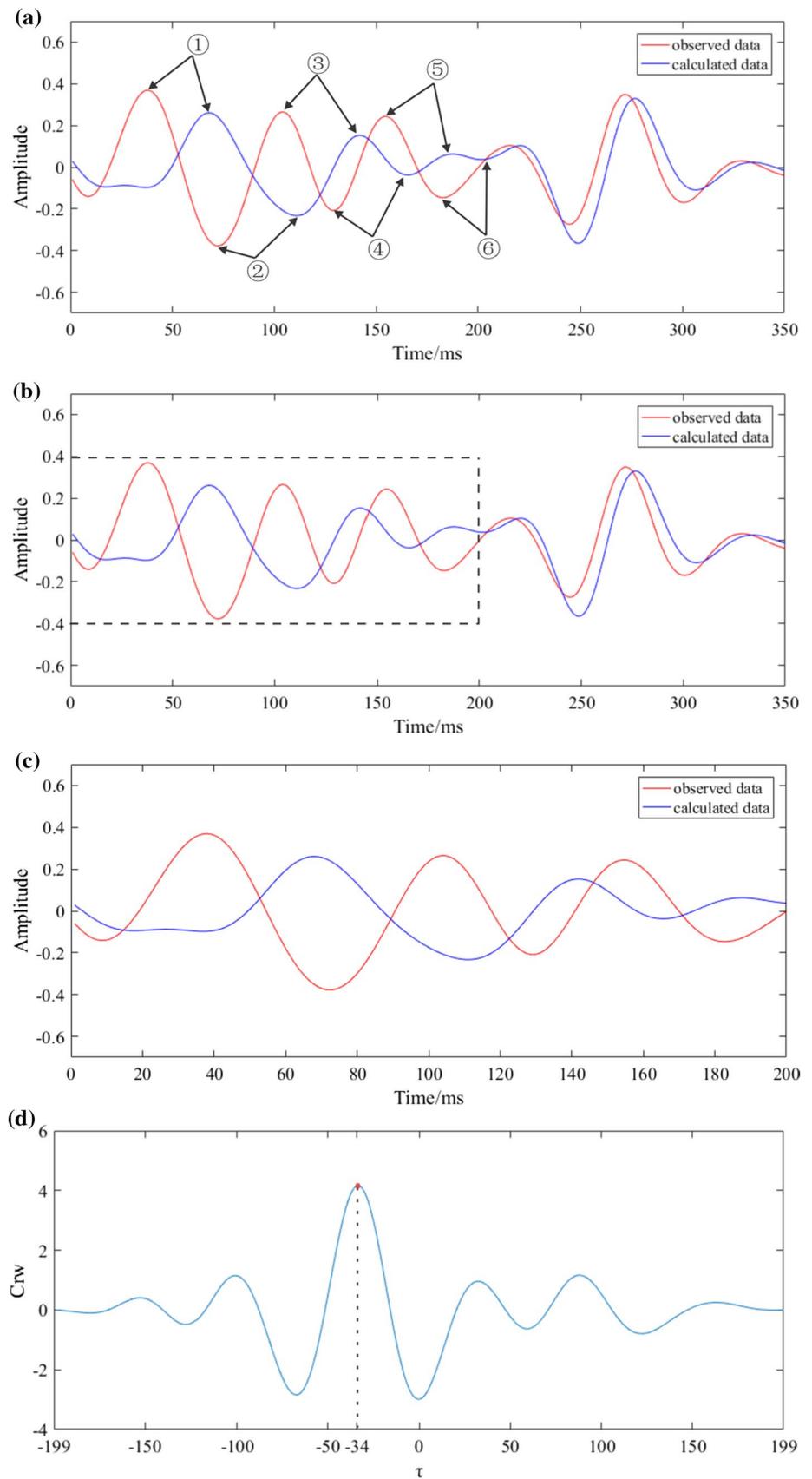


Fig. 2 (continued)

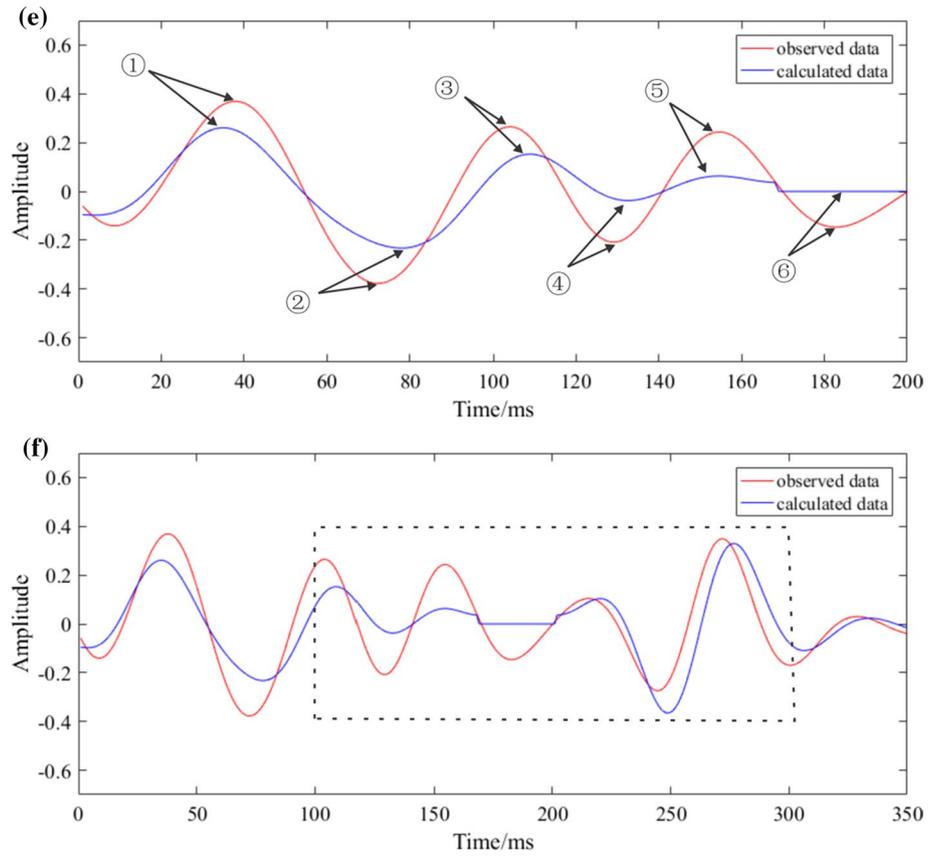
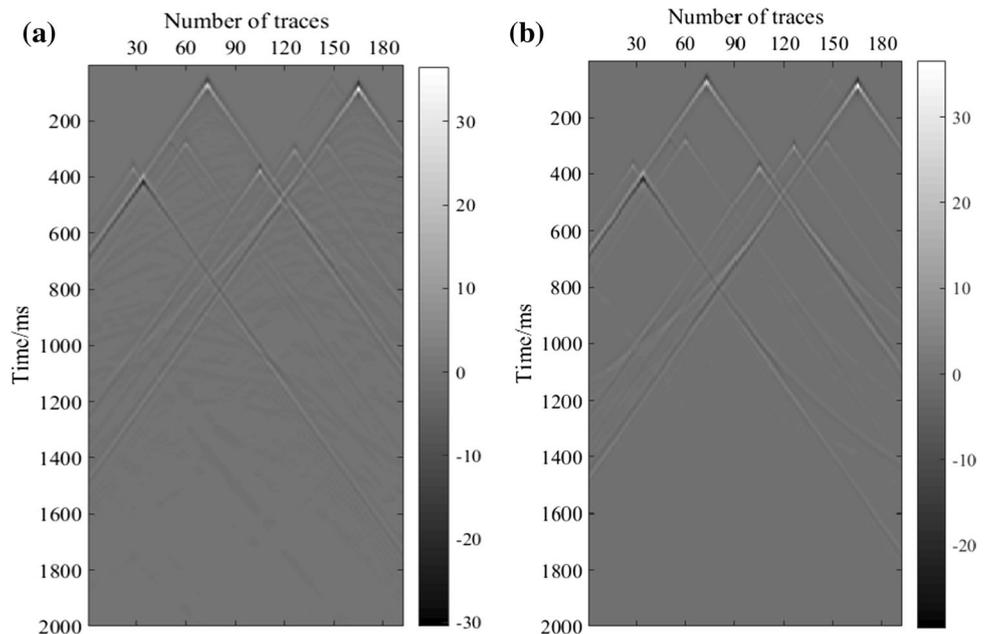


Fig. 3 Seismic data with 9 blended sources of **a** observed data; and **b** calculated data



because the traveltime differences between the data at each corresponding time sample will be negligible and will not reveal the kinematic characteristics of the wave propagation. However, if we use a long time window, the traveltime

differences at many times between the observed and calculated data will be ignored, and the time shifting will only focus on the events with strong energy. In our tests, the effects of the time window length with 50 ms, 30 ms and

Fig. 4 A ricker wavelet with a peak frequency of 20 Hz

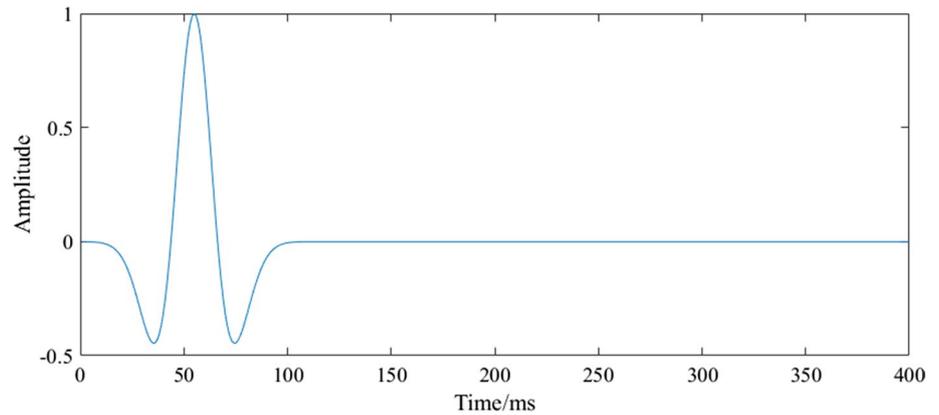
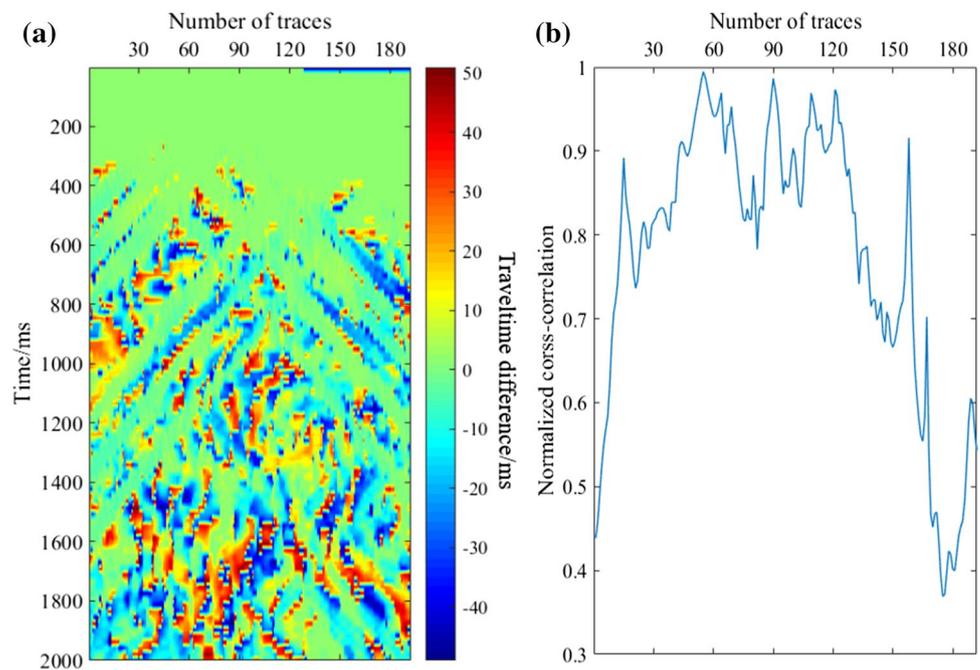


Fig. 5 **a** Traveltime differences between the observed and calculated data ($l=50$ ms, $s=20$ ms) and **b** the normalized cross-correlation between the observed and calculated data ($l=50$ ms, $s=20$ ms)

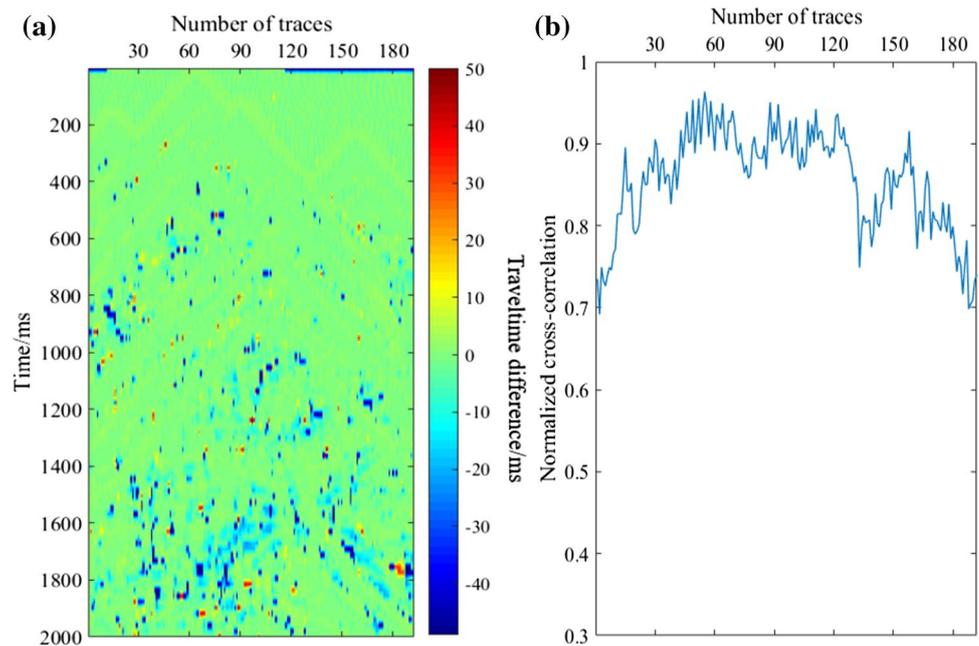


20 ms are similar, and all these time window lengths can be used for the local traveltime correction. Thus, in order to determine the optimal choice of the time window length, we calculate the amplitude spectrum of each trace from the observed data and sum them up to obtain the amplitude spectrum of the entire observed data (Fig. 8). Figure 8 shows the peak frequency of the observed data is 20 Hz, which is equal to the peak frequency of the wavelet, then we calculate the cycle corresponding to this peak frequency

$$H = \frac{1}{f_{\text{peak}}}, \quad (9)$$

where H represents the cycle and f_{peak} represents the peak frequency of the observed data. In this example, the cycle is 50 ms, which is equal to one of the alternative time window lengths. In addition, the seismic data can be regarded as a convolution between the wavelet and the reflected impulse response of subsurface medium. Therefore, we recommend that the time window length should be close to the cycle length of the wavelet, or near the cycle length of the peak frequency corresponding to the amplitude spectrum of the whole observed data. Considering the wavelet might contain low-frequency components, the cycle of the wavelet is longer than that of the peak frequency, so the length of time window can be extended appropriately. Besides, if there is a time difference between the true wavelet and the

Fig. 6 **a** Traveltime differences between the observed and calculated data after traveltime correction ($l=50$ ms, $s=20$ ms) and **b** the normalized cross-correlation between the observed and calculated data after correction ($l=50$ ms, $s=20$ ms)



modeling wavelet, we can obtain the delay time difference approximately by calculating the cross-correlation between the observed and calculated direct waves, then add this time difference to the time window length, which is chosen based on the theory introduced above, and ultimately obtain a new length for the time window. As for the choice of the time step, by comparing Figs. 5a and 9a–e, it shows that the smaller the time step, the higher the accuracy of the traveltime differences between the observed and calculated data. The traveltime differences shown in Figs. 5a and 9a, b are similar, which implies that the traveltime differences will not change significantly over this short time. Another important factor to be considered is the computational efficiency, and Table 1 shows that the smaller the time step, the larger the computational cost. Thus, based on a comprehensive consideration of the computational efficiency and correction accuracy, we recommend that the time step should be as small as possible, if our computational resources allow for that.

After the local traveltime correction, the phase information of the calculated data is more similar to that of the

function to reduce the interference of the incorrect amplitude. The global-correlation misfit function can be expressed as

$$E(v) = - \sum_r \frac{\int_t [d_{cal}^*(r, t, v)d_{obs}(r, t)]dt}{\sqrt{\int_t [d_{cal}^*(r, t, v)]^2 dt} \sqrt{\int_t [d_{obs}(r, t)]^2 dt}}, \quad (10)$$

where d_{cal}^* is the calculated data after the local traveltime correction, and we describe in detail the derivation of the gradient of Eq. (10) in “Appendix”.

Zero-mean normalized cross-correlation misfit function and its gradient

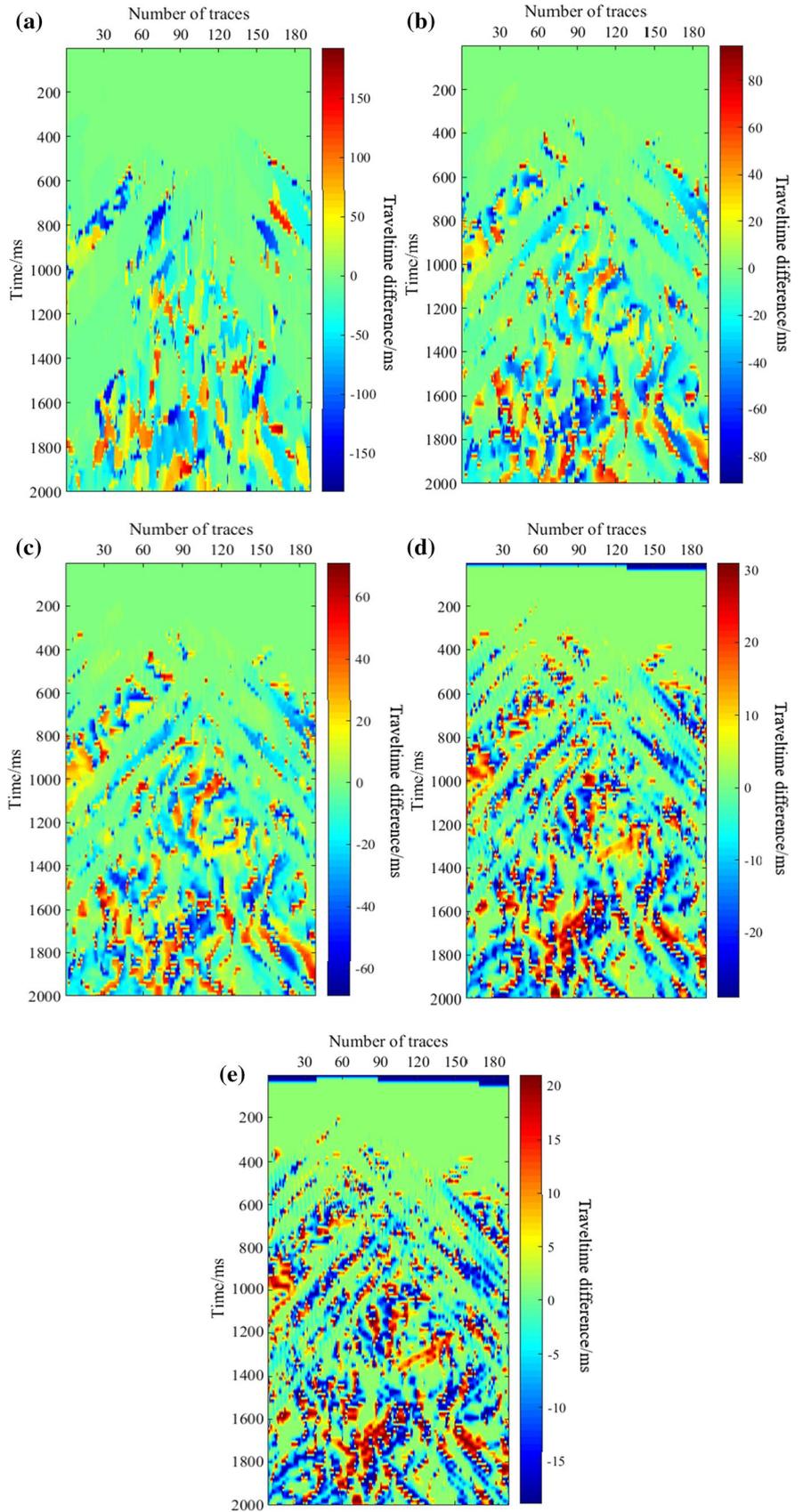
The observed data may contain the low-frequency components, but they are often contaminated with the strong non-zero-mean low-frequency noise. In order to improve the results of FWI under such a circumstance, we propose a zero-mean normalized cross-correlation misfit function, and derive the gradient. This misfit function can be expressed as:

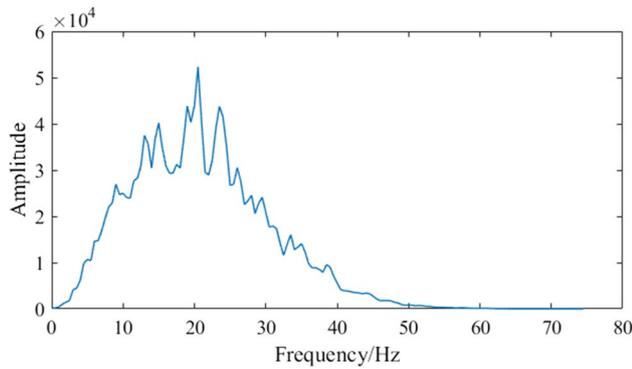
$$J(v) = - \sum_r \frac{\int_t [d_{cal}(r, t, v) - \overline{d_{cal}(r, t, v)}][d_{obs}(r, t) - \overline{d_{obs}(r, t)}]dt}{\sqrt{\int_t [d_{cal}(r, t, v) - \overline{d_{cal}(r, t, v)}]^2 dt} \sqrt{\int_t [d_{obs}(r, t) - \overline{d_{obs}(r, t)}]^2 dt}}, \quad (11)$$

observed data, but the amplitude information between them is significantly different since some parts of the calculated data are set to zero. So we use a global-correlation misfit

where J represents the proposed misfit function and the superscript “ $\overline{\quad}$ ” represents the average of the data. The partial derivative of the misfit function on v is

Fig. 7 Traveltime differences between the observed and calculated data with **a** $l=200$ ms, $s=20$ ms; **b** $l=100$ ms, $s=20$ ms; **c** $l=70$ ms, $s=20$ ms; **d** $l=30$ ms, $s=20$ ms; **e** $l=20$ ms, $s=20$ ms





Note:

$$\widetilde{d}_{cal} = \frac{[d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]}{\sqrt{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]^2 dt}} \tag{14}$$

$$\widetilde{d}_{obs} = \frac{[d_{obs}(r, t) - \overline{d}_{obs}(r, t)]}{\sqrt{\int_t [d_{obs}(r, t) - \overline{d}_{obs}(r, t)]^2 dt}} \tag{15}$$

Fig. 8 The amplitude spectrum of the observed data shown in Fig. 2a

$$\frac{\partial J(v)}{\partial v} = - \sum_r \int_t \left\{ \begin{array}{l} \frac{\frac{\partial [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]}{\partial v} \frac{[d_{obs}(r, t) - \overline{d}_{obs}(r, t)]}{\sqrt{\int_t [d_{obs}(r, t) - \overline{d}_{obs}(r, t)]^2 dt}} \sqrt{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]^2 dt}}{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]^2 dt} - \\ \frac{\frac{\partial [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]}{\partial v} \frac{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)][d_{obs}(r, t) - \overline{d}_{obs}(r, t)] dt}{\sqrt{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]^2 dt} \sqrt{\int_t [d_{obs}(r, t) - \overline{d}_{obs}(r, t)]^2 dt}}}{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]^2 dt} [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)] \end{array} \right\} dt. \tag{12}$$

We reorganize Eq. (12) into:

$$\frac{\partial J(v)}{\partial v} = \sum_r \int_t \frac{\partial [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]}{\partial v} \frac{1}{\sqrt{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]^2 dt}} \left\{ \frac{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)][d_{obs}(r, t) - \overline{d}_{obs}(r, t)] dt}{\sqrt{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]^2 dt} \sqrt{\int_t [d_{obs}(r, t) - \overline{d}_{obs}(r, t)]^2 dt}} \frac{[d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]}{\sqrt{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]^2 dt}} - \frac{[d_{obs}(r, t) - \overline{d}_{obs}(r, t)]}{\sqrt{\int_t [d_{obs}(r, t) - \overline{d}_{obs}(r, t)]^2 dt}} \right\} dt, \tag{13}$$

$$\widetilde{d}_{cal} \widetilde{d}_{obs} = \frac{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)][d_{obs}(r, t) - \overline{d}_{obs}(r, t)] dt}{\sqrt{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]^2 dt} \sqrt{\int_t [d_{obs}(r, t) - \overline{d}_{obs}(r, t)]^2 dt}} \tag{16}$$

So Eq. (13) can be expressed as:

$$\frac{\partial J(v)}{\partial v} = \sum_r \int_t \frac{\partial [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]}{\partial v} \frac{1}{\sqrt{\int_t [d_{cal}(r, t, v) - \overline{d}_{cal}(r, t, v)]^2 dt}} (\widetilde{d}_{cal} \widetilde{d}_{cal} \widetilde{d}_{obs} - \widetilde{d}_{obs}) dt. \tag{17}$$

Fig. 9 Traveltime differences between observed data and calculated data with **a** $l = 50$ ms, $s = 1$ ms; **b** $l = 50$ ms, $s = 10$ ms; **c** $l = 50$ ms, $s = 30$ ms; **d** $l = 50$ ms, $s = 40$ ms; **e** $l = 50$ ms, $s = 50$ ms

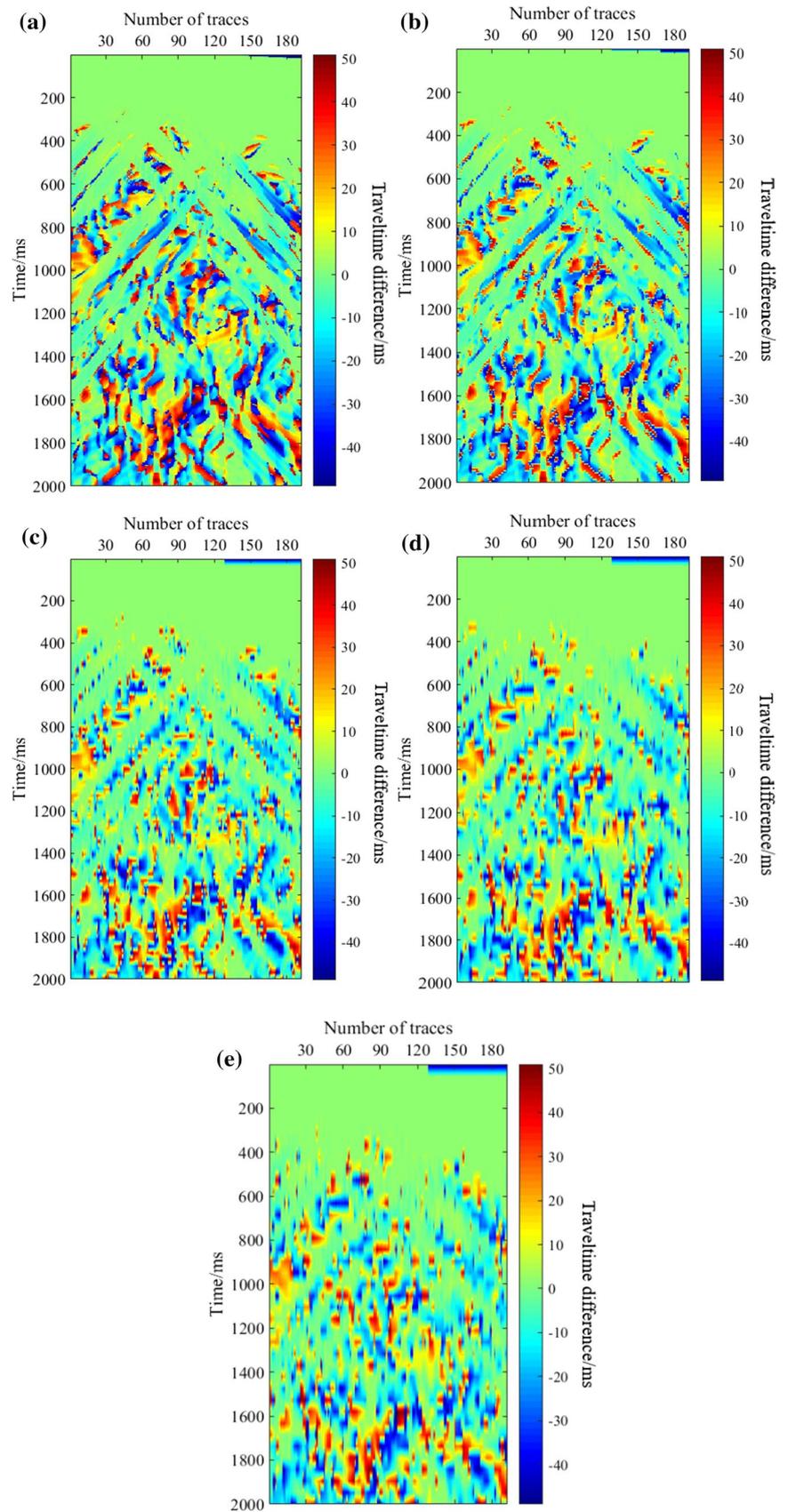


Table 1 Computational cost of different time steps

l, s (ms)	Computational cost (s)
$l=50, s=1$	26.423
$l=50, s=10$	2.563
$l=50, s=20$	1.415
$l=50, s=30$	0.874
$l=50, s=40$	0.696
$l=50, s=50$	0.548

Finally, the gradient can be expressed as:

$$\frac{\partial J(v)}{\partial v} = \sum_r \int_t \frac{\partial d_{\text{cal}}(r, t, v)}{\partial v} \frac{1}{\sqrt{\int_t [d_{\text{cal}}(r, t, v) - \bar{d}_{\text{cal}}(r, t, v)]^2 dt}} \left[\left(\overline{\widetilde{d}_{\text{cal}} \widetilde{d}_{\text{cal}} \widetilde{d}_{\text{obs}} - \widetilde{d}_{\text{obs}}} \right) - \left(\overline{\widetilde{d}_{\text{cal}} \widetilde{d}_{\text{cal}} \widetilde{d}_{\text{obs}} - \widetilde{d}_{\text{obs}}} \right) \right] dt. \quad (18)$$

According to the adjoint state method, Eq. (18) can be simplified to

$$\frac{\partial J(v)}{\partial v} = \sum_r \int_t \frac{\partial d_{\text{cal}}(r, t, v)}{\partial v} \lambda dt, \quad (19)$$

where λ represents the adjoint source, and it is expressed as:

$$\lambda = \frac{1}{\sqrt{\int_t [d_{\text{cal}}(r, t, v) - \bar{d}_{\text{cal}}(r, t, v)]^2 dt}} \left[\left(\overline{\widetilde{d}_{\text{cal}} \widetilde{d}_{\text{cal}} \widetilde{d}_{\text{obs}} - \widetilde{d}_{\text{obs}}} \right) - \left(\overline{\widetilde{d}_{\text{cal}} \widetilde{d}_{\text{cal}} \widetilde{d}_{\text{obs}} - \widetilde{d}_{\text{obs}}} \right) \right]. \quad (20)$$

Therefore, the gradient in the time domain can be expressed as:

$$\frac{\partial J(v)}{\partial v} = \frac{2}{v^3} \sum_r \int_t \frac{\partial^2 u_f}{\partial t^2} u_b^\lambda dt, \quad (21)$$

where u_b^λ is the adjoint wavefield with λ as the adjoint source. The gradient is similar to that of the standard FWI except for the content of the adjoint source, which is a simple modification to the standard adjoint source, and the computational cost should not increase. The workflow of the entire algorithm of FWI based on the zero-mean normalized cross-correlation misfit function and FWI based on the local traveltimes correction method is shown in Fig. 10.

Numerical tests

We demonstrate the waveform local traveltimes correction method and the characteristics of the zero-mean normalized cross-correlation misfit function on the modified Marmousi

model (Fig. 11a). We add a 50-m water layer on the top of the Marmousi model and will not update the velocity of the water layer during the inversion. The initial model is a linearly increasing velocity model (Fig. 11b). The grid dimensions are 69×192 , and the grid spacing in each dimension is 10 m. Each grid point on the surface acts as a receiver, and the Ricker wavelet with a peak frequency of 20 Hz is used as a source. We use the encoded blended sources scheme to reduce the computational cost (Krebs et al. 2009; Schuster et al. 2011; Choi and Alkhalifah 2011), and the blended sources consist of 9 shots. The blended sources are obtained by random phase and amplitude coding. In these tests, we calculate the observed and calculated data using a 10th-order staggered-grid finite-difference method with a 10-layer PML (perfectly matched layer) absorbing boundary. The data total recording time is 2 s with a sample rate of 0.001 s. We use a multi-scale scheme (Bunks et al. 1995; Shin and Ho Cha 2009; Chen et al. 2015) to obtain better inversion results, and we use the gradient normalization formula (Gauthier et al. 1986; Bai et al. 2014) for illumination compensation. The L-BFGS optimization algorithm is used to update the velocity models, and the Wolfe criterion is used to find the updating step length.

Characteristics of the zero-mean normalized cross-correlation misfit function

Firstly, we use the observed data with the strong low-frequency non-zero-mean noise to perform standard FWI using the norm-2 misfit function, the global-correlation misfit function and the zero-mean normalized cross-correlation misfit function. Figure 12a shows the original observed data, which are obtained by modeling in the true model. Figure 12b shows the low-frequency band (0–20 Hz) noise with a mean of 60 dB, and Fig. 12c shows the non-zero-mean noise, which is a kind of random oscillation around the mean value. Next, we add the noise to the observed data (Fig. 12d). The inversion is divided into two steps, that is we first obtain the inverted results for the low-frequency band with a cutoff frequency of 20 Hz ($f_{\text{cut}} = 20$ Hz), and then use it as an initial model to obtain the full-frequency band inverted results (which means the original observed data are used in the inversion directly without low-pass filtering). We regard the information in the data below the peak frequency of the wavelet as low-frequency components, so the cutoff frequency of the low-pass filter in the first step is 20 Hz. Figure 13a, b shows that FWI based on the norm-2 misfit function and the global cross-correlation misfit function are negatively affected by the strong low-frequency non-zero-mean noise, and the inverted results did not accurately represent the long-wavelength components of the true velocity model. The proposed misfit function, on the other hand, behaved well in spite of the strong low-frequency

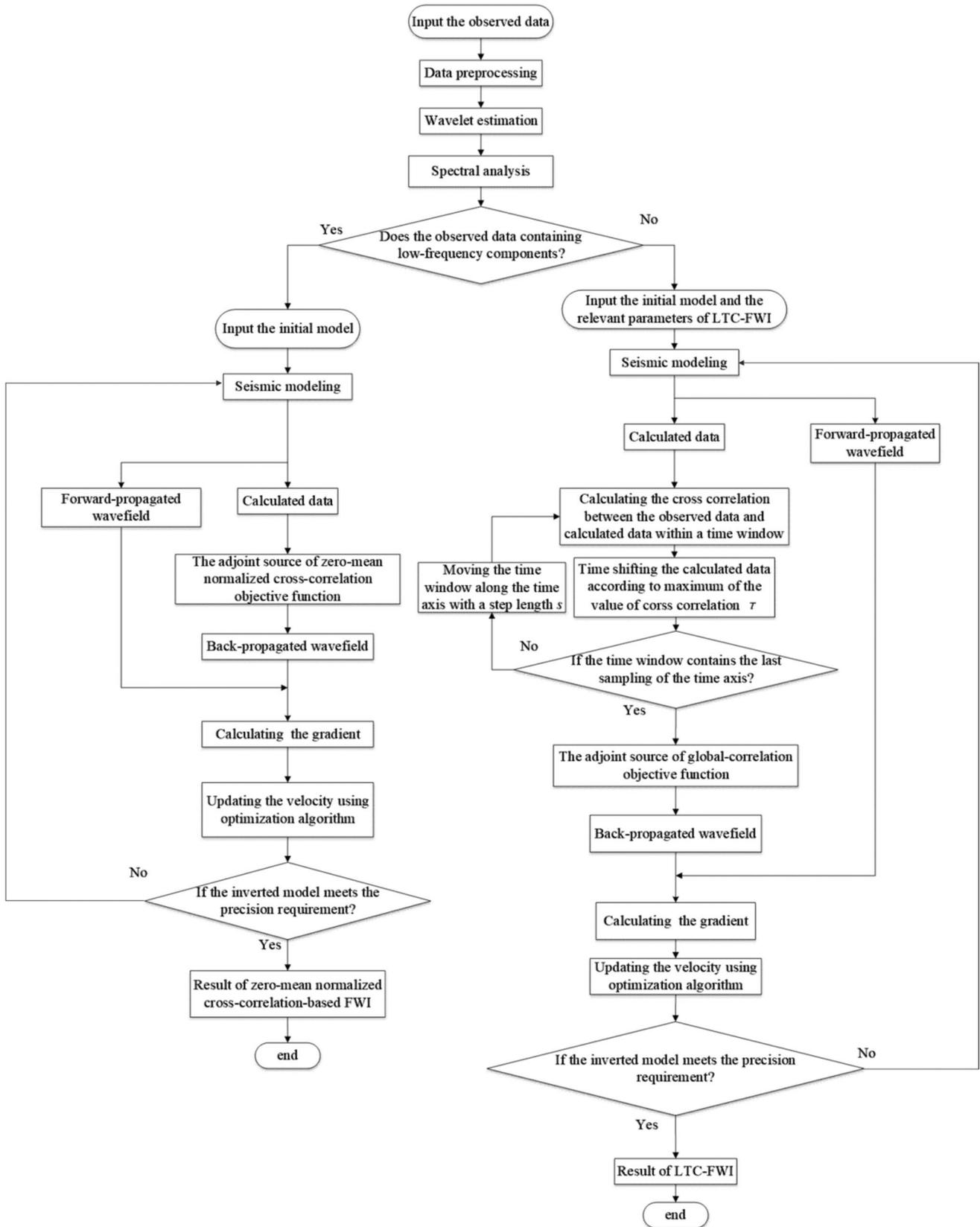


Fig. 10 A flowchart of full waveform inversion based on the zero-mean normalized cross-correlation misfit function and the local traveltme correction method

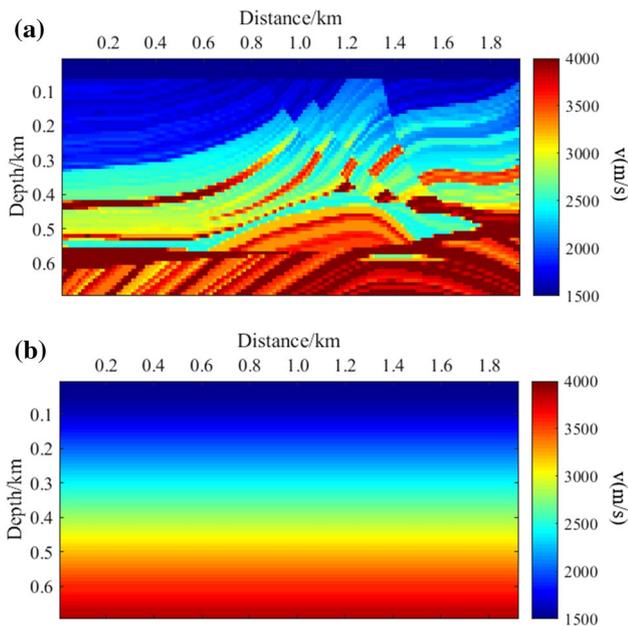


Fig. 11 **a** The Marmousi model and **b** the initial model with linearly increasing velocity

non-zero-mean noise because of its ability in adjusting the mean value in the adjoint source (Fig. 13c). Although the full-frequency band inversion can reduce the influence of the low-frequency noise to some extent, its influence cannot be completely eliminated, and the initial models provided by FWI based on the norm-2 misfit function and global-correlation misfit function are much different from the true model, which causes velocity updating errors in many areas (Fig. 14a, b). In addition, the inverted results based on the zero-mean normalized cross-correlation misfit function can better restore the structure of the true model, and most of the resulting model seem accurate (Fig. 14c). Therefore, the proposed misfit function can handle the strong low-frequency non-zero-mean noise during FWI, which is important in practical applications. Next, we test the performance of zero-mean normalized cross-correlation misfit function when the observed data contain the strong white noise. We use the observed data with a SNR (signal to noise ratio) of -4.46 for testing (Fig. 12e). The low-frequency band (Fig. 15a) and the full-frequency band inverted models (Fig. 15b), both of which start from the initial model shown in Fig. 15a, show good results and no velocity anomalies. This test proves that the misfit function we proposed has the ability to combat white noise, as well.

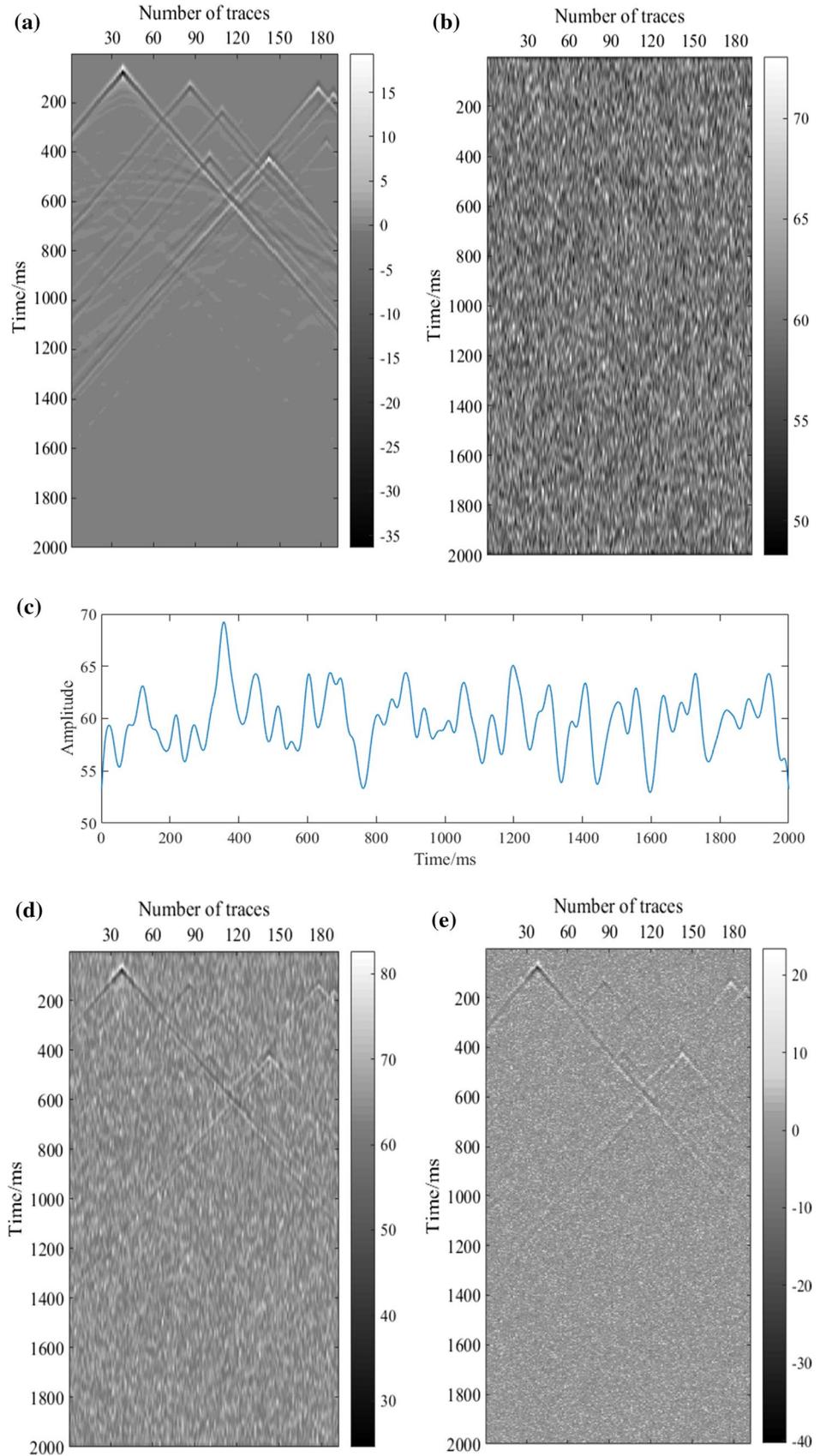
Behavior of the waveform local traveltime correction in FWI (LTC-FWI)

Firstly, we use the observed data, which lack information below 10 Hz to compare the inverted results after LTC-FWI

and the standard FWI for the low-frequency band and both methods use a global-correlation misfit function. Then, we perform a standard FWI using the inverted models for the low-frequency band to provide initial models for the full-frequency band inversion. Figure 16a shows the inverted model using a standard FWI for the low-frequency band. The left side of the model is obviously quite different from the true model due to the cycle skipping. However, the inverted model after the local traveltime correction largely restored the long-wavelength components of the true model, and no cycle skipping occurred (Fig. 16b). Figure 17a shows the inverted model based on the full-frequency band data starting from the initial model shown in Fig. 16a. The areas of velocity updating error caused by cycle skipping cannot be corrected by the high-frequency components in the data. Instead, as the inversion progresses, the incorrect velocity updated values accumulate and eventually become illusions. However, the full-frequency band inverted model starting from the initial model shown in Fig. 16b is nearly the same as the true model, and there is no illusion caused by velocity updating errors (Fig. 17b). For a more intuitive comparison, we extract velocity-depth curves in two different distances from the true model, the initial model and the inverted models. Figure 18a shows that the velocity variation differences between the inverted models after a standard FWI and LTC-FWI at 1.4 km on the right side of the models are nearly the same, and both curves are consistent with the velocity trends of the true model. This finding is primarily observed because the right side of the true model is a sloping structure, and the seismic waves exhibit large-angle scattering during subsurface propagation. Due to layout of our receivers, this large-angle scattering energy can be recorded by the receivers, so the observed data contain rich low-wavenumber components from the right side of the model. However, the left side of the model is given by structure with mainly horizontal bedding, so the observed data mainly include weak reflected information from thin layers, which causes cycle skipping. Figure 18b shows the velocity-depth curves at a distance of 0.4 km on the left side of models, which show that the variation in the velocity of a standard FWI is different from that of the true model in the shallow layers, and obvious velocity updating errors are generated by cycle skipping. The velocity variation of the inverted model after LTC-FWI is nearly the same as the true model except for some small differences, which implies that after the local traveltime correction, the traveltime differences between most of the waveforms in observed and calculated data have been reduced to less than half a cycle. These results show that our method effectively helps FWI avoid cycle skipping when the observed data lack low-frequency components.

Next, we compare the influences of different time window parameters on the LTC-FWI results. If the length of the time window or time step is too large, the traveltime

Fig. 12 The observed data and noises. **a** The original observed data; **b** the low-frequency band (0–20 Hz) noise with a mean of 60 dB. **c** Trace 100 chosen from **b**; **d** the observed data containing the noise shown in **b**; and **e** the observed data containing white noise, SNR = -4.46



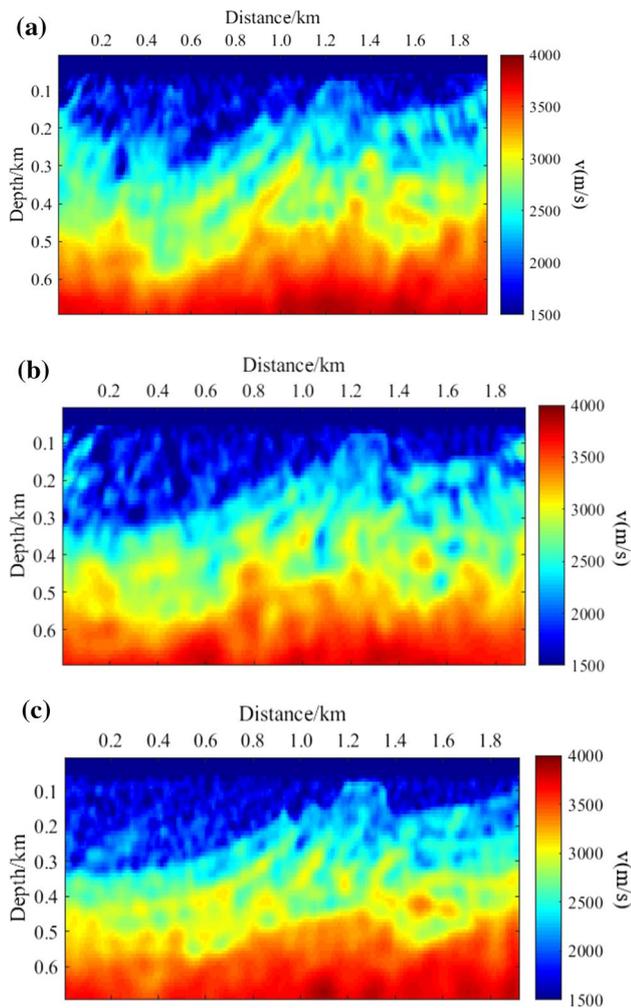


Fig. 13 Inverted models of standard FWI based on the low-frequency observed data ($\text{fluc}=20$ Hz) using **a** the norm-2 misfit function; **b** the global cross-correlation misfit function; and **c** the zero-mean normalized cross-correlation misfit function

correction error between the observed and calculated data will be large (Figs. 7a, 9e), which will directly reduce the accuracy of the waveform matching and ultimately affect the inverted results. Therefore, due to an inappropriate choice of the time window parameters, obvious cycle skipping occurs for the full-frequency band FWI results, with an initial model obtained from LTC-FWI ($l=50$ ms, $s=50$ ms and $l=200$ ms, $s=20$ ms) results shown in Fig. 19a, b. The amount of the velocity updating produced by cycle skipping causes the velocity curves to deviate from those of the true model (Fig. 20a, b). Therefore, this test demonstrates that the criterion for the optimal time window length and time step is effective.

Then, we compared the inverted models when the wavelet used for modeling is incorrect, and all the inverted results are obtained based on the global-correlation misfit function. Firstly, we use a zero time delayed and a 20-ms time delayed

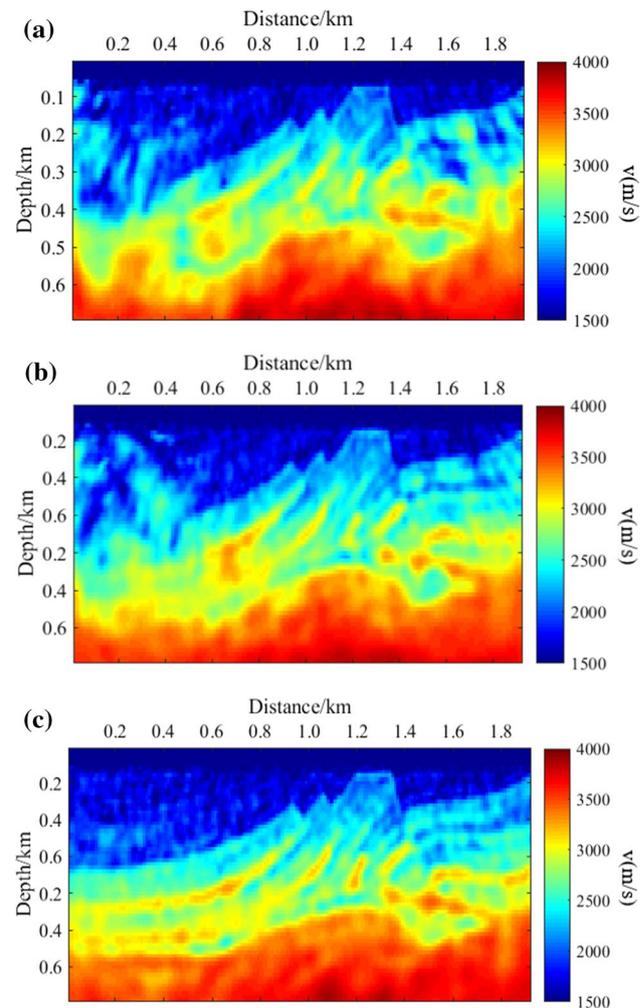


Fig. 14 Inverted models of standard FWI based on the full-frequency band observed data using **a** the norm-2 misfit function; **b** the global cross-correlation misfit function; and **c** the zero-mean normalized cross-correlation misfit function starting from the initial model shown in Fig. 13a–c, respectively

Ricker wavelet with a peak frequency of 20 Hz to produce the observed and calculated data, respectively (Fig. 21). Figure 22a shows the inverted model of the standard FWI for the low-frequency band ($\text{fluc}=20$ Hz). Due to the error between the wavelets, the traveltimes difference will increase during the propagation of wavelets. Thus, the waveform between the observed and calculated data cannot match correctly, which causes an inaccurate velocity updating and the inverted model deviates from the true model eventually. However, the local traveltimes correction method can mitigate the traveltimes difference caused by the wavelet's delay time error and improve the cross-correlation between the waveforms to avoid cycle skipping. Therefore, LTC-FWI result for the low-frequency band can restore the shallow structure of the true model when the modeling wavelet has a large time difference from the true wavelet (Fig. 22b). Next,

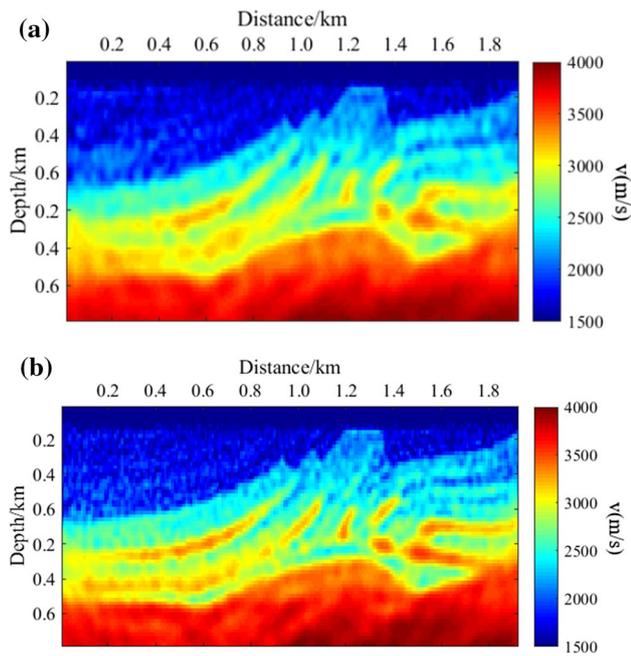


Fig. 15 **a** Low-frequency band inverted model (fluc=20 Hz) based on the observed data containing strong white noise; **b** full-frequency band inverted model starting from the initial model shown in **a**

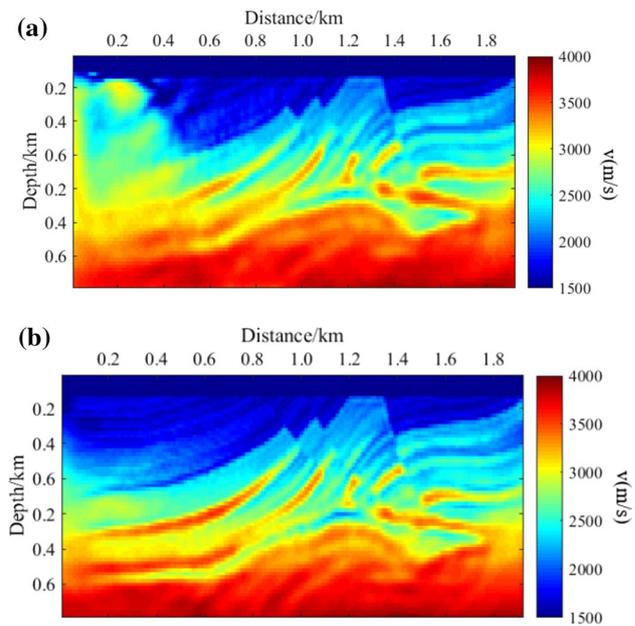


Fig. 17 Inverted models after the standard FWI of full-frequency band observed data starting from the initial model shown in **a** Fig. 16a and **b** Fig. 16b, respectively

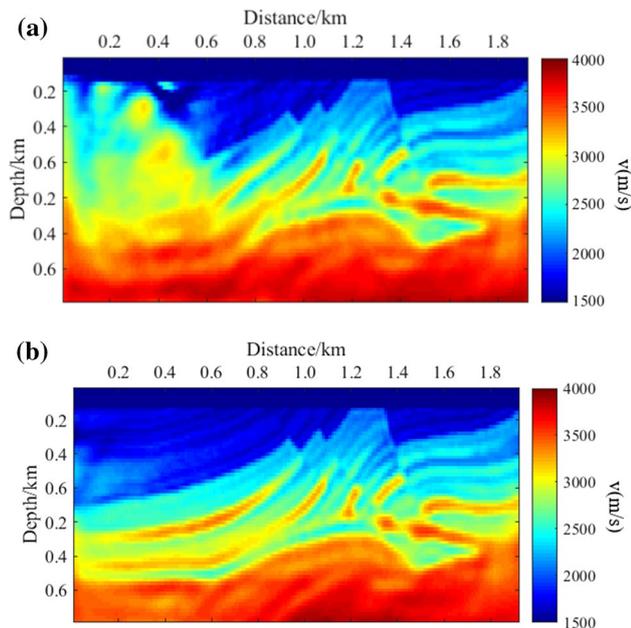


Fig. 16 Inverted models based on the low-frequency band observed data (fluc=20 Hz) after **a** standard FWI and **b** LTC-FWI ($l=50$ ms, $s=20$ ms)

we use an original and a 125° phase-rotated Ricker wavelet with a peak frequency of 20 Hz to produce the observed and calculated data, respectively (Fig. 23). Figure 24a shows the inverted model after a standard FWI for the low-frequency

band (fluc=20 Hz). Because of the phase difference between wavelets, both the amplitude and phase information in the observed data are obviously different from those of the calculated data, which causes severe cycle skipping. Therefore, the inverted result after a standard FWI fails completely. However, by using the local traveltimes correction method, the waveforms with strong energy in the observed and calculated data can be matched well because of their large contribution to the cross-correlation, and the unmatched waveforms with weak energy are mostly set to zero in the time shifting calculation. Thus, the unmatched waveforms have a small influence on the cross-correlation-based gradient. Therefore, Fig. 24b shows that LTC-FWI for the low-frequency band can restore the shallow structure of the true Marmousi model when there is a 125° phase shift between the modeling wavelet and true wavelet. Through these two tests, we demonstrate that LTC-FWI is able to accommodate the interference of inaccurate wavelet to some extent.

Discussion

It is very difficult to calculate the traveltimes differences between the observed and calculated data correctly when the observed data contain the strong low-frequency non-zero-mean noise, because the noise alters the amplitude information of the original data. The amplitude information determines the contribution of each time sample to the

Fig. 18 A comparison of velocity-depth profiles at distances of **a** 1.4 km and **b** 0.4 km, including the true model (red), initial model (black), final inverted model shown in Fig. 17a (green), and the final inverted model shown in Fig. 17b (blue)

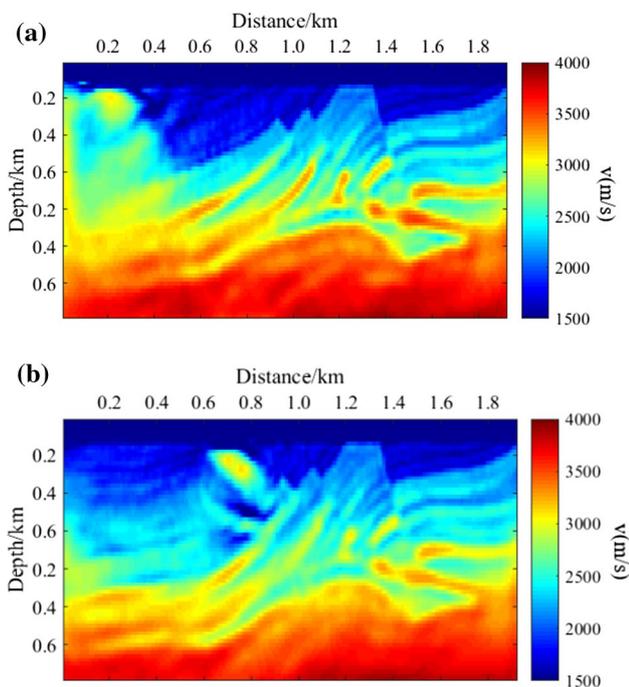
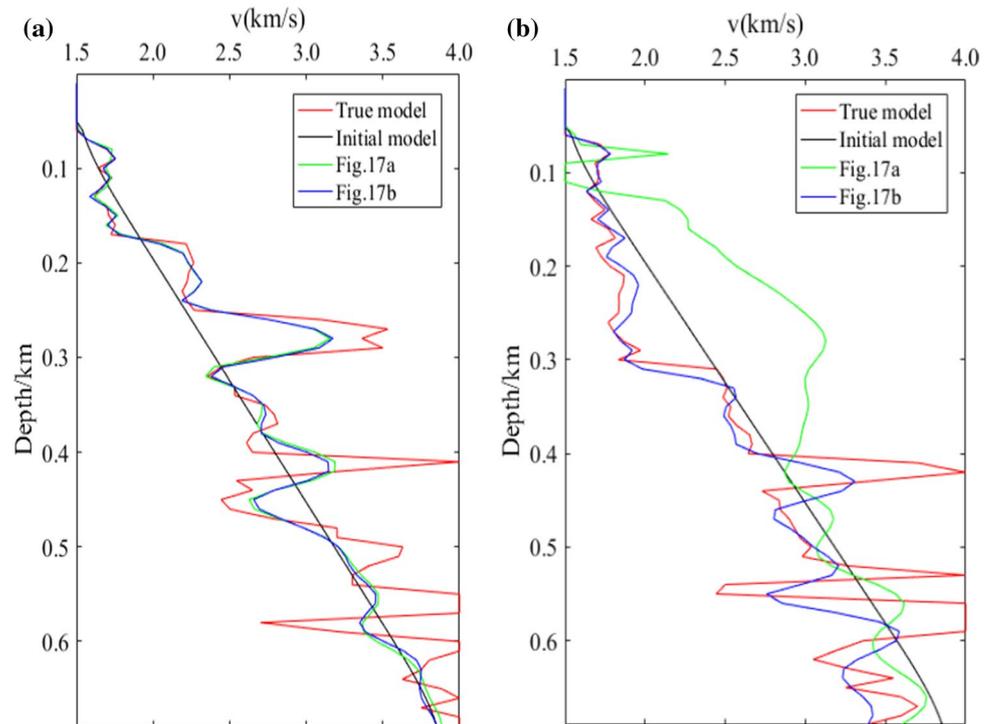


Fig. 19 Inverted models based on **a** LTC-FWI ($l=50$ ms, $s=50$ ms)+standard FWI and **b** LTC-FWI ($l=200$ ms, $s=20$ ms)+standard FWI

cross-correlation-based traveltimes calculation. Thus, the obtained traveltimes are incorrect under such a circumstance. Therefore, we can't combine the local traveltimes correction

method and the zero-mean normalized cross-correlation misfit function to solve the situation when the observed data lack low-frequency components and contain the strong low-frequency noise. However, when the observed data lack low-frequency components, the low-frequency noise does not usually exist. Thus, these two situations often happen separately. Therefore, in this paper, we propose the local traveltimes correction method to solve the cycle skipping problem when the observed data lack low-frequency components, and we propose the zero-mean normalized cross-correlation misfit function to improve the anti-noise capability of FWI when the observed data contain the strong low-frequency non-zero-mean noise.

As for the criterion for the optimal time window length and time step, the suggested criterion in Sect. 2.2 can be further discussed. The recommended criterion is given based on the traveltimes difference accuracy analysis and computational cost analysis, and we use a constant time window length and time step during the inversion. However, the traveltimes difference often increases with time. The traveltimes differences between the observed and calculated data for the direct waves and early arrivals are usually small. Thus, we should use a small window length and time step to process this part of the data. However, the traveltimes differences for the later arrivals are usually large, especially with reflections, which is the main cause of cycle skipping. Thus, in order to calculate the traveltimes differences more accurately, using a gradually increasing time window length and time step along the time axis should produce better

Fig. 20 A comparison of velocity-depth profiles at distances of **a** 0.3 km and **b** 0.7 km, including the true model (red), initial model (black), final inverted model shown in Fig. 17b (blue), the final inverted model shown in Fig. 19a (green), and the final inverted model shown in Fig. 19b (cyan)

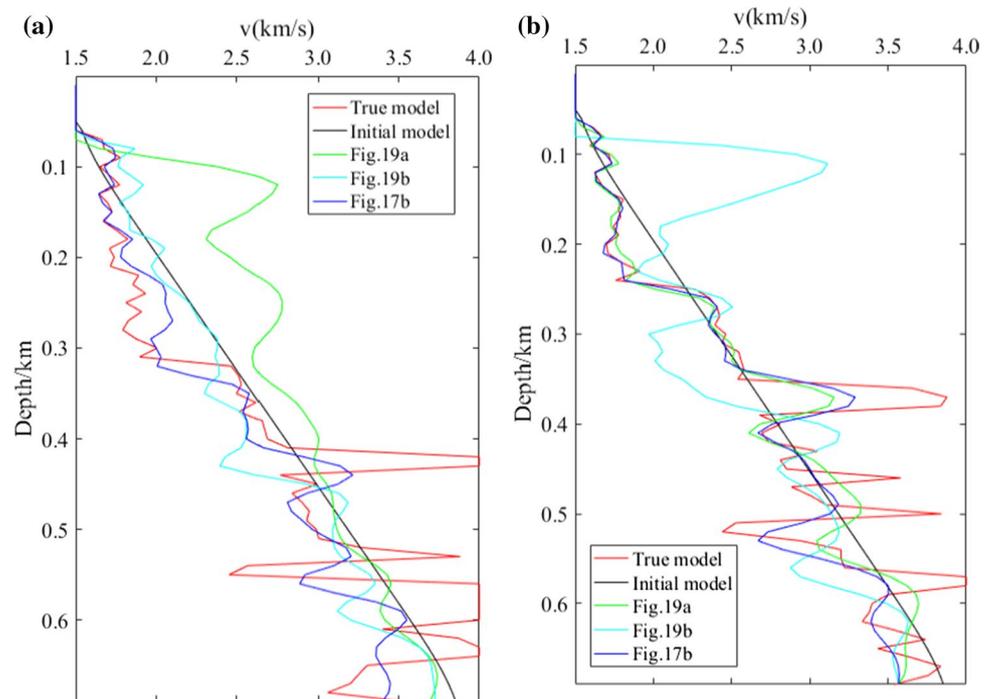
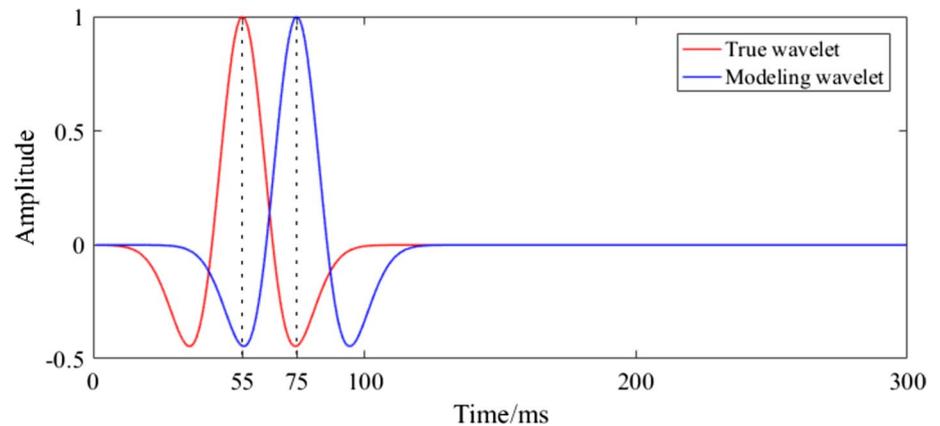


Fig. 21 The wavelets with 20-ms traveltimes difference



LTC-FWI results. The difficulty in the varying window is how to choose the increment in the time window length and time step at different times. We will focus on solving this difficulty in our future works.

In addition, we will try to combine our method with other techniques, such as source-independent schemes or direct envelope inversion, to further solve the cycle skipping problem under a more complex circumstance.

Conclusions

We proposed a waveform local traveltimes correction method to improve the matching between the observed data and the calculated data. After the local traveltimes correction,

cycle skipping can be significantly mitigated, even if the observed data lack low-frequency components. In addition, by choosing the appropriate time window length and time step, the inversion accuracy can be improved without too much additional computational cost. As for the large difference between the modeling wavelet and the true wavelet, the local traveltimes correction method can handle the mismatch caused by the wavelet error between the observed data and the calculated data. Although a high-precision inverted model might not be possible in this case, the shallow structure can be obtained after performing LTC-FWI and no cycle skipping occurs. In addition, FWI based on the zero-mean normalized cross-correlation misfit function can effectively mitigate the interference of the strong low-frequency non-zero-mean noise, which helps FWI avoid cycle skipping, as

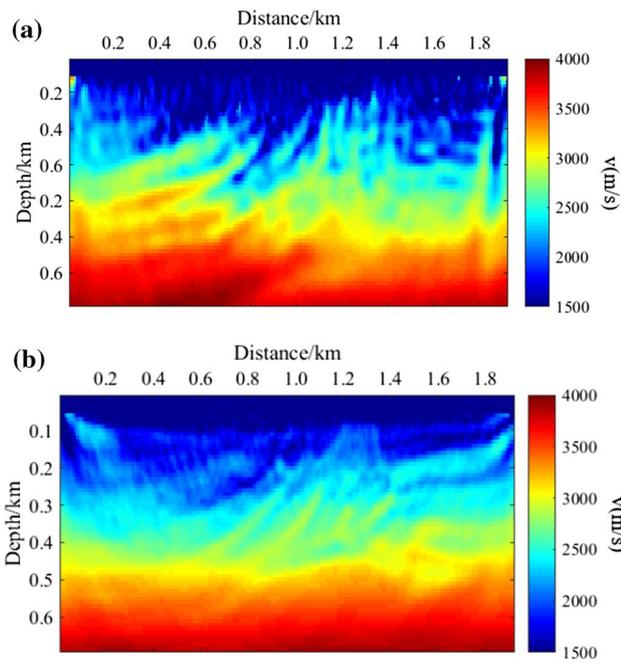


Fig. 22 Inverted models with incorrect wavelets shown in Fig. 21 based on the low-frequency band observed data (fluc=20 Hz) after **a** standard FWI and **b** LTC-FWI ($l=70$ ms, $s=20$ ms)

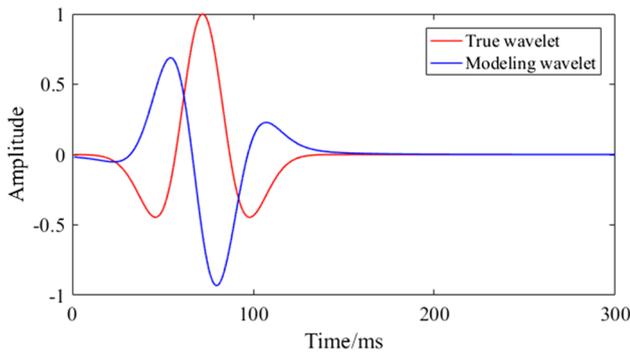


Fig. 23 The wavelets with 125° phase difference

well. Moreover, there is hardly no additional computational cost, because for the proposed misfit function the only difference is the adjoint source. Therefore, whether or not the observed data contain low-frequency components, we introduce a strategy to obtain a better inverted result.

Acknowledgments We are grateful for the support of the National Natural Science Foundation of China (No.41674124).

Compliance with ethical standards

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

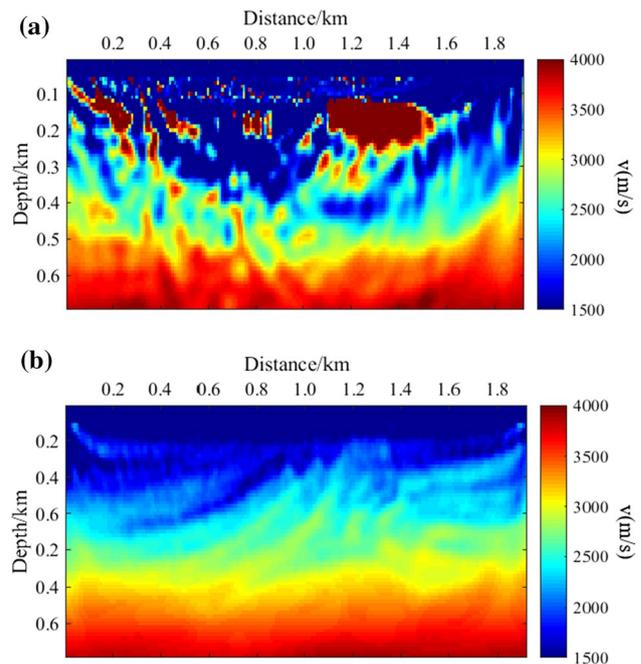


Fig. 24 Inverted models with incorrect wavelets shown in Fig. 23 based on the low-frequency band observed data (fluc=20 Hz) after **a** standard FWI and **b** LTC-FWI ($l=50$ ms, $s=20$ ms)

Appendix

The global-correlation misfit function based on the calculated data after the local traveltimes correction can be expressed as

$$E(v) = - \sum_r \frac{\int_t [d_{cal}^*(r, t, v) d_{obs}(r, t)] dt}{\sqrt{\int_t [d_{cal}^*(r, t, v)]^2 dt} \sqrt{\int_t [d_{obs}(r, t)]^2 dt}}, \quad (22)$$

where d_{cal}^* represents the calculated data after the local traveltimes correction. The partial derivative of Eq. (22) with respect to v is:

$$\frac{\partial E(v)}{\partial v} = - \sum_r \int_t \left\{ \frac{\frac{\partial d_{cal}^*(r, t, v)}{\partial v} \frac{d_{obs}(r, t)}{\sqrt{\int_t [d_{obs}(r, t)]^2 dt}} \sqrt{\int_t [d_{cal}^*(r, t, v)]^2 dt}}{\int_t [d_{cal}^*(r, t, v)]^2 dt} - \frac{\frac{\partial d_{cal}^*(r, t, v)}{\partial v} \frac{\int_t [d_{cal}^*(r, t, v)] [d_{obs}(r, t)] dt}{\sqrt{\int_t [d_{cal}^*(r, t, v)]^2 dt} \sqrt{\int_t [d_{obs}(r, t)]^2 dt}}}{\int_t [d_{cal}^*(r, t, v)]^2 dt} d_{cal}^*(r, t, v)} \right\} dt, \quad (23)$$

We reorganize Eq. (23) into:

$$\frac{\partial E(v)}{\partial v} = \sum_r \int_t \frac{\partial d_{\text{cal}}^*(r, t, v)}{\partial v} \frac{1}{\sqrt{\int_t [d_{\text{cal}}^*(r, t, v)]^2 dt}} \left\{ \frac{\int_t [d_{\text{cal}}^*(r, t, v)] [d_{\text{obs}}(r, t)] dt}{\sqrt{\int_t [d_{\text{cal}}^*(r, t, v)]^2 dt} \sqrt{\int_t [d_{\text{obs}}(r, t)]^2 dt}} \left(\frac{d_{\text{cal}}^*(r, t, v)}{\sqrt{\int_t [d_{\text{cal}}^*(r, t, v)]^2 dt}} - \frac{d_{\text{obs}}(r, t)}{\sqrt{\int_t [d_{\text{obs}}(r, t)]^2 dt}} \right) \right\} dt, \quad (24)$$

According to the adjoint state method, Eq. (24) can be simplified to

$$\frac{\partial E(v)}{\partial v} = \sum_r \int_t \frac{\partial d_{\text{cal}}^*(r, t, v)}{\partial v} \chi dt, \quad (25)$$

where χ represents the adjoint source, and it is expressed as:

$$\chi = \frac{\int_t [d_{\text{cal}}^*(r, t, v) d_{\text{obs}}(r, t)] dt \cdot d_{\text{cal}}^*(r, t, v)}{\sqrt{\int_t [d_{\text{cal}}^*(r, t, v)]^2 dt} \sqrt{\int_t [d_{\text{obs}}(r, t)]^2 dt}} - \frac{d_{\text{obs}}(r, t)}{\sqrt{\int_t [d_{\text{cal}}^*(r, t, v)]^2 dt} \sqrt{\int_t [d_{\text{obs}}(r, t)]^2 dt}}, \quad (26)$$

Therefore, the gradient in the time domain can be simplified to:

$$\frac{\partial E(v)}{\partial v} = \frac{2}{v^3} \sum_r \int_t \frac{\partial^2 u_f}{\partial t^2} u_b^\chi dt. \quad (27)$$

where u_b^χ is the adjoint wavefield with χ as the adjoint source.

References

- Alkhalifah T (2016) Full-model wavenumber inversion: an emphasis on the appropriate wavenumber continuation. *Geophysics* 81:R89–R98. <https://doi.org/10.1190/geo2015-0537.1>
- Alkhalifah T, Choi Y (2014) From tomography to full-waveform inversion with a single objective function. *Geophysics* 79:R55–R61. <https://doi.org/10.1190/geo2013-0291.1>
- Bai J, Yingst D, Bloor R, Leveille J (2014) Viscoacoustic waveform inversion of velocity structures in the time domain. *Geophysics* 79:R103–R119. <https://doi.org/10.1190/geo2013-0030.1>
- Biondi B, Symes WW (2004) Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging. *Geophysics* 69:1283–1298. <https://doi.org/10.1190/1.1801945>
- Bishop TN et al (1985) Tomographic determination of velocity and depth in laterally varying media. *Geophysics* 50:903–923. <https://doi.org/10.1190/1.1441970>
- Bozdağ E, Trampert J, Tromp J (2011) Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. *Geophys J Int* 185:845–870. <https://doi.org/10.1111/j.1365-246X.2011.04970.x>
- Bunks G, Saleck FM, Zaleski S, Chavent G (1995) Multiscale seismic waveform inversion. *Geophysics* 60:1457–1473. <https://doi.org/10.1190/1.1443880>
- Chen G, Chen S, Wu R-S (2015) Full Waveform Inversion in time domain using time-damping filters. *SEG Tech Program Expanded Abstr.* <https://doi.org/10.1190/segam2015-5878658.1>
- Chi B, Dong L, Liu Y (2014) Full waveform inversion method using envelope objective function without low frequency data. *J Appl Geophys* 109:36–46. <https://doi.org/10.1016/j.jappgeo.2014.07.010>
- Chi B, Dong L, Liu Y (2015) Correlation-based reflection full-waveform inversion. *Geophysics* 80:R189–R202. <https://doi.org/10.1190/geo2014-0345.1>
- Choi Y, Alkhalifah T (2011) Source-independent time-domain waveform inversion using convolved wavefields: application to the encoded multisource waveform inversion. *Geophysics* 76:R125–R134. <https://doi.org/10.1190/geo2010-0210.1>
- Choi Y, Alkhalifah T (2012) Application of multi-source waveform inversion to marine streamer data using the global correlation norm. *Geophys Prospect* 60:748–758. <https://doi.org/10.1111/lj.1365-2478.2012.01079.x>
- Choi Y, Alkhalifah T (2015) Unwrapped phase inversion with an exponential damping. *Geophysics* 80:R251–R264. <https://doi.org/10.1190/geo2014-0498.1>
- Choi Y, Alkhalifah T (2018) Time-domain full-waveform inversion of exponentially damped wavefield using the deconvolution-based objective function. *Geophysics* 83:R77–R88. <https://doi.org/10.1190/geo2017-0057.1>
- Datta D, Sen MK (2016) Estimating a starting model for full-waveform inversion using a global optimization method. *Geophysics* 81:R211–R223. <https://doi.org/10.1190/geo2015-0339.1>
- Engquist B, Froese BD, Yang Y (2016) Optimal transport for seismic full waveform inversion. *Commun Math Sci* 14:2309–2330. <https://doi.org/10.4310/CMS.2016.v14.n8.a9>
- Fomel S (2007) Local seismic attributes. *Geophysics* 72:A29–A33. <https://doi.org/10.1190/1.2437573>
- Gauthier O, Virieux J, Tarantola A (1986) Two-dimensional nonlinear inversion of seismic waveforms: numerical results. *Geophysics* 51:1387–1403. <https://doi.org/10.1190/1.1442188>
- Guo Q, Alkhalifah T (2017) Elastic reflection-based waveform inversion with a nonlinear approach. *Geophysics* 82:R309–R321. <https://doi.org/10.1190/geo2016-0407.1>
- Hale D (2006) Fast local cross-correlations of images. *SEG Tech Program Expanded Abstr.* <https://doi.org/10.1190/1.2370185>
- Hu W (2014) FWI without low frequency data-beat tone inversion. *SEG Tech Program Expanded Abstr.* <https://doi.org/10.1190/segam2014-0978.1>
- Hu Y, Han L, Xu Z, Zhang F, Zeng J (2017) Adaptive multi-step full waveform inversion based on waveform mode decomposition. *J Appl Geophys* 139:195–210. <https://doi.org/10.1016/j.jappgeo.2017.02.017>
- Hu Y, Han L, Zhang P, Ge Q (2018) Time-frequency domain multi-scale full waveform inversion based on adaptive non-stationary phase correction. *EAGE Tech Program Ext Abstr.* <https://doi.org/10.3997/2214-4609.201800891>
- Justice JH et al (1989) Acoustic tomography for enhancing oil recovery. *Lead Edge* 8:12–19. <https://doi.org/10.1190/1.1439605>
- Krebs JR, Anderson JE, Hinkley D, Neelamani R, Lee S, Baumstein A, Lacasse MD (2009) Fast full-wavefield seismic inversion using encoded sources. *Geophysics* 74:WCC177–WCC188. <https://doi.org/10.1190/1.3230502>
- Lailly P (1983) The seismic inverse problem as a sequence of before stack migrations. In: *Conference on inverse scattering: theory and application*, pp 206–220

- Lee KH, Kim HJ (2003) Source-independent full-waveform inversion of seismic data. *Geophysics* 68:2010–2015. <https://doi.org/10.1190/1.1635054>
- Li Y, Choi Y, Alkhalifah T, Li Z, Zhang K (2018) Full-waveform inversion using a nonlinearly smoothed wavefield. *Geophysics* 83:R117–R127. <https://doi.org/10.1190/geo2017-0312.1>
- Lian S, Yuan S, Wang G, Liu T, Liu Y, Wang S (2018) Enhancing low-wavenumber components of full-waveform inversion using an improved wavefield decomposition method in the time-space domain. *J Appl Geophys* 157:10–22. <https://doi.org/10.1016/j.jappgeo.2018.06.013>
- Liu Y, Teng J, Xu T, Wang Y, Liu Q, Badal J (2016) Robust time-domain full waveform inversion with normalized zero-lag cross-correlation objective function. *Geophys J Int* 209:106–122. <https://doi.org/10.1093/gji/ggw485>
- Liu Q, Peter D, Tape C (2019) Square-root variable metric based elastic full-waveform inversion—part 1: theory and validation. *Geophys J Int* 218:1121–1135. <https://doi.org/10.1093/gji/ggz188>
- Luo Y, Schuster GT (1991) Wave-equation traveltime inversion. *Geophysics* 56:645–653. <https://doi.org/10.1190/1.1443081>
- Mikesell TD, Malcolm AE, Yang D, Haney MM (2015) A comparison of methods to estimate seismic phase delays: numerical examples for coda wave interferometry. *Geophys J Int* 202:347–360. <https://doi.org/10.1093/gji/ggv138>
- Pan W, Huang L (2018) Elastic-wave-equation traveltime tomography in anisotropic media using differential measurements of long-offset data. *SEG Tech Program Expand Abstr* 2018:5233–5237. <https://doi.org/10.1190/segam2018-2997226.1>
- Pratt RG, Shin C, Hicks GJ (1998) Gauss–Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophys J Int* 133:341–362. <https://doi.org/10.1046/j.1365-246X.1998.00498.x>
- Sajeva A, Aleari M, Stucchi E, Bienati N, Mazzotti A (2016) Estimation of acoustic macro models using a genetic full-waveform inversion: applications to the Marmousi model. *Geophysics* 81:R173–R184. <https://doi.org/10.1190/geo2015-0198.1>
- Schuster GT, Wang X, Huang Y, Dai W, Boonyasiriwat C (2011) Theory of multisource crosstalk reduction by phase-encoded statics. *Geophys J Int* 184:1289–1303. <https://doi.org/10.1111/j.1365-246X.2010.04906.x>
- Shin C, Ho Cha Y (2009) Waveform inversion in the Laplace–Fourier domain. *Geophys J Int* 177:1067–1079. <https://doi.org/10.1111/j.1365-246X.2009.04102.x>
- Shin C, Min D-J (2006) Waveform inversion using a logarithmic wavefield. *Geophysics* 71:R31–R42. <https://doi.org/10.1190/1.2194523>
- Sun B, Alkhalifah T (2019) Adaptive traveltimes inversion. *Geophysics* 84:U13–U29. <https://doi.org/10.1190/geo2018-0595.1>
- Symes WW, Carazzone JJ (1991) Velocity inversion by differential semblance optimization. *Geophysics* 56:654–663. <https://doi.org/10.1190/1.1443082>
- Tarantola A (1984) Linearized inversion of seismic reflection data. *Geophys Prospect* 32:998–1015. <https://doi.org/10.1111/j.1365-2478.1984.tb00751.x>
- Virieux J, Operto S (2009) An overview of full-waveform inversion in exploration geophysics. *Geophysics* 74:WCC1–WCC26. <https://doi.org/10.1190/1.3238367>
- Wang G, Yuan S, Wang S (2019) Retrieving low-wavenumber information in FWI: an efficient solution for cycle skipping. *IEEE Geosci Remote Sens Lett* 16:1125–1129. <https://doi.org/10.1109/Lgrs.2019.2892998>
- Warner M, Guasch L (2014) Adaptive waveform inversion: theory. *SEG Tech Program Expand Abstr*. <https://doi.org/10.1190/segam2014-0371.1>
- Wu R-S, Luo J, Wu B (2014) Seismic envelope inversion and modulation signal model. *Geophysics* 79:WA13–WA24. <https://doi.org/10.1190/geo2013-0294.1>
- Xu S, Wang D, Chen F, Zhang Y, Lambare G (2012) Full waveform inversion for reflected seismic data. *EAGE Tech Program Ext Abstr*. <https://doi.org/10.3997/2214-4609.20148725>
- Yang H, Han L, Zhang F, Sun H, Bai L (2016) Full waveform inversion without low frequency using wavefield phase correlation shifting method. *J Seism Explor* 25:45–55
- Yuan S, Wang S, Luo Y, Wei W, Wang G (2019) Impedance inversion by using the low-frequency full-waveform inversion result as an a priori model. *Geophysics* 84:R149–R164. <https://doi.org/10.1190/geo2017-0643.1>
- Zhang P, Han L, Xu Z, Zhang F, Wei Y (2017) Sparse blind deconvolution based low-frequency seismic data reconstruction for multiscale full waveform inversion. *J Appl Geophys* 139:91–108. <https://doi.org/10.1016/j.jappgeo.2017.02.021>
- Zhang Z, Alkhalifah T, Wu Z, Liu Y, He B, Oh J (2019) Normalized nonzero-lag crosscorrelation elastic full-waveform inversion. *Geophysics* 84:R1–R10. <https://doi.org/10.1190/geo2018-0082.1>
- Zhou B, Greenhalgh SA (2003) Crosshole seismic inversion with normalized full-waveform amplitude data. *Geophysics* 68:1320–1330. <https://doi.org/10.1190/1.1598125>
- Zhu H, Fomel S (2016) Building good starting models for full-waveform inversion using adaptive matching filtering misfit. *Geophysics* 81:U61–U72. <https://doi.org/10.1190/geo2015-0596.1>



Possibility of an earthquake prediction based on monitoring crustal deformation anomalies and thermal anomalies at the epicenter of earthquakes with oblique thrust faulting

Zahra Alizadeh Zakaria¹ · Farshid Farnood Ahmadi¹

Received: 26 April 2019 / Accepted: 30 November 2019 / Published online: 6 December 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

Is it possible to predict location, time and magnitude of earthquakes through identifying their precursors based on remotely sensed data? Earthquakes are usually preceded by unusual natural incidents that are considered as earthquake precursors. With the recent advances in remote sensing techniques which have made it possible monitoring the earth's surface with different sensors, scientists are now able to better study earthquake precursors. Thus, the present study aims at developing the algorithm of classic PS-InSAR processing for obtaining crustal deformation values at the epicenter of earthquakes with magnitude larger than 5.0 on the Richter scale and with oblique thrust faulting and then after calculating temperature values using remotely sensed thermal imagery at the epicenter of same earthquakes; thermal and crustal deformation anomalies were calculated using data mining techniques before earthquake occurrence. In the next stage, taking the correlation between thermal anomalies and crustal deformation anomalies at the epicenter of the study earthquakes into account, an integrated technique was proposed to predict probable magnitude and time of oblique thrust earthquakes occurrence over the earthquake-prone areas. Eventually, the validity of the proposed algorithm was evaluated for an earthquake with a different focal mechanism. The analysis results of the thermal anomalies and crustal deformation anomalies at the epicenter of April 16, 2016, Japan-Kumamoto earthquake of magnitude 7.0 with strike-slip faulting, showed completely different trends than the suggested patterns by the proposed algorithm.

Keywords Crustal deformation anomalies · Data mining · Earthquake precursors · Earthquake prediction · Extended-PS method · Surface thermal anomalies

Introduction

Earthquake prediction with purposes of reducing death and destruction caused by earthquakes is a significant issue that has attracted the attention of many experts and researchers in the field of disaster management (Tsai and Chen 2010; Rikitake 1968). While most seismologists insist on the improbability of earthquake prediction in terms of location and time of occurrence using the current knowledge and facilities, they believe that there are certain precursors that may lead to finding more reliable methods by investigating them effectively (Rikitake 1975). Therefore, considering the

significance of this issue, extensive researches from various scientific fields have been concentrated on identifying these precursors. The most common predictive methods based on identifying precursors include monitoring the ground displacement along a fault using ground-based tools such as global positioning systems (GPS) (Yue and Lay 2011; Wang et al. 2013; Calais et al. 2003) or radar interferometry on radar images of satellite sensors (Tomás et al. 2014; Moro et al. 2017; Graham 1974; Bamler and Hartl 1998; Gabriel et al. 1989; Massonnet and Feigl 1998), earthquake prediction using geophysical techniques such as determining probable location or, sometimes, probable time of earthquake occurrence through foreshocks (Ogata and Katsura 2012; Moreno et al. 2010; Lippiello et al. 2012), earthquake prediction by means of assessing variations in velocity of seismic waves of S and P prior to earthquake (Pio Lucente et al. 2010; Peacock et al. 1988), earthquake prediction through monitoring thermal anomalies on the earth surface using

✉ Farshid Farnood Ahmadi
f.farnood.tabrizu@iran.ir

¹ Department of Geomatics Engineering, University of Tabriz, 29 Bahman Boulevard, Tabriz, Iran

thermal images obtained by remote sensors in fault-exposed areas (Pulinets et al. 2006; Saraf et al. 2009; Saradjian and Akhoondzadeh 2011), monitoring changes in ionospheric precursors analyzing the collected data from remote sensing DEMETER satellites (Molchanov et al. 2006; Akhoondzadeh et al. 2010) or ground-based systems (Yao et al. 2012; Pulinets and Davidenko 2014), earthquake prediction through monitoring earthquake clouds along fault range with the aid of satellite imagery (Thomas et al. 2015; Guo and Wang 2008), earthquake prediction using geochemical and hydrological precursors like monitoring variations in concentration of ions or dissolved gases including radon, helium, etc. (Thomas 1988; Ingebritsen and Manga 2014), earthquake prediction through studying changes in animal behaviors (Grant et al. 2015) and earthquake prediction by measuring electromagnetic changes in ULF range (Hayakawa 2013; Han et al. 2011).

Most of the above-mentioned methods require the installation of the equipment on the ground or direct contact with the ground. Since the long-term surveying installations within an area are time- and cost-consuming, the remote sensing techniques such as monitoring thermal anomalies and land surface deformation prior to earthquake within earthquake-prone areas using satellite images are of higher priority because they do not need physical presence at the site, and also the required data are continuously collected from a wider range with high spatiotemporal resolution and lower costs. But, remote sensing methods, similar to other predictive approaches, have a single-parametrical orientation toward the issue. On the other hand, since the inherent behavioral variations of the precursors don't follow a common pattern in all earthquakes, and the changes show peculiar behaviors to each region, it seems unfeasible to provide an obvious prediction in an area solely based on behavioral variations in a single precursor. Even in the case of the studies which have focused on investigating several precursors simultaneously, the correlation among latent structural patterns in the collected data related to different precursors has been ignored. Thus, these methods are not able to provide a reliable prediction about the location and time of earthquakes. However, these limitations never underestimate the precursors. Besides, since each precursor contains some information about earthquakes, it seems more likely to find an effective technique or an acceptable indicator for earthquake prediction. Therefore, the present study aims at developing the algorithm of classic PS-InSAR processing for obtaining crustal deformation values at the epicenter of earthquakes with magnitude larger than 5.0 on the Richter scale and with oblique thrust faulting, and then after calculating temperature values using remotely sensed thermal imagery at the epicenter of same earthquakes; thermal and crustal deformation anomalies were calculated using

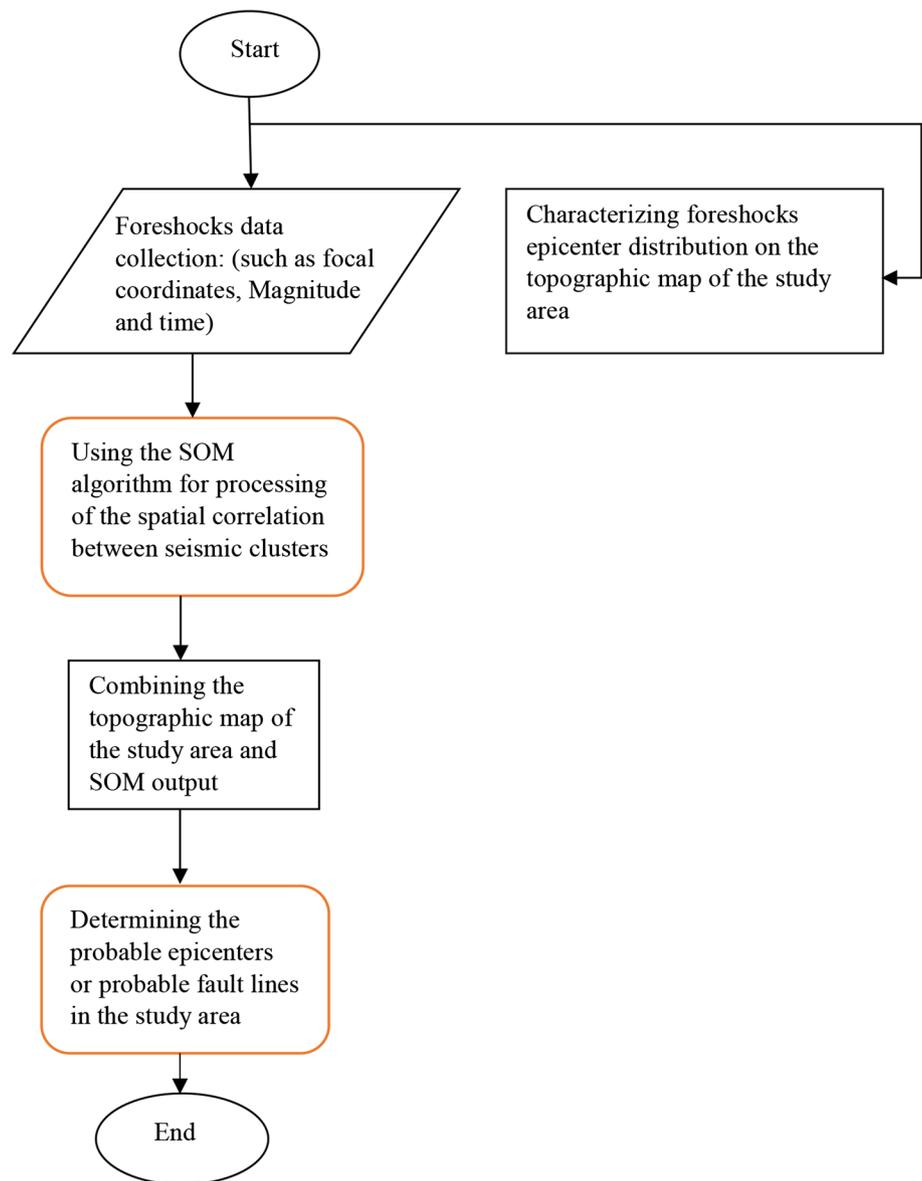
data mining techniques in the period of two months before earthquake occurrence; then, similarities or discrepancies among behavioral variations of these anomalies were analyzed. In the next stage, this approach was initially carried out in a unidimensional mode and then in an integrated form to offer a certain reliable method to predict the probable magnitude and time of earthquake based on identifying thermal anomalies and crustal deformation anomalies within earthquake-prone regions. Eventually, the validity of the proposed algorithm was evaluated for an earthquake with a different focal mechanism. The analysis results of the thermal anomalies and crustal deformation anomalies at the epicenter of April 16, 2016, Japan-Kumamoto earthquake of magnitude 7.0 with strike-slip faulting, showed completely different trends than the suggested patterns by the proposed algorithm. So the proposed algorithm was unable to determine the probable time and magnitude of an earthquake occurrence with strike-slip faulting.

Methodology

Recognition of the earthquake-prone areas

It should be noted that the remote sensing approaches to monitoring anomalies caused by earthquake precursors carry out the monitoring operation globally and over a wide area. Therefore, this issue comes into question that to monitor anomalies caused by earthquake precursors, which areas of the earth surface have to be monitored? In the present study, thermal anomalies and crustal deformation anomalies have been monitored regarding the already happened earthquakes. In other words, the epicenter of earthquakes to perform analyses was already identifiable. According to serious incidents, earthquake-prone areas are usually preceded by some turbulences such as minor shocks in the area, which are completely random and do not follow a certain pattern or principle, whereas the predictive locations of such shocks follow a certain order despite the high level of irregularity. But, it would be possible to estimate the probable location of earthquakes using modern statistical techniques like self-organizing map (SOM) neural networks (Mohankumar and Sangeetha 2018). This method is not only capable of identifying risky regions but also can give ideas about places that are vulnerable to earthquakes. As discussed above in the present study, the epicenter of earthquakes to perform analyses was already identifiable, but for evaluating the ratibility of SOM networks to recognize earthquake-prone areas, we applied the SOM algorithm for identifying earthquake-prone areas in one of the study areas, whose stages are depicted in detail in Fig. 1.

Fig. 1 SOM algorithm for analyzing spatiotemporal correlation between seismic clusters



Calculation of the crustal deformation anomalies at the epicenter of an earthquake

Monitoring of the ground surface deformations at the epicenter of an earthquake using radar interferometry

Ground surface deformations are considered as one of the direct impacts caused by tension exerted on the rocks, so by measuring length and width of a certain region of the ground surface located on an active plate, it is possible to map movements of the plate in relation to time, which can play a significant role in predicting earthquake probable time and location (Parsons 2017; Huang et al. 2017). A wide range of methods has been developed so far to measurement crustal

movements. However, using radar interferometry technique has some advantages over other methods due to its wide coverage over ground surface, repeatability of observations over time, high spatiotemporal resolution and saving time and costs (Bürgmann et al. 2000; Rosen et al. 1998; Massonnet et al. 1993; Peltzer and Rosen 1995).

Advanced techniques of InSAR involving the simultaneous processing of multiple SAR acquisitions in time increase the number of locations where a deformation signal can be extracted and reduce associated error. Currently, there are two broad categories of algorithms for processing multiple acquisitions, persistent scatterer (PS-INSAR) and small baseline methods (SBAS), which are optimized for different models of scattering. Each of these two methods gives temporal solution to the uncorrelated phenomena based on the

principle of dominant reflectivity of centers of permanent scatterers (Hooper 2008). Considering that the SBAS technique relies on an appropriate combination of differential interferograms created by using image pairs characterized by a small orbital separation (baseline), using of the SBAS method leads to low spatial decorrelation results. Therefore, in the present study according to the PS method (Tu et al. 2012) which uses large stacks of SAR images to generate differential interferograms concerning common one master, we used classic PS-InSAR principles in our analyses. But classic PS technique to calculate the displacements of selected points faces this restriction that the amount of displacements is only measurable at the PS points around our study area which may not correspond to the target points (e.g., Earthquake epicenter). Thus, the classic PS method was not responsive to the purpose of this study to monitoring surface deformations of the earthquake's epicenter. So considering the limitations of the classic PS method due to lack of concentration on a selective point with specific geographical coordinates, in the present study the problem of determining the displacements at the selective points has been solved through spatiotemporal interpolation methods (Li and Heap 2014; Yang et al. 2004) based on calculated displacements for the PS points, whose stages are depicted in detail in Fig. 2a.

Behavioral analysis of the earthquake epicenter deformations and calculating crustal deformation anomalies using data mining

In the presented study, data mining methods (Hand 2007) were used to analyze the behavior of the crustal deformations at the epicenter of the earthquakes in order to calculate crustal deformation anomalies.

It is noteworthy that data mining uses a set of statistical and mathematical methods which can lead to recognition of the latent patterns and meaningful relationships within databases. Once a behavioral pattern has been identified among the databases and this pattern has been learned through mathematical and statistical models including regression, artificial neural networks, time series, etc., variations in the phenomenon behavior can be predicted, evaluated and interpreted based on its preceding behavior. Since the ARIMA model (Brockwell et al. 1991) involves a special set of univariate modeling for analysis and prediction of time series, the present study used the ARIMA (p , d and q) model to analyze time series of the calculated displacements for the epicenter of the earthquakes (as presented in Fig. 2b).

According to that, the measured data considered an anomaly when the difference between the predicted and actual values of the data goes beyond the range of RMSE defined by the statistical models; thus, the amount of crustal

deformation anomalies of three months before the earthquake occurrence was calculated according to Eq. (1):

$$\text{Anomaly: } z_{\text{Real}} - z_{\text{calculated}} \mp \text{RMSE} \quad (1)$$

Identifying thermal anomalies at the epicenter of the earthquake

As already acknowledged by the previous researches, several phenomena are believed to trigger the increase in land surface temperature before the earthquake occurrence. Despite that the nature of this phenomenon and the mechanism of increasing land surface temperature before the earthquake are still vaguely understood (Qiang et al. 1997), monitoring thermal anomalies caused by earthquake can be one of the major approaches to estimate parameters related to earthquakes of magnitude larger than 4.5 on the Richter scale, which can be monitored by various thermal sensors and satellites in different spatial and temporal resolutions at TIR wavelengths (Tronin 2000; Huang et al. 2008).

Different studies have claimed different occurrence times of the LST anomalies with a wide range 1–24 days before the earthquake in the form of an increase in the temperature of 2–12 °C (Zhang et al. 2014; Pohn et al. 1974). It should be noted that in recent researches there are also other cases with a different range of 30–40 days before the earthquake that have been reported (Lu et al. 2016). On the other hand, considering the fact that the earth acts like an isothermal and homogenous surface during night and is exposed to thick clouds, shade, and direct sunlight during daytime, this could lead to the emergence of a temperature difference in the range of 15°–20° (Spencer et al. 1999), which would conceal thermal anomalies during the day. Thus, in the present study for providing reliable analysis of thermal anomalies and their relationship with earthquakes, average temperature day and night products were used. Also, considering that the temperature increase happens occasionally and locally along a fault, if there is no anomaly or turbulence within the area, all points (focal area and around it) will follow definite increasing or decreasing temperature patterns in each season. Therefore, it will be expectable that anomalies or turbulences such as the earthquake phenomenon may disrupt this compliance at the epicenter both spatially and temporally in comparison with other points within the area. So, in the present study as shown in Fig. 3b for calculating thermal anomalies unlike most of the previous investigations that have used time-series analysis methods to model temperature variations around the earthquake epicenter, the modeling of thermal variations was proposed taking into account the spatiotemporal correlations between epicenter thermal data and adjacent areas using regression functions Eq. (2). It should be noted that thermal anomalies detection

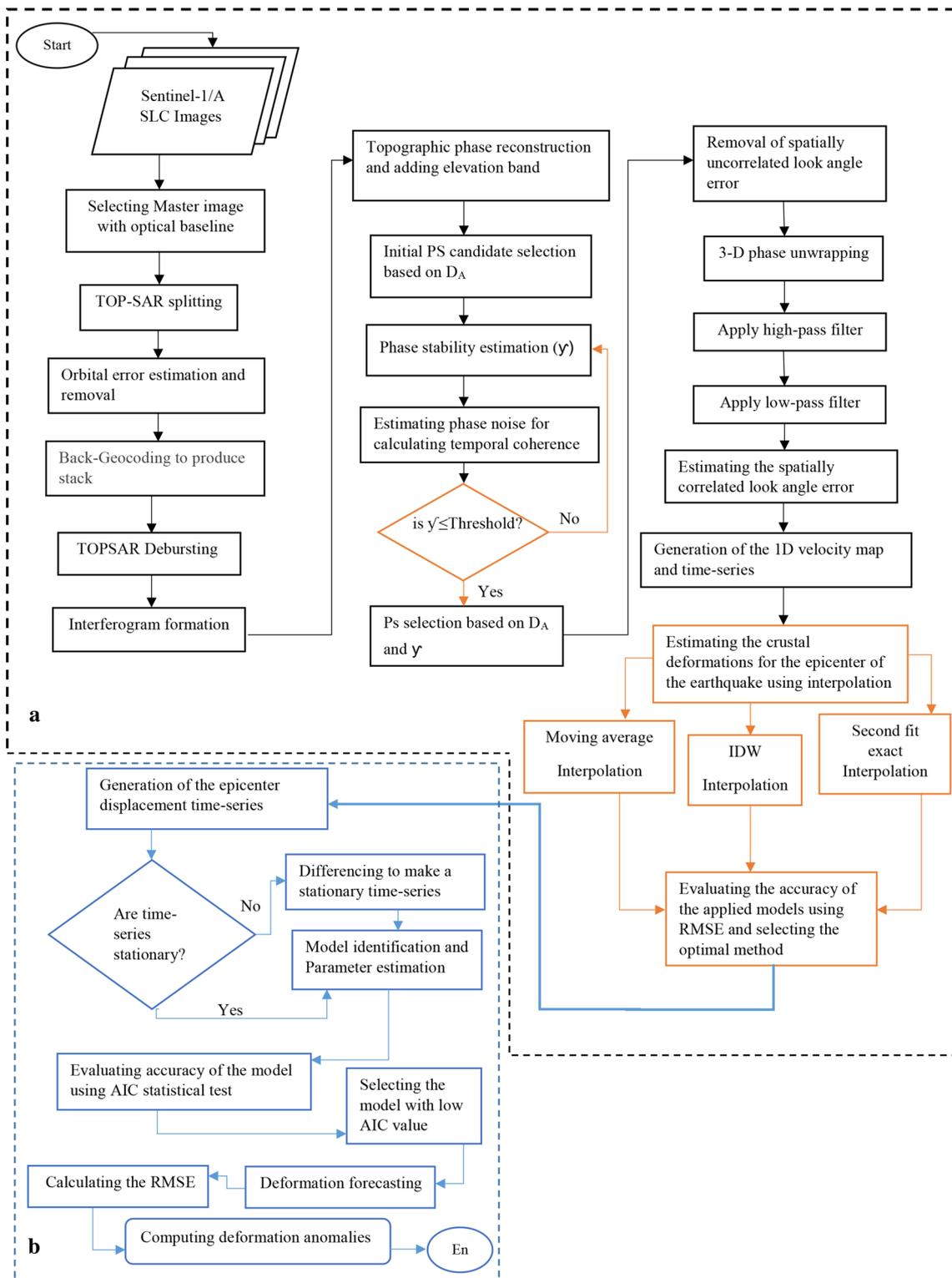
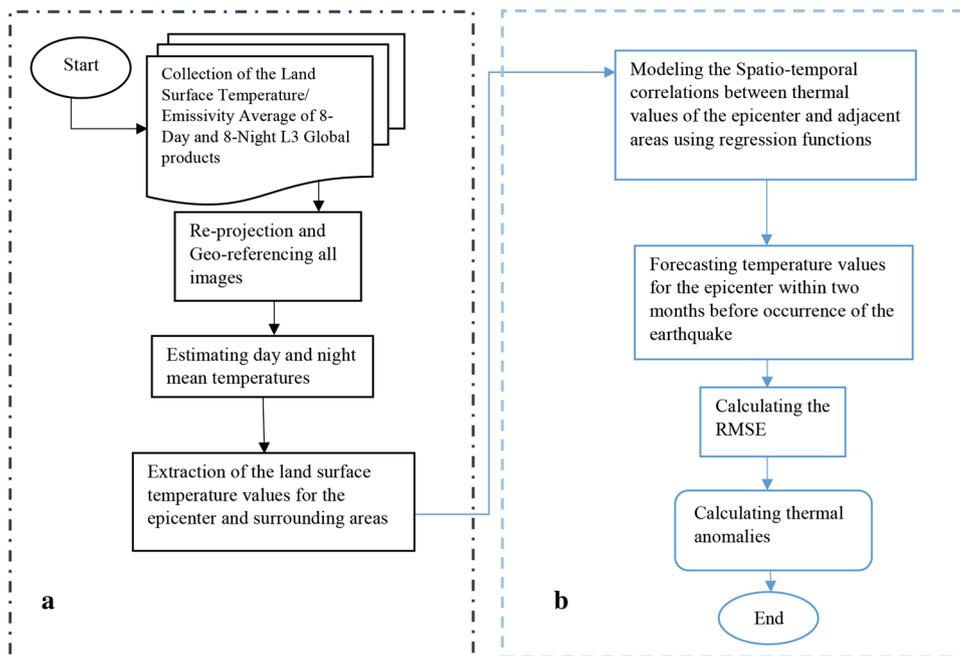


Fig. 2 Proposed algorithm for measuring surface displacement at the epicenter of the earthquake and calculating crustal deformation anomalies

Fig. 3 Proposed algorithm for calculating thermal anomalies at the earthquake epicenter



through time series is merely based on temporal dependency among observation and without considering spatial dependency among them.

$$z(t) = a_0 + a_1T_1 + a_2T_2 + \dots + a_pT_p \tag{2}$$

where a_0, a_1, \dots, a_p are the model coefficients; T_1, T_2, \dots, T_p represent the temperature values related to each area; and $Z(t)$ represents the calculated temperature for time t .

Implementation

Properties of the study area

According to a large number of earthquakes with magnitude lower than 5.0 and considering that each of them can be considered as a foreshock for large magnitude earthquakes, predicting the occurrence of an earthquake of magnitude lower than 5.0 is

not very valuable, both from a scientific and a practical point of view. On the other hand, considering that we implemented our proposed methods to monitor earthquake precursor anomalies on earthquakes with the same seismic properties and focal mechanism, to have an optimal distribution of sample regions, earthquakes with magnitude larger than 5.0 with nearly identical seismic properties from different regions of the world were selected. Table 1 presents the regional and seismic properties of each earthquake examined in the present study.

SOM results

As discussed in Sect. 2.1, in the present study thermal anomalies and crustal deformation anomalies have been monitored regarding the already happened earthquakes. But in this section results related to the implementation of the SOM algorithm to identify probable epicenters of the 2017 Kermanshah earthquake are presented (Figs. 4, 5).

Table 1 Properties of the earthquakes, extracted from the USGS Web site

Region	Focal geographic coordinates		Mag-nitude (Mw)	Depth (km)	Date	Time	Moment tensor
	Longitude	Latitude					
Iran, Kermanshah Province, 5 Km of Ezgeleh	45.90E	34.91N	7.3	19	2017/11/12	18:18:17(UTC)	Oblique thrust faulting
Mexico, 3km of San Pedro	97.97W	16.38N	7.2	22	2018/02/16	23:39:39(UTC)	Shallow-thrust faulting
Iran, Mashhad Province, 5 Km of Sefid Sang	60.43E	35.77N	6.1	13	2017/04/05	06:09:12(UTC)	Oblique thrust faulting
Iran, Bushehr Province, 4 Km of Kaki	51.61E	28.32N	5.5	10	2018/04/19	06:34:47(UTC)	Oblique thrust faulting

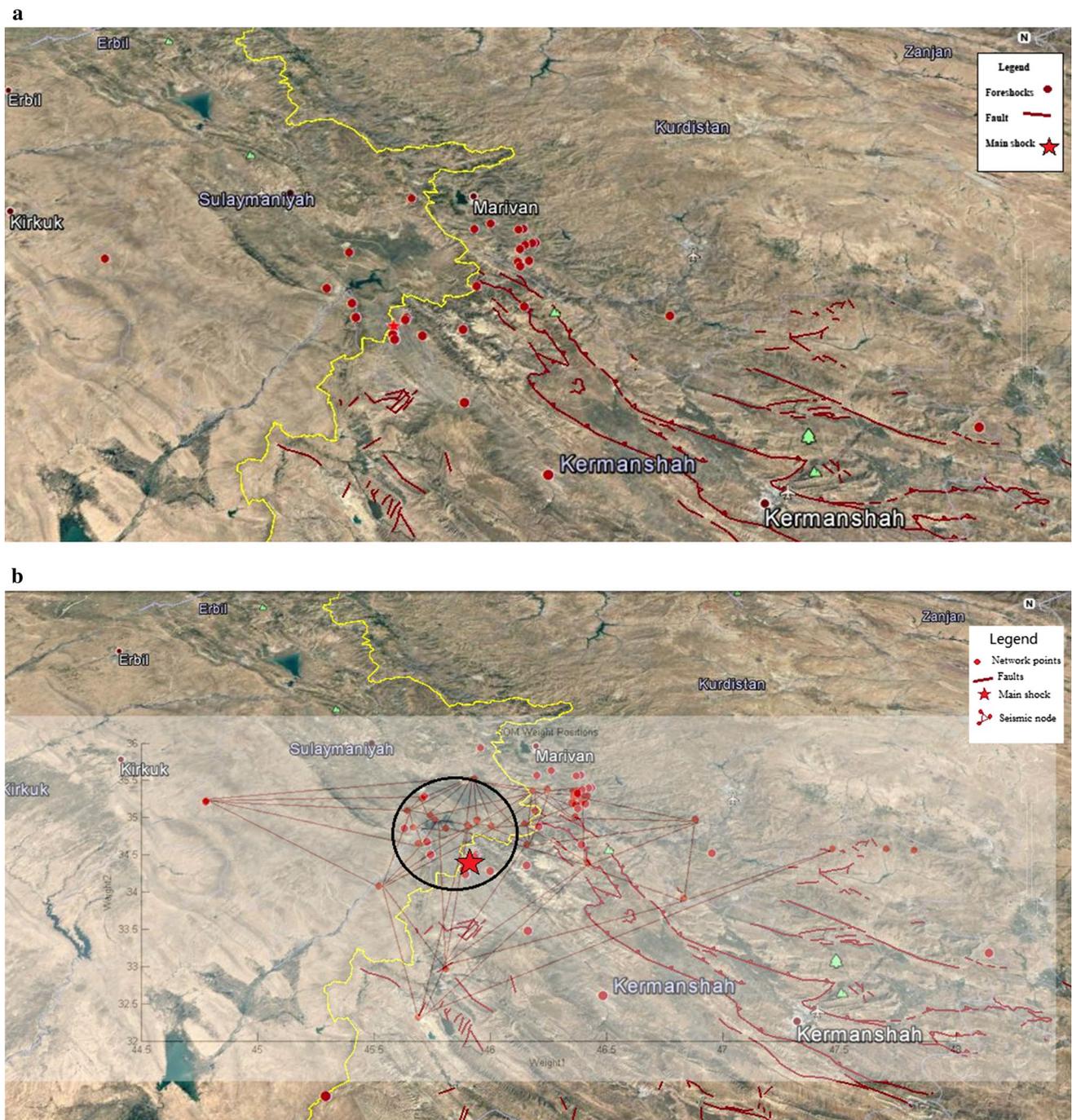


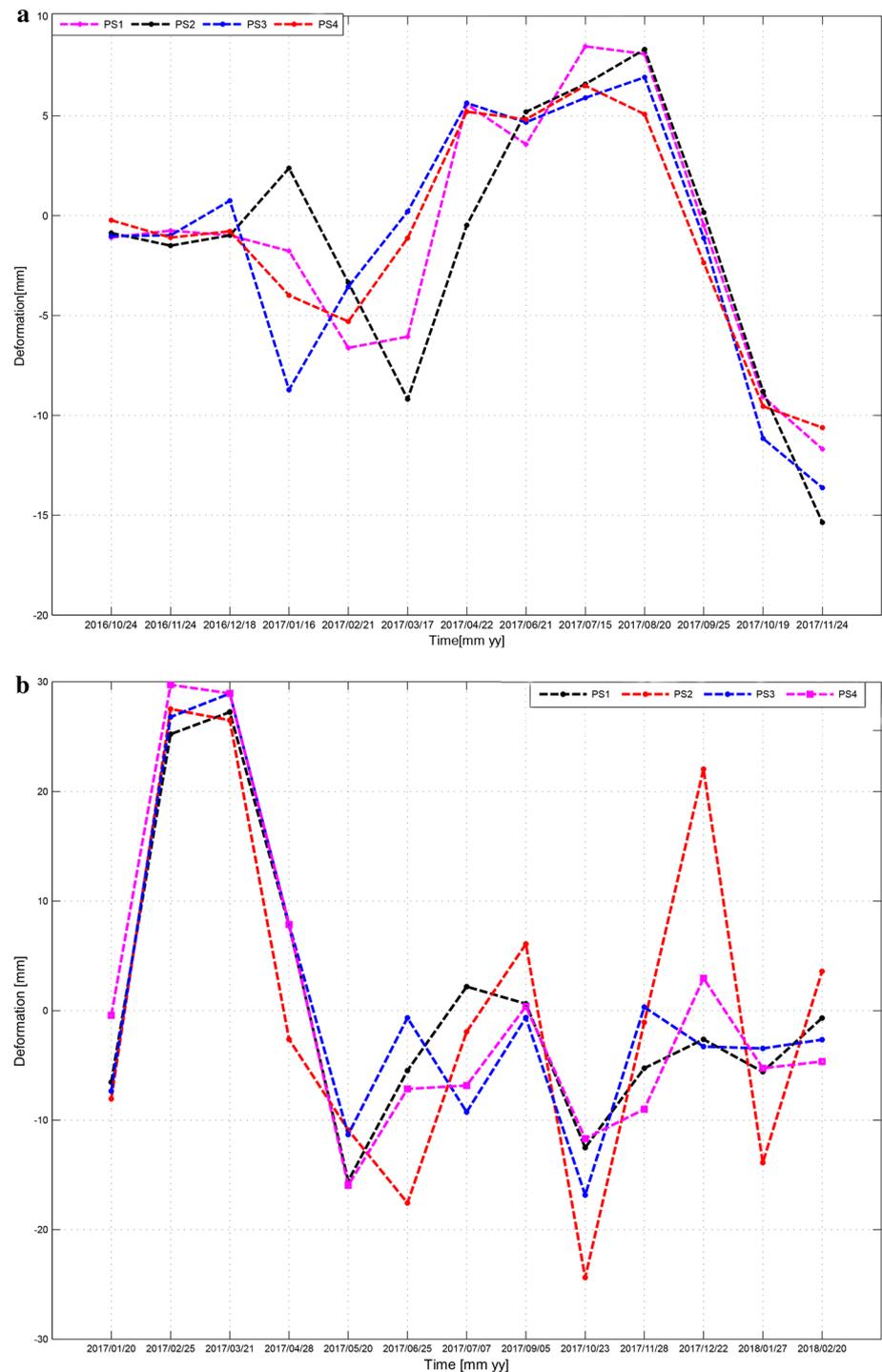
Fig. 4 **a** Topographic map of the Kermanshah Province. **b** Integration map of the SOM network output and topographic map of the Kermanshah Province

Determination of the crustal deformation anomalies at the earthquake epicenter

As discussed in Sect. 2.2, this study aimed at inspecting any significant relationship between crustal deformation

anomalies at the epicenter of the earthquakes and probability of an earthquake through radar interferometry, so Sentinel-1 SAR data during a period of 14 months before each earthquake were used. Tables 2, 3, 4 and 5 present the properties of the SAR imagery used for this study.

Fig. 5 **a** Time series of Line Of Sight (LOS) displacements for the selected PS points at the focal area of the Kermanshah Earthquake. **b** Time series of Line Of Sight (LOS) displacements for the selected PS points at the focal area of the Mexico Earthquake. **c** Time series of Line Of Sight (LOS) displacements for the selected PS points at the focal area of the Mashhad Earthquake. **d** Time series of Line Of Sight (LOS) displacements for the selected PS points at the focal area of the Bushehr Earthquake



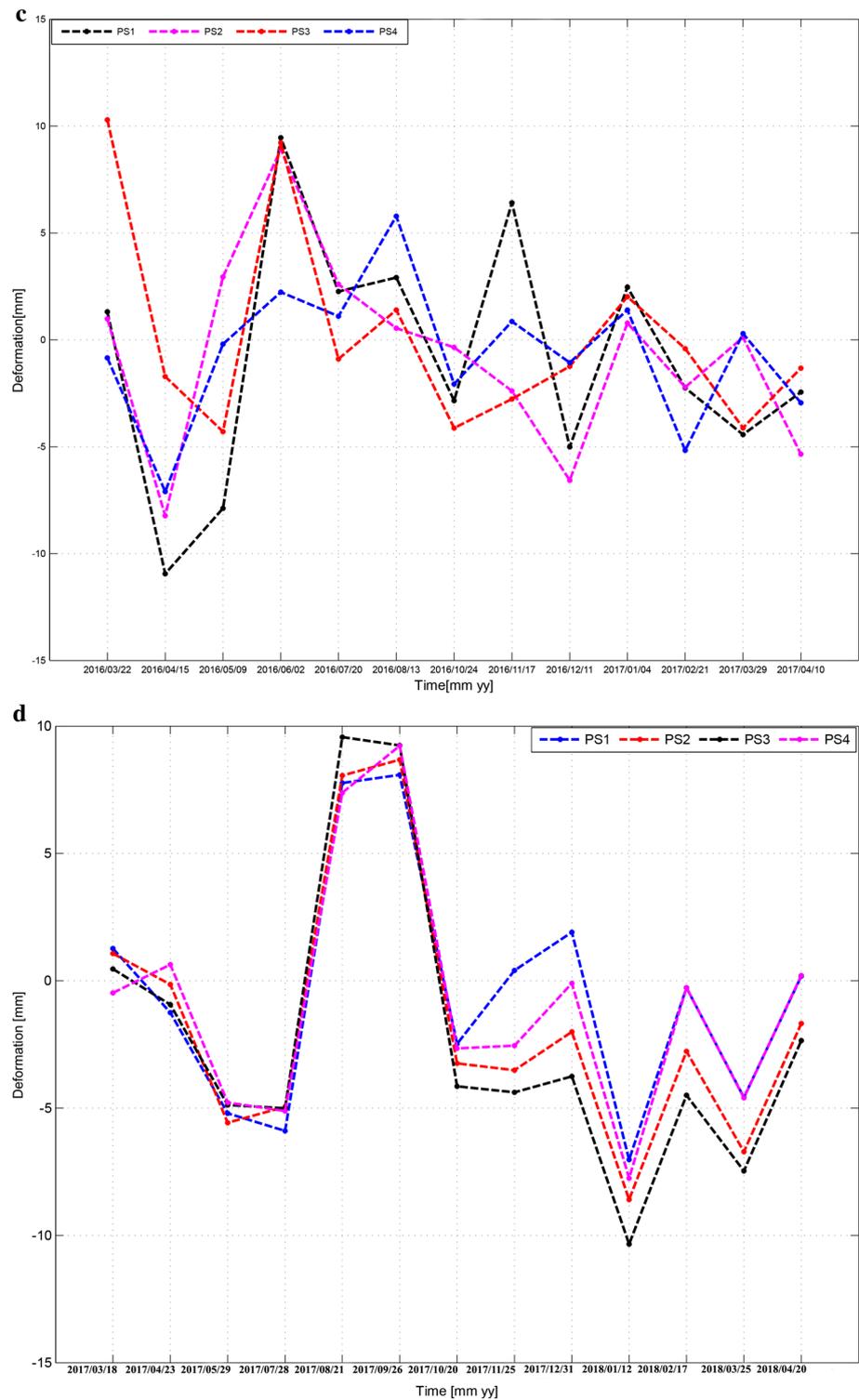
Initial results of the extended-PS method

Final results of the extended-PS-InSAR

For the purposes of the present study aiming at estimating reliable values for the crustal deformations at the epicenter of the earthquakes based on the estimated displacements for PS points around the epicenter, three methods of

interpolation were used including inverse distance weighted interpolation, moving average point interpolation and second fit exact interpolation. Then, for evaluating whether the accuracy of spatial sampling of the extended-PS method is adequate enough or not, obtained results from the above models were validated through RMSE statistical testing. By doing this, it will be possible to select the most accurate model so that surface deformations can be calculated for

Fig. 5 (continued)



the epicenter of the earthquake. Table 6 presents the results of the accuracy of each method. By comparing all results, inverse distance weighting method was identified as the most desirable approach. Thus, it was used to extract land surface displacements at the epicenter of the earthquakes. Figure 6a–d illustrates these results.

Calculation of the crustal deformation anomalies using data mining approaches

As already described in Sect. 2.2.2, in order to perform behavioral analysis on epicenter deformation time series and calculate crustal deformation anomalies for earthquakes

Table 2 Properties of the SAR images used for Kermanshah region

Row	Satellite platform	Product type	Instrument mode	Time period	Orbit	Geometry
1	S1-A	SLC	IW	2016/10/24	13451	Descending
2	S1-A	SLC	IW	2016/11/24	14078	Descending
3	S1-A	SLC	IW	2016/12/18	14428	Descending
4	S1-A	SLC	IW	2017/01/16	14851	Descending
5	S1-A	SLC	IW	2017/02/21	15376	Descending
6	S1-A	SLC	IW	2017/03/17	15726	Descending
7	S1-A	SLC	IW	2017/04/22	16251	Descending
8	S1-A	SLC	IW	2017/05/16	16601	Descending
9	S1-A	SLC	IW	2017/06/21	17126	Descending
10	S1-A	SLC	IW	2017/07/15	17476	Descending
11	S1-A	SLC	IW	2017/08/20	18001	Descending
12	S1-A	SLC	IW	2017/09/25	18526	Descending
13	S1-A	SLC	IW	2017/10/19	18876	Descending
14	S1-A	SLC	IW	2017/11/24	19401	Descending

Table 3 Properties of the SAR images used for Mexico region

Row	Satellite platform	Product type	Instrument mode	Time period	Orbit	Geometry
1	S1-A	SLC	IW	2017/01/20	17190	Descending
2	S1-A	SLC	IW	2017/02/25	15440	Descending
3	S1-A	SLC	IW	2017/03/21	15790	Descending
4	S1-A	SLC	IW	2017/04/26	16315	Descending
5	S1-A	SLC	IW	2017/05/20	16665	Descending
6	S1-A	SLC	IW	2017/06/25	17190	Descending
7	S1-A	SLC	IW	2017/07/07	17365	Descending
8	S1-A	SLC	IW	2017/08/12	17890	Descending
9	S1-A	SLC	IW	2017/09/05	18240	Descending
10	S1-A	SLC	IW	2017/10/23	18940	Descending
11	S1-A	SLC	IW	2017/11/28	19465	Descending
12	S1-A	SLC	IW	2017/12/22	19815	Descending
13	S1-A	SLC	IW	2018/01/27	20340	Descending
14	S1-A	SLC	IW	2018/02/20	20690	Descending

epicenter, the ARIMA model with different orders of p and q was fitted to PS time series. Tables 7, 8, 9 and 10 present the modeling results, and Fig. 7a–d illustrates the crustal deformation anomalies for each region.

Estimation of the thermal anomalies at the epicenter of the earthquakes before the occurrence

For the purpose of finding out any significant relationship between behavioral variations of the thermal precursor and earthquake occurrence, considering the fact that maximum interval of this precursor has been reported in the range of 30–40 days before the earthquake, in this

research, for each study area MODIS/Terra (MOD11A2) Land Surface Temperature/Emissivity average of 8-Day and 8-Night products were used for 2 months before the earthquake occurrence.

As already discussed in Sect. 2.3, in the present study, surface temperature variations at the epicenter were modeled based on temperatures of areas around the focal point using the proposed regression model (Eq. 2), during a period of 2 months before the earthquake. Finally, by comparing the estimated values of the model and the actual temperature available for the epicenter, thermal anomalies were calculated. Figure 8a–d illustrates the results of thermal anomalies.

Table 4 Properties of the SAR images used for Mashhad region

Row	Satellite platform	Product type	Instrument mode	Time period	Orbit	Geometry
1	S1-A	SLC	IW	2016/03/22	10483	Ascending
2	S1-A	SLC	IW	2016/04/15	10833	Ascending
3	S1-A	SLC	IW	2016/05/09	11183	Ascending
4	S1-A	SLC	IW	2016/06/02	11533	Ascending
5	S1-A	SLC	IW	2016/07/20	12233	Ascending
6	S1-A	SLC	IW	2016/08/13	12583	Ascending
7	S1-A	SLC	IW	2016/09/06	12933	Ascending
8	S1-A	SLC	IW	2016/10/24	13633	Ascending
9	S1-A	SLC	IW	2016/11/17	13983	Ascending
10	S1-A	SLC	IW	2016/12/11	14333	Ascending
11	S1-A	SLC	IW	2017/01/04	14683	Ascending
12	S1-A	SLC	IW	2017/02/21	15383	Ascending
13	S1-A	SLC	IW	2017/03/29	15908	Ascending
14	S1-A	SLC	IW	2017/04/10	16083	Ascending

Table 5 Properties of the SAR images used for Bushehr region

Row	Satellite platform	Product type	Instrument mode	Time period	Orbit	Geometry
1	S1-A	SLC	IW	2017/03/18	15748	Ascending
2	S1-A	SLC	IW	2017/04/23	16273	Ascending
3	S1-A	SLC	IW	2017/05/29	16798	Ascending
4	S1-A	SLC	IW	2017/06/22	17148	Ascending
5	S1-A	SLC	IW	2017/07/28	17673	Ascending
6	S1-A	SLC	IW	2017/08/21	18023	Ascending
7	S1-A	SLC	IW	2017/09/26	18548	Ascending
8	S1-A	SLC	IW	2017/10/20	18898	Ascending
9	S1-A	SLC	IW	2017/11/25	19423	Ascending
10	S1-A	SLC	IW	2017/12/31	19948	Ascending
11	S1-A	SLC	IW	2018/01/12	20123	Ascending
12	S1-A	SLC	IW	2018/02/17	20648	Ascending
13	S1-A	SLC	IW	2018/03/25	21173	Ascending
14	S1-A	SLC	IW	2018/04/20	21523	Ascending

Table 6 Accuracy of the applied interpolation methods for calculating crustal deformation at the epicenter of the earthquakes

Interpolation method	Modeling RMSE (mm) for Kermanshah earthquake	Modeling RMSE (mm) for Mexico earthquake	Modeling RMSE (mm) for Mashhad earthquake	Modeling RMSE (mm) for Bushehr earthquake
Linear second exact fit interpolation	± 8.55 mm	± 6.35 mm	± 5.55 mm	± 6.55 mm
Moving average interpolation	± 4.58 mm	± 3.28 mm	± 4.53 mm	± 2.58 mm
Inverse distance weighted interpolation	± 1.69 mm	± 1.23 mm	± 1.75 mm	± 0.69 mm

Discussion and Interpretation of the results

For the purposes of the present study, dealing with finding out a correlation between crustal deformation anomalies

and thermal anomalies with an earthquake occurrence in the focal area is initially carried out in a unidimensional mode and then in an integrated form in order to offer a more reliable method to predict probable magnitude and

Table 7 Computational process of the crustal deformation anomalies of Kermanshah earthquake

Samples period	Monthly deformations (mm)	Stationary deformations (mm)	Fitted results (mm)	Residuals (mm)	Forecasted results (mm)
2016/10/24	−0.2	*	*	*	
2016/11/24	−3.0	−2.8	−0.14	−2.65	
2016/12/24	−4.9	−1.9	0.66	−2.55	
2017/01/24	−4.8	0.1	3.45	−3.35	
2017/02/24	−4.5	0.3	3.59	−3.29	
2017/03/24	4.1	8.6	4.07	4.52	
2017/04/24	5.1	1.0	2.07	−1.07	
2017/05/24	5.1	0.0	−1.41	1.41	
2017/06/24	7.1	2.0	1.46	0.53	
2017/07/24	7.5	0.4	0.12	0.27	
2017/08/24	6.1	−1.4	0.74	−2.14	
2017/09/24	0.7				1.51
2017/10/24	−7.5				2.49
2017/11/24	−12.5				1.14

Table 8 Computational process of the crustal deformation anomalies of Mexico earthquake

Samples' period	Monthly deformations (mm)	Stationary deformations (mm)	Fitted results (mm)	Residuals (mm)	Forecasted results (mm)
2017/01/20	−7.30	*	*	*	
2017/02/20	22.12	29.42	30.13	−0.71	
2017/03/20	27.00	4.88	8.37	−3.49	
2017/04/20	6.83	−20.17	−25.48	5.31	
2017/05/20	−14.38	−21.21	−22.79	1.58	
2017/06/20	−10.38	4.00	2.48	1.51	
2017/07/20	−2.35	8.03	20.65	−12.62	
2017/08/20	1.39	3.74	7.60	−3.86	
2017/09/20	−3.10	−4.49	−2.67	−1.81	
2017/10/20	−14.80	−11.70	−6.61	−5.08	
2017/11/20	−6.25	8.55	−1.28	9.83	
2017/12/20	9.38				9.73
2018/01/20	−15.3				−5.61
2018/02/20	1.39				−15.17

time of earthquake occurrence anomalies within earthquake-prone regions.

Interpretation of the detected crustal deformation anomalies

Figure 7a shows the crustal deformation anomalies of Kermanshah earthquake; the behavior of the anomalous structure shows that after observation subsidence within 3 months before the earthquake and by increasing the rate of this subsidence in the future months, an earthquake of magnitude of 7.3 occurs during a period of 19 days after observing the first subsidence. Behavioral analysis is performed on

crustal deformation anomalies related to another earthquake of magnitude of 7.2 and similar focal mechanism. As shown in Fig. 7b, the crustal deformation anomalies at the epicenter of the Mexico earthquake follow a half cycle of the sinusoidal waveform before the earthquake. For the earthquakes with magnitude of lower than 7.0 and larger than 5.0, as shown by crustal deformation anomalies of Mashhad earthquake (Fig. 7c), the behavior of the anomalous structure has been viewed as subsidence for 2 months before the earthquakes. while the behavioral variations of the crustal deformation anomalies of an earthquake with similar seismic properties, as shown in diagram d (Fig. 7), shows a different pattern of subsidence at the epicenter before the

Table 9 Computational process of the crustal deformation anomalies of Mashhad earthquake

Samples' period	Monthly deformations (mm)	Stationary deformations (mm)	Fitted results (mm)	Residuals (mm)	Forecasted results (mm)
2016/03/22	2.93	*	*	*	
2016/04/22	-5.78	-8.71	-3.89	-4.81	
2016/05/22	1.1	6.88	3.86	3.01	
2016/06/22	7.1	6.00	4.61	1.38	
2016/07/22	1.1	-6.00	-5.51	-0.48	
2016/08/22	1.2	0.10	-1.72	1.82	
2016/09/22	-0.82	-2.02	-0.75	-1.26	
2016/10/22	-2.1	-1.28	-1.75	0.47	
2016/11/22	-0.54	1.56	1.38	0.17	
2016/12/22	-2.9	-2.36	-0.90	-1.45	
2017/01/22	0.10	3.00	0.77	2.22	
2017/02/22	-2.20				0.34
2017/03/22	-2.60				-3.30
2017/04/22	-2.40				-0.07

Table 10 Computational process of the crustal deformation anomalies of Bushehr earthquake

Samples period	Monthly deformations (mm)	Stationary deformations (mm)	Fitted results (mm)	Residuals (mm)	Forecasted results (mm)
2017/03/18	0.36	*	*	*	
2017/04/18	-0.62	-0.98	1.30	-2.28	
2017/05/18	-4.22	-3.60	-0.63	-2.96	
2017/06/18	-5.01	-0.79	3.07	-3.86	
2017/07/18	-5.01	0.00	4.48	-4.48	
2017/08/18	8.50	13.51	6.05	7.45	
2017/09/18	8.90	0.40	0.70	-0.30	
2017/10/18	-3.61	-12.51	-11.80	-0.70	
2017/11/18	-3.90	-0.29	-0.84	0.55	
2017/12/18	-2.78	1.12	-0.94	2.06	
2018/01/18	-8.30	-5.52	-3.47	-2.04	
2018/02/18	-3.22				-3.42
2018/03/18	-6.10				1.20
2018/04/18	-2.40				-1.63

earthquake. Considering that the amounts of crustal deformation anomalies and their behaviors are different in earthquakes with a magnitude of larger than or equal to 7.0 as well as earthquakes with a magnitude of lower than or 7.0, we cannot deduce a general index for predicting the probable magnitude or time of earthquake occurrence.

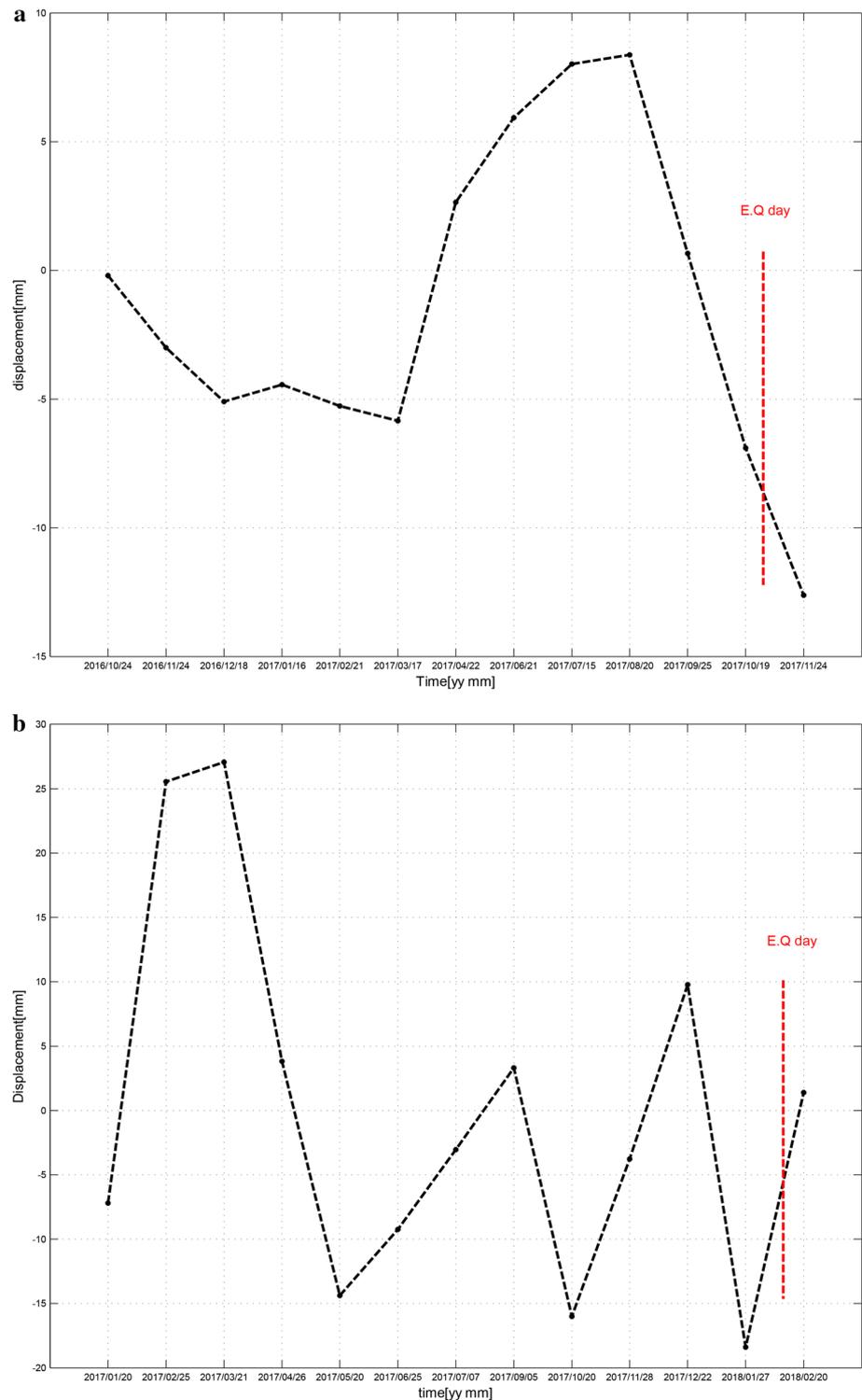
Interpretation of the detected thermal anomalies

In the case of interpretation of the thermal anomalies at the epicenter of the earthquakes, unlike previous works in this field, the present study by modeling spatiotemporal dependence of thermal variations using regression functions proved

that earthquake prediction based on the thermal anomalies will be unreliable by identifying the positive thermal anomalies and correlating them with the main fault of the earthquake merely, because the thermal anomalies before the earthquake may also have a negative values. In other words, an unusual increase or decrease in temperature variations of the probable epicenter can be considered as a thermal anomaly.

In the case of behavioral analysis on thermal anomalies, as shown in Fig. 8a, b, it is not possible to detect a consistent and constant pattern related to earthquake occurrence with a magnitude of 7.0 or higher because of irregularities seen in the behavior of thermal anomalies two months before the

Fig. 6 **a** Time series of line-of-sight (LOS) displacements at the epicenter of Kermanshah earthquake with the accuracy of ± 1.69 mm. **b** Time series of line-of-sight (LOS) displacements at the epicenter of Mexico earthquake with the accuracy of ± 1.23 mm. **c**: Time series of line-of-sight (LOS) displacements at the epicenter of Mashhad earthquake with the accuracy of ± 1.75 mm. **d** Time series of line-of-sight (LOS) displacements at the epicenter of Bushehr earthquake with the accuracy of ± 0.69 mm

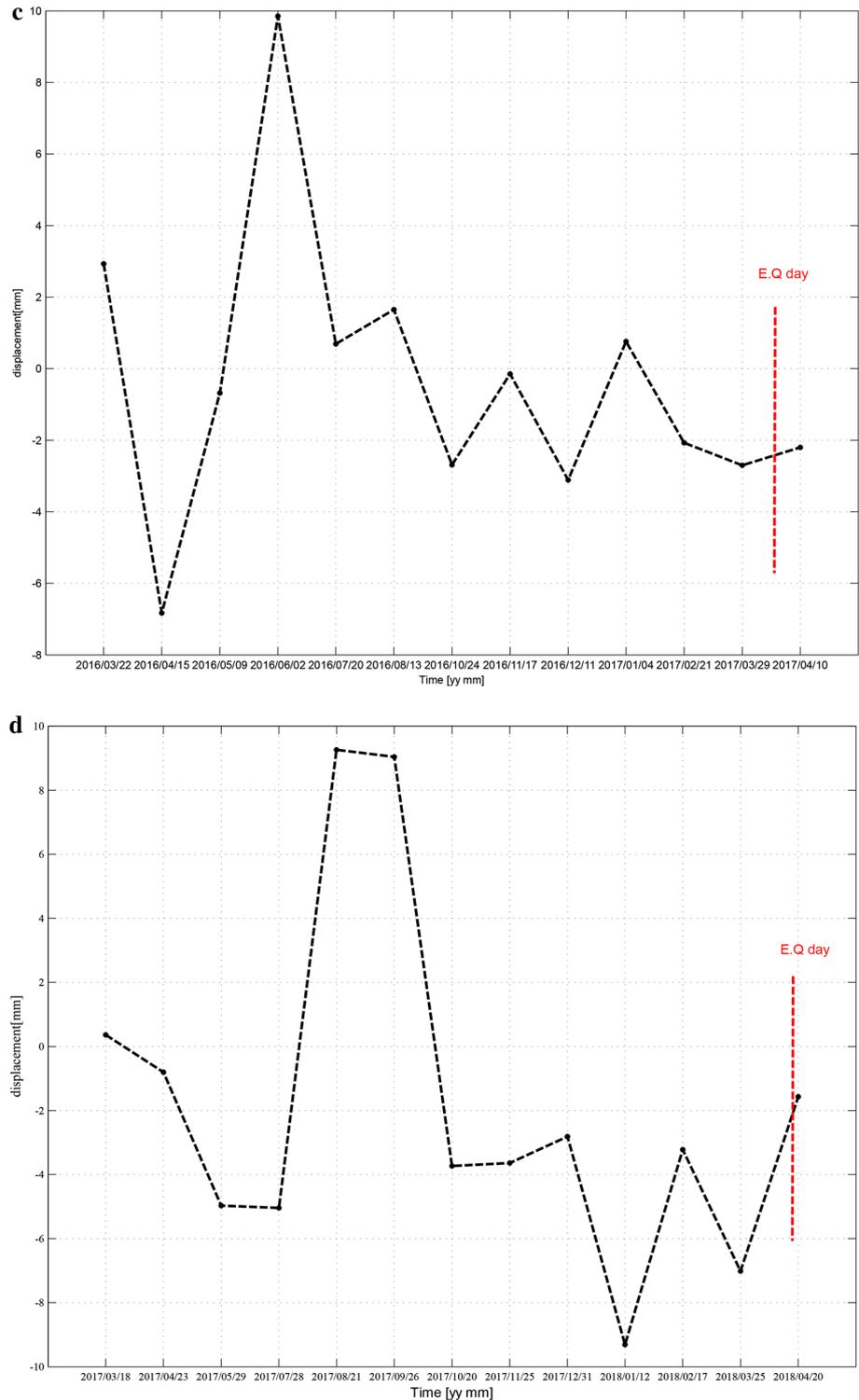


earthquake. Moreover, frequent irregularities in the variation rate of thermal anomalies estimated for earthquakes with a magnitude of lower than 7.0 and larger than 5.0 (Fig. 8c, d), prevent identification of the significant correlation between focal thermal anomalies and occurrence of an earthquake.

Integrated interpretation of detected thermal anomalies and crustal deformation anomalies at the epicenter of the earthquake

However, the absence of any significant correlation between calculated thermal and crustal deformation

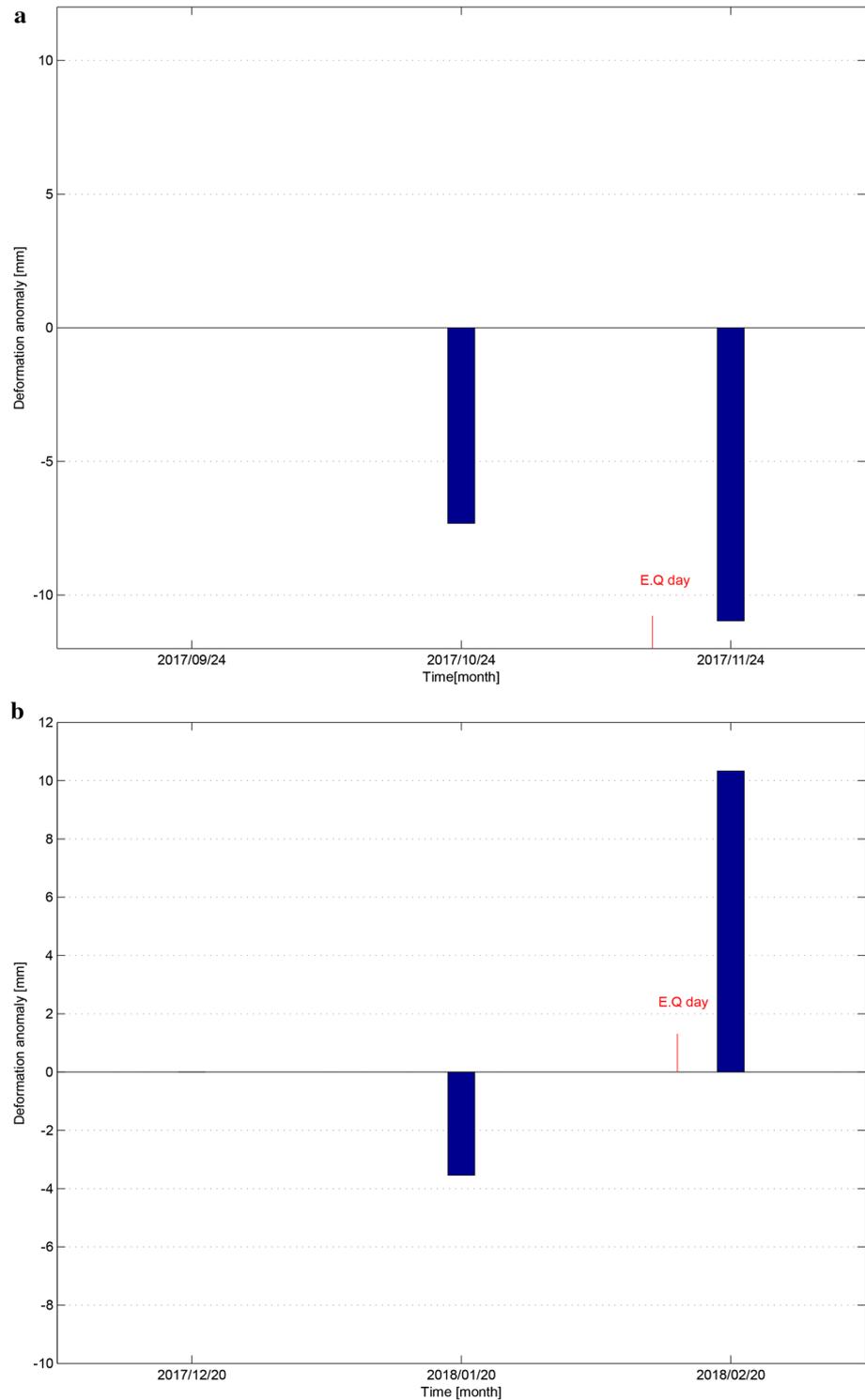
Fig. 6 (continued)



anomalies and earthquakes, in unidimensional mode, never underestimates precursors and estimated anomalies for each precursor. On the other hand, in the research process, after the extraction of crustal deformation anomalies and thermal anomalies before the earthquake occurrence, the temporal correlation analyzing between these two data

sets was analyzed. The results of the analyses showed that there is no one-to-one temporal correspondence between these two data sets, and the relationship between them is only related to the maximum amount of the anomalies. Accordingly, the use of spatiotemporal models was not effective. Therefore, in the present study, instead of

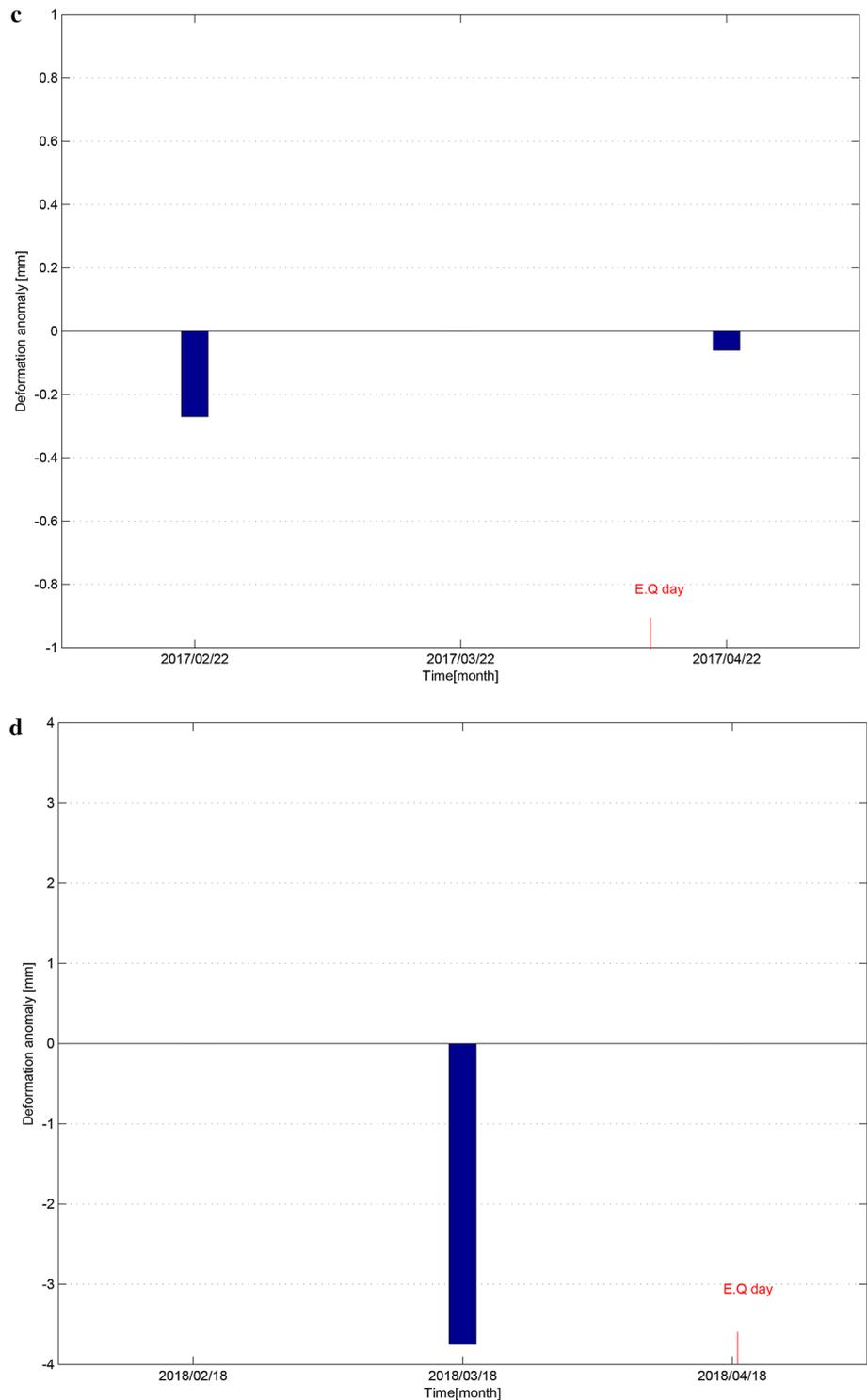
Fig. 7 **a** Crustal deformation anomalies at the epicenter of Kermanshah earthquake within 3 months before the occurrence. **b**: Crustal deformation anomalies at the epicenter of Mexico earthquake within 3 months before the occurrence. **c** Crustal deformation anomalies at the epicenter of Mashhad earthquake within 3 months before the occurrence. **d** Crustal deformation anomalies at the epicenter of Bushehr earthquake within 3 months before the occurrence



examining general and unidimensional behavior of the detected anomalies, maximum crustal deformation anomaly and maximum thermal anomaly at the earthquake's epicenter in a period of two months before earthquake based on their value and interval from the day of the event were extracted (as shown in Table 11). Then, a descriptive

and logical rule as earthquake prediction algorithm based on amount and time of observing maximum crustal deformation anomaly and the maximum thermal anomaly of earthquakes with oblique thrust faulting was proposed (as shown in Fig. 9).

Fig. 7 (continued)

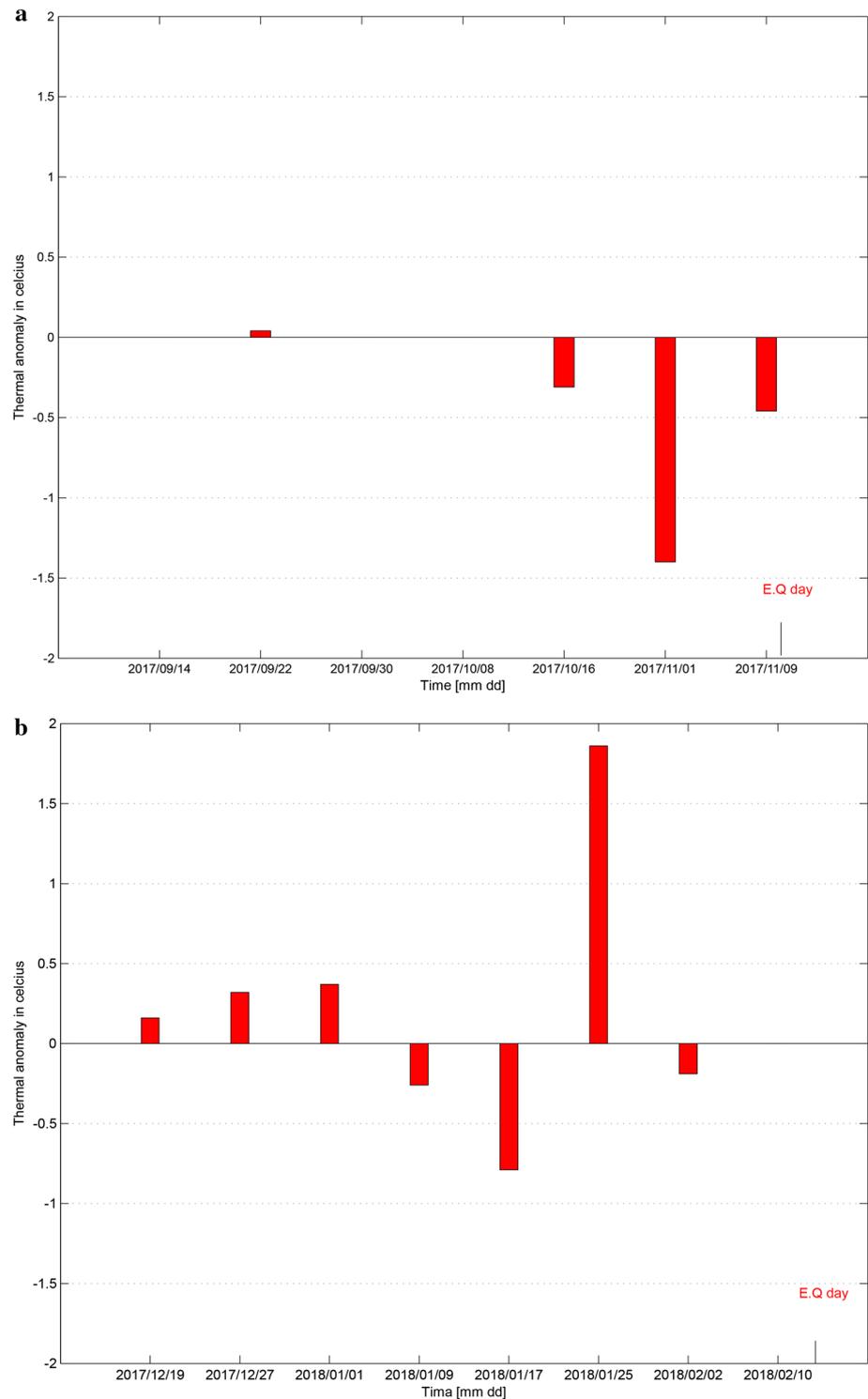


Evaluation of the efficiency of the proposed earthquake prediction algorithm for an earthquake with a different focal mechanism

In this section, the efficiency of the proposed algorithm in Sect. 4.3 was evaluated for an earthquake with a different

focal mechanism. We selected April 16, 2016, Japan-Kumamoto earthquake of magnitude 7.0 with strike-slip faulting for performing this test. Figure 10a–c illustrates the results.

Fig. 8 **a** Thermal anomalies during a period of 2 months at the epicenter of Kermanshah earthquake. **b** Thermal anomalies during a period of 2 months at the epicenter of Mexico earthquake. **c** Thermal anomalies during a period of 2 months at the epicenter of Mashhad earthquake. **d** Thermal anomalies during a period of 2 months at the epicenter of Bushehr earthquake

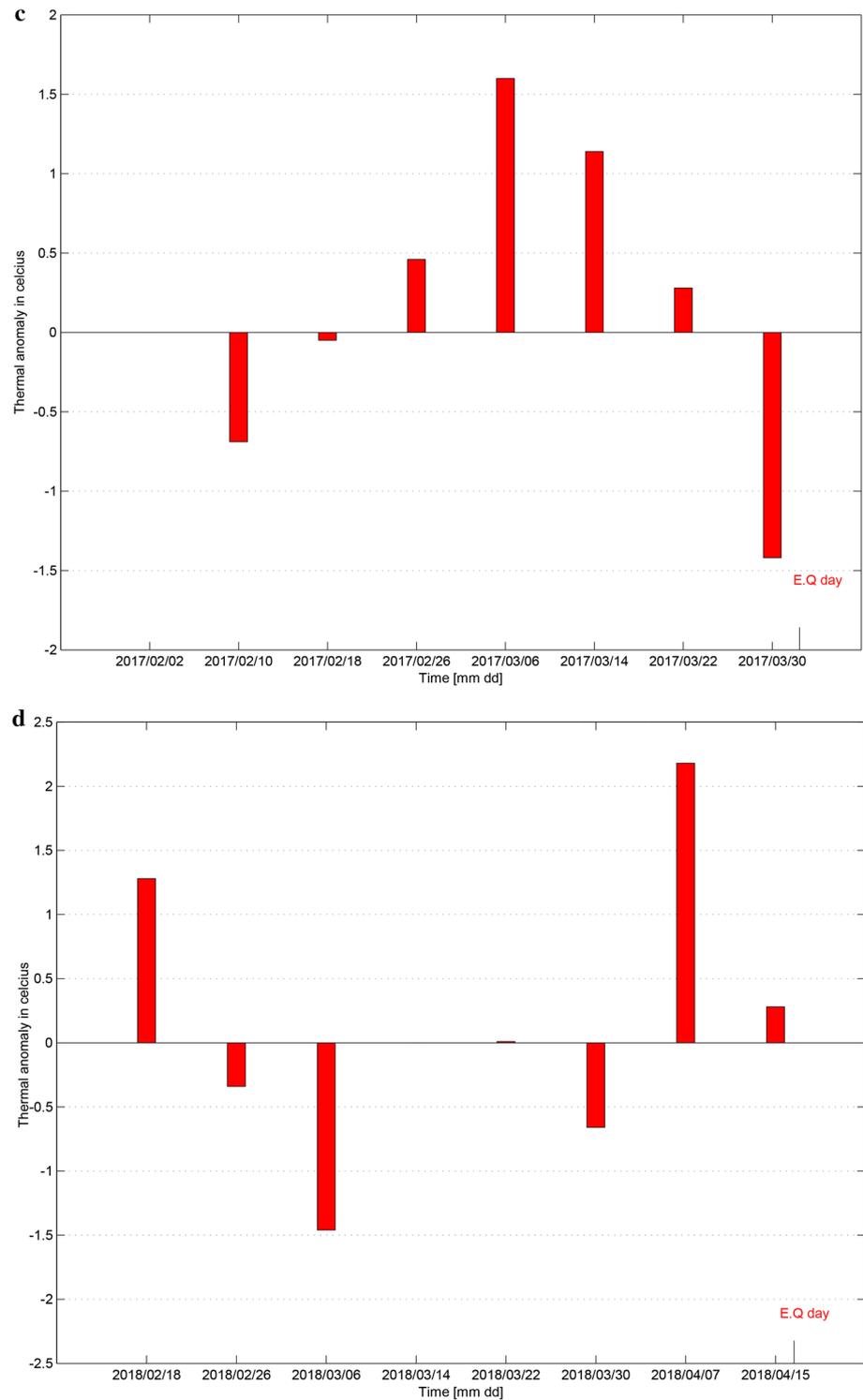


Conclusion

According to the obtained results from this study, thermal anomalies and crustal deformation anomalies at the epicenter of earthquakes with similar focal mechanism and similar seismic characteristics show different structural

patterns before earthquake occurrence. Therefore, making deterministic earthquake prediction based on individual monitoring of thermal anomalies in the focal area, as well as monitoring of the crustal deformation anomalies, is practically impossible. Also, during the research process, separate time series and regression analyses were

Fig. 8 (continued)



carried out on the extracted thermal anomalies and crustal deformation anomalies. Unfortunately, none of the above-mentioned methods resulted in a deterministic index for the prediction of an earthquake occurrence. Despite the fact that the existence of different structures in the crustal deformation anomalies and thermal anomalies in the

focal area prevented modeling unidimensional correlation between these precursors with earthquake occurrence, in this study, by considering the maximum amount of thermal anomalies and crustal deformation anomalies in a period of 2 months before earthquake occurrence, a descriptive and logical rule as an earthquake prediction algorithm in

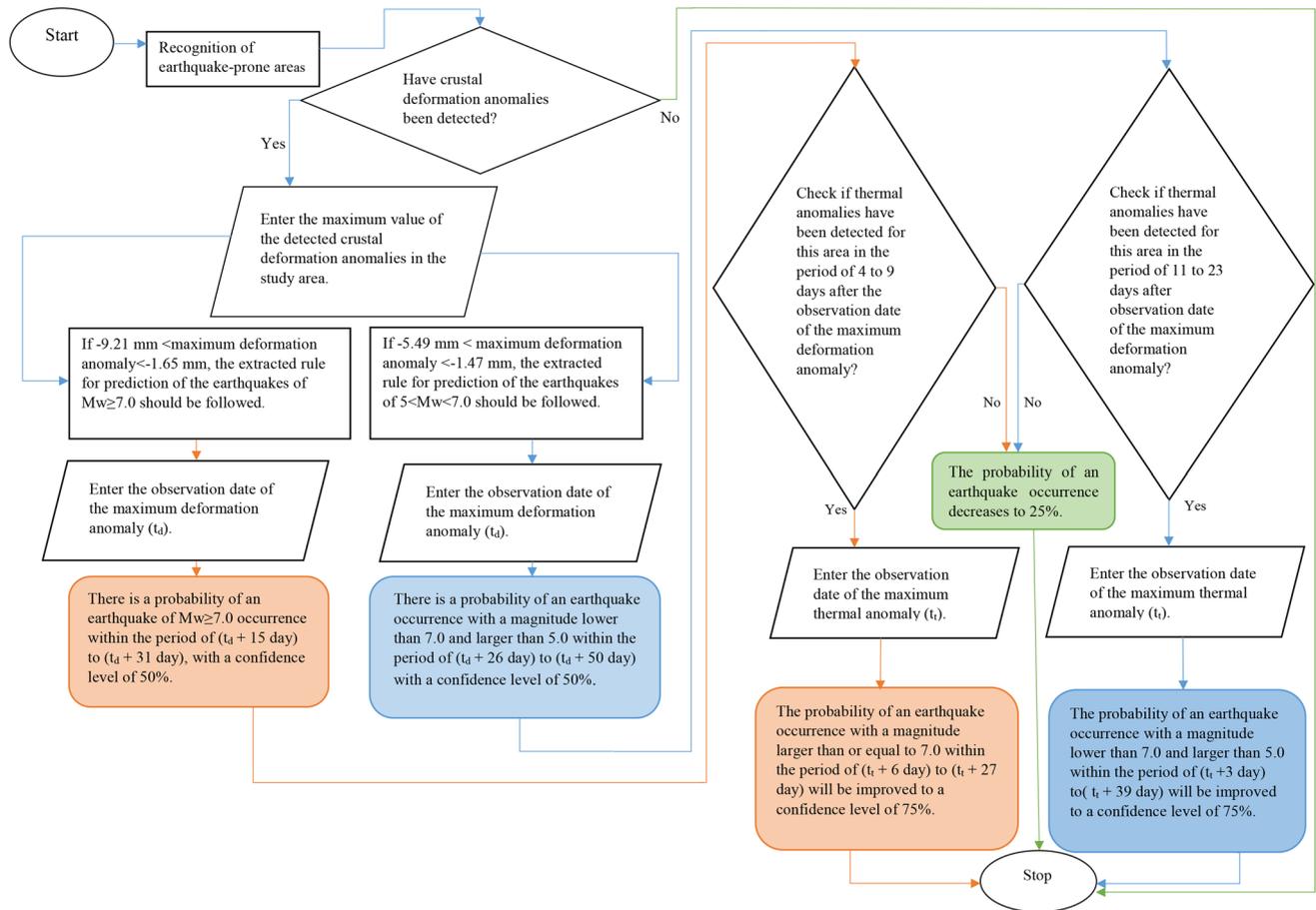


Fig. 9 Proposed algorithm to investigate the probability of an earthquake occurrence based on observing maximum anomalies of crustal deformation and surface temperature of earthquakes with oblique thrust faulting

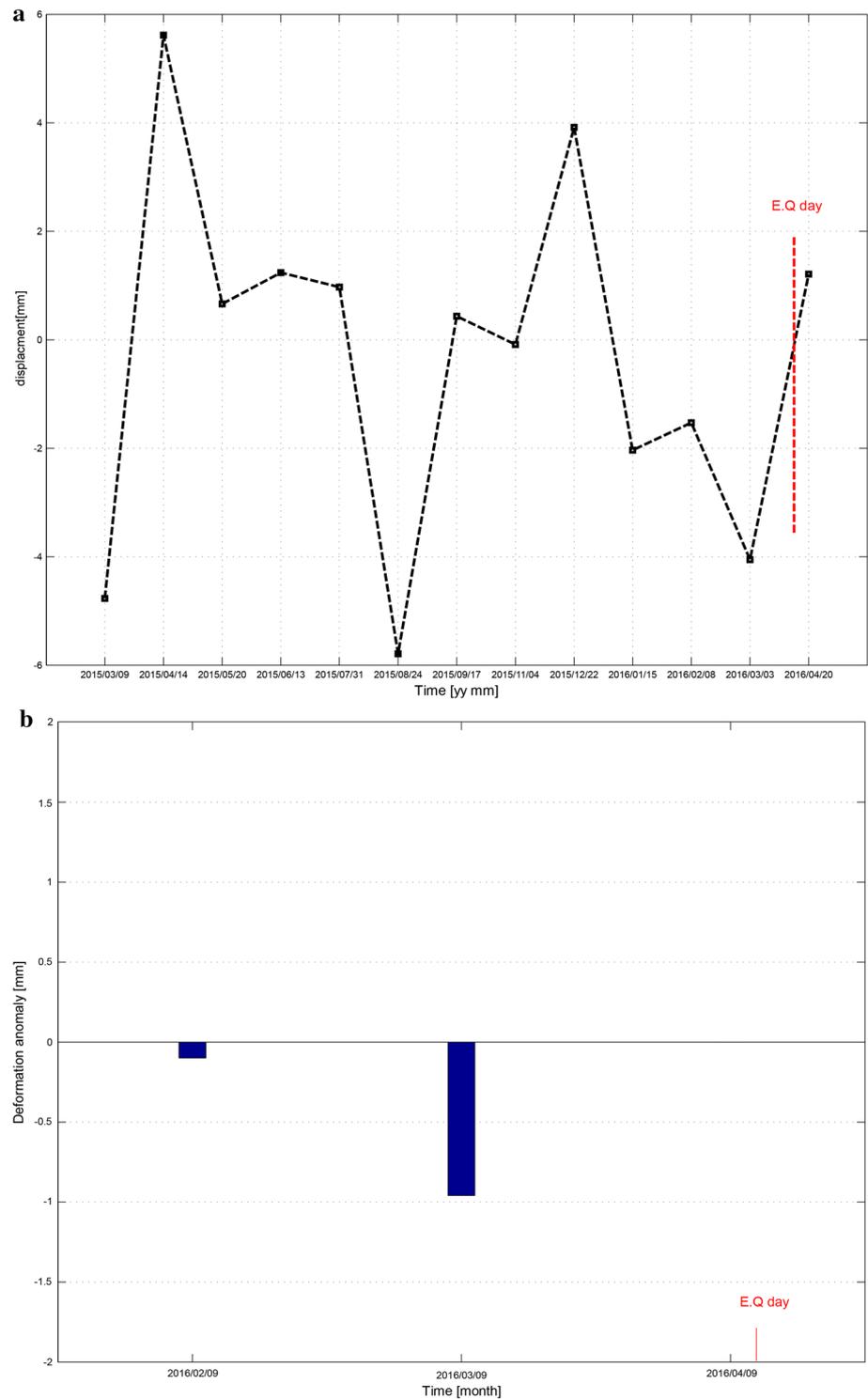
Table 11 Maximum amounts of thermal anomalies and crustal deformation anomalies at the epicenter of the earthquakes within two months before the earthquake

Region	Maximum value of the crustal deformation anomaly within 2 months before the earthquake occurrence (mm)	Time interval between the maximum crustal deformation anomaly and the earthquake day (day)	Maximum value of the thermal anomaly within 2 months before the earthquake occurrence (c°)	Time interval between the maximum thermal anomaly and the earthquake day (day)
Iran, Kermanshah Province, 5 km of Ezgeleh	- 7.32	19	- 1.40	11
Mexico, 3 km of San Pedro	- 3.54	27	+ 1.86	22
Iran, Mashhad Province, 5 km of Sefid Sang	- 0.27	44	+ 1.61	30
Iran, Bushehr Province, 4 km of Kaki	- 3.75	32	+ 2.18	12

the case of earthquakes with oblique thrust faulting was proposed (see Fig. 9). Since the investigated earthquakes for the purposes of this study were oblique thrust earthquakes, the validity of the proposed algorithm was evaluated for an earthquake with a different focal mechanism.

The analysis results of the thermal anomalies and crustal deformation anomalies at the epicenter of April 16, 2016, Japan-Kumamoto earthquake of magnitude 7.0 with strike-slip faulting, showed completely different trends than the suggested patterns by the proposed algorithm. So the

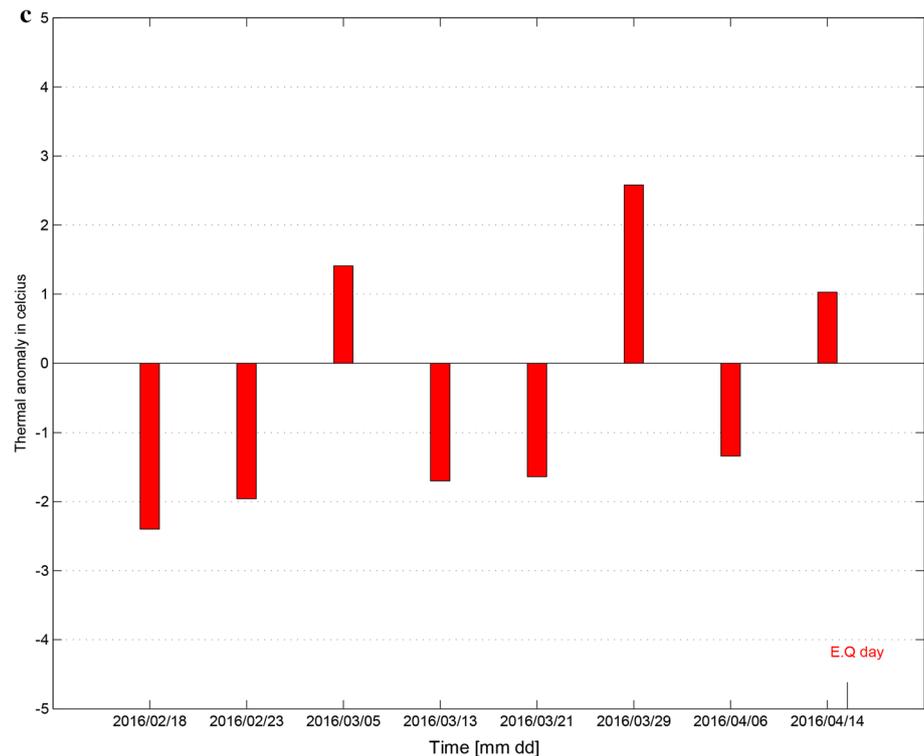
Fig. 10 **a** Time series of line-of-sight (LOS) displacements at the epicenter of Japan-Kumamoto earthquake with the accuracy of ± 1.23 mm. **b** Crustal deformation anomalies at the epicenter of Japan-Kumamoto earthquake within 3 months before the earthquake. **c** Thermal anomalies during a period of 2 months at the epicenter of Japan-Kumamoto earthquake



proposed algorithm was unable to determine the probable time and magnitude of an earthquake occurrence with strike-slip faulting. Therefore, it is recommended that, in

order to complete the proposed prediction algorithm by this study, its performance must be tested with more study samples.

Fig. 10 (continued)



References

- Akhoondzadeh M, Parrot M, Saradjian M (2010) Electron and ion density variations before strong earthquakes ($M \geq 6.0$) using DEMETER and GPS data. *Nat Hazards Earth Syst Sci* 10:7–18
- Bamler R, Hartl P (1998) Synthetic aperture radar interferometry. *Inverse Prob* 14(4):R1
- Brockwell PJ, Davis RA, Fienberg SE (1991) *Time-series: theory and methods*. Springer, Berlin
- Bürgmann R, Rosen PA, Fielding EJ (2000) Synthetic aperture radar interferometry to measure Earth's surface topography and its deformation. *Annu Rev Earth Planet Sci* 28(1):169–209
- Calais E, Vergnolle M, San'Kov V et al (2003) GPS measurements of crustal deformation in the Baikal-Mongolia area (1994–2002): implications for current kinematics of Asia. *J Geophys Res Solid Earth* 108:2501
- Gabriel AK, Goldstein RM, Zebker HA (1989) Mapping small elevation changes over large areas: differential radar interferometry. *J Geophys Res Solid Earth* 94(B7):9183–9191
- Graham LC (1974) Synthetic interferometer radar for topographic mapping. *Proc IEEE* 62(6):763–768
- Grant RA, Raulin JP, Freund FT (2015) Changes in animal activity prior to a major ($M=7$) earthquake in the Peruvian Andes. *Phys Chem Earth Parts A/B/C* 85:69–77
- Guo G, Wang B (2008) Cloud anomaly before Iran earthquake. *Int J Remote Sens* 29:1921–1928
- Han P, Hattori K, Huang Q et al (2011) Evaluation of ULF electromagnetic phenomena associated with the 2000 Izu Islands earthquake swarm by wavelet transform analysis. *Nat Hazards Earth Syst Sci* 11:965–970
- Hand DJ (2007) Principles of data mining. *Drug Saf* 30:621–622
- Hayakawa M (2013) Possible electromagnetic effects on abnormal animal behavior before an earthquake. *Animals* 3:19–32
- Hooper A (2008) A multi-temporal InSAR method incorporating both persistent scatterer and small baseline approaches. *Geophys Res Lett* 35(16):16302
- Huang J, Mao F, Zhou W, Zhu X (2008) Satellite thermal IR associated with Wenchuan earthquake in China using MODIS data. In: Paper presented at the proceeding of the 14th world conference on earthquake engineering.
- Huang F, Li M, Ma Y et al (2017) Studies on earthquake precursors in China: a review for recent 50 years. *Geod Geodyn* 8:1–12
- Ingebritsen SE, Manga M (2014) Earthquakes: hydrogeochemical precursors. *Nat Geosci* 7:697
- Li J, Heap AD (2014) Spatial interpolation methods applied in the environmental sciences: a review. *Environ Modell Softw* 53:173–189
- Lippiello E, Marzocchi W, de Arcangelis L, Godano C (2012) Spatial organization of foreshocks as a tool to forecast large earthquakes. *Sci Rep* 2:846
- Lu X, Meng Q, Gu X et al (2016) Thermal infrared anomalies associated with multi-year earthquakes in the Tibet region based on China's FY-2E satellite data. *Adv Space Res* 58:989–1001
- Massonnet D, Feigl KL (1998) Radar interferometry and its application to changes in the Earth's surface. *Rev Geophys* 36:441–500
- Massonnet D, Rossi M, Carmona C, Adragna F, Peltzer G, Feigl K, Rabaute T (1993) The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* 364(6433):138
- Mohankumar K, Sangeetha K (2018) A study on earthquake prediction using neural network algorithms
- Molchanov O, Rozhnoi A, Solovieva M et al (2006) Global diagnostics of the ionospheric perturbations related to the seismic activity using the VLF radio signals collected on the DEMETER satellite. *Nat Hazards Earth Syst Sci* 6:745–753

- Moreno M, Rosenau M, Oncken O (2010) 2010 Maule earthquake slip correlates with pre-seismic locking of Andean subduction zone. *Nature* 467:198
- Moro M, Saroli M, Stramondo S et al (2017) New insights into earthquake precursors from InSAR. *Sci Rep* 7:12035
- Ogata Y, Katsura K (2012) Prospective foreshock forecast experiment during the last 17 years. *Geophys J Int* 191:1237–1244
- Parsons, B. (2017) From geodesy to tectonics: observing earthquake processes from space (Augustus Love Medal Lecture). In: Paper presented at the EGU general assembly conference abstracts.
- Peacock S, Crampin S, Booth DC, Fletcher JB (1988) Shear wave splitting in the Anza seismic gap, southern California: temporal variations as possible precursors. *J Geophys Res Solid Earth* 93:3339–3356
- Peltzer G, Rosen P (1995) Surface displacement of the 17 May 1993 Eureka Valley, California, earthquake observed by SAR interferometry. *Science* 268(5215):1333–1336
- Pio Lucente F, De Gori P, Margheriti L et al (2010) Temporal variation of seismic velocity and anisotropy before the 2009 MW 6.3 L'Aquila earthquake Italy. *Geology* 38:1015–1018
- Pohn H, Offield T, Watson K (1974) Thermal inertia mapping from satellite-discrimination of geologic units in Oman. *J Res US Geol Surv* 2:147–158
- Pulinets S, Davidenko D (2014) Ionospheric precursors of earthquakes and global electric circuit. *Adv Space Res* 53:709–723
- Pulinets S, Ouzounov D, Karelin A, Boyarchuk K, Pokhmelnikh L (2006) The physical nature of thermal anomalies observed before strong earthquakes. *Phys Chem Earth Parts A/B/C* 31:143–153
- Qiang Z-J, Kong L-C, Zheng L-Z et al (1997) An experimental study on temperature increasing mechanism of satellitic thermo-infrared. *Acta Seismol Sin* 10:247–252
- Rikitake T (1968) Earthquake prediction. *Earth Sci Rev* 4:245–282
- Rikitake T (1975) Earthquake precursors. *Bull Seismol Soc Am* 65:1133–1162
- Rosen PA, Hensley S, Joughin IR, Li F, Madsen SN, Rodriguez E, Goldstein RM (1998) Synthetic aperture radar interferometry. *Proceed IEEE* 88(3):333–382
- Saradjian M, Akhoondzadeh M (2011) Thermal anomalies detection before strong earthquakes ($M \geq 6.0$) using interquartile, wavelet and Kalman filter methods. *Nat Hazards Earth Syst Sci* 11:1099–1108
- Saraf AK, Rawat V, Choudhury S, Dasgupta S, Das J (2009) Advances in understanding of the mechanism for generation of earthquake thermal precursors detected by satellites. *Int J Appl Earth Obs Geoinf* 11:373–379
- Spencer JR, Tamppari LK, Martin TZ, Travis LD (1999) Temperatures on Europa from Galileo photopolarimeter-radiometer: nighttime thermal anomalies. *Science* 284(5419):1514–1516
- Thomas D (1988) Geochemical precursors to seismic activity. *Pure Appl Geophys* 126:241–266
- Thomas J, Masci F, Love JJ (2015) On a report that the 2012 M6: 0 earthquake in Italy was predicted after seeing an unusual cloud formation. *Nat Hazards Earth Syst Sci* 15:1061–1068
- Tomás R, Romero R, Mulas J et al (2014) Radar interferometry techniques for the study of ground subsidence phenomena: a review of practical issues through cases in Spain. *Environ Earth Sci* 71:163–181
- Tronin A (2000) Thermal IR satellite sensor data application for earthquake research in China. *Int J Remote Sens* 21(16):3169–3177
- Tsai C-H, Chen C-W (2010) An earthquake disaster management mechanism based on risk assessment information for the tourism industry—a case study from the island of Taiwan. *Tour Manag* 31:470–481
- Tu PF, Xia L, Li HT, Wu SC (2012) PSInSAR technology research and its application on landslide monitoring. In: Paper presented at the applied mechanics and materials
- Wang T, Zhuang J, Kato T, Bebbington M (2013) Assessing the potential improvement in short-term earthquake forecasts from incorporation of GPS data. *Geophys Res Lett* 40:2631–2635
- Yang C-S, Kao S-P, Lee F-B, Hung P-S (2004) Twelve different interpolation methods: a case study of Surfer 8.0. In: Paper presented at the Proceedings of the XXth ISPRS congress
- Yao Y, Chen P, Wu H, Zhang S, Peng W (2012) Analysis of ionospheric anomalies before the 2011 M w 9.0 Japan earthquake. *Chin Sci Bull* 57:500–510
- Yue H, Lay T (2011) Inversion of high-rate (1 sps) GPS data for rupture process of the 11 March 2011 Tohoku earthquake (Mw 9.1). *Geophys Res Lett* 38:142
- Zhang Y, Jiang Z, Huang H et al (2014) Thermal anomalies detection before 2013 Songyuan earthquake using MODIS LST data. In: Paper presented at the 2014 IEEE international geoscience and remote sensing symposium (IGARSS)



Peak ground acceleration prediction by fuzzy logic modeling for Iranian plateau

Babak Karimi Ghalehjough¹ · Reza Mahinroosta²

Received: 26 September 2018 / Accepted: 10 December 2019 / Published online: 16 December 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

In this study, fuzzy logic modeling is applied to a complex and nonlinear set of data to predict both horizontal and vertical peak ground accelerations in Iranian plateau. The data used for the model include an up-to-date seismic catalogue from earthquakes in Iran for prediction of both horizontal and vertical acceleration of a probable earthquake. Fuzzy logic toolbox on MATLAB program was used for modeling. Earthquake magnitude ranging from 4 to 7.4, source-to-site distance from 7 to 80 km and three different site conditions were considered: rock, stiff soil and soft soil. Results are compared with those from worldwide and regional attenuation relationships, which show the higher capability of the model in comparison with the other models. After training the model, testing of the fuzzy model with the remaining data set was performed to confirm the accuracy of the model. Changes in the peak ground accelerations in connection with changes in input parameters are studied which are in agreement with basic characteristics of earthquake input motions.

Keywords Attenuation relationships · Fuzzy logic modeling · Earthquake input motion · Peak ground acceleration

Introduction

Earthquake input motions are of paramount importance in engineering seismology. Attenuation relationships developed from records of ground motions usually used to estimate these input motions. In these relations, some independent factors such as magnitude, distance from source of the earthquake to site, faulting mechanism and site conditions are related to other dependent parameters such as peak ground acceleration through regression analysis. There are many factors that have been neglected on suggested equation for estimating ground motions such as subsurface topography variations, dynamic and static stress drop in the crust, radiation pattern and different decay rate for different types of waves (Anderson 1991; Boore 1983; Douglas 2001; Joyner

1987; Joyner and Boore 1988). Lack of information on these factors in the equations causes a large deviation in the developed relationships, which makes engineers use a large safety factors at their designs (Douglas 2002) or in some cases, use high input motion.

Ground motion prediction equations (GMPE) are improving every day. New earthquake data following with modern data analysis help researchers to review GMPEs and improve their reliability (Abrahamson et al. 2014; Boore et al. 2014; Campbell and Bozorgnia 2014; Chiou and Youngs 2014; Idriss 2014). Different attenuation relationships have been suggested for prediction of ground motion characteristics for worldwide, a country or a special zone (Ozmen and Babsbug Erkan 2014; Shoushtari et al. 2016). Attenuation relationships developed by Ambraseys and Douglas (2000), Bozorgnia et al. (2000), Campbell and Bozorgnia (2000) and Campbell and Bozorgnia (2003) are some examples for worldwide use, and those developed by Ambraseys et al. (1996), Ambraseys and Simpson (1996) and Bommer et al. (1998) are examples for Europe and Middle east. In connection with this study, attenuation relationships developed by Khademi (2002), Zare and Sabzali (2006) and Amiri et al. (2007) are those developed to predict ground motions in Iranian plateau.

✉ Babak Karimi Ghalehjough
karimi.babak@gmail.com
Reza Mahinroosta
rmahinroosta@csu.edu.au

¹ Civil Engineering Department, Erzurum Technical University, Room: K1-08, 25050 Erzurum, Turkey

² CSU Engineering, Faculty of Business, Justice and Behavioral Sciences, Charles Sturt University, Panorama Avenue, Bathurst, NSW 2795, Australia

In the common methods of developing attenuation relationships, first recorded acceleration histories are analyzed and short- and long-period noises are removed from data. At the next step, useful dependent variable such as spectral acceleration or peak ground acceleration is derived from filtered acceleration histories, while independent variables such as source-to-site distance or magnitude are collected. At last complex analyses with the aim of best regression are performed to get an acceptable result for dependent variables (Douglas 2002). Some researchers have used magnitude, source-to-site distance and soil conditions in their relations (Ambraseys and Douglas 2000), and some others applied additional terms like source mechanism (Campbell and Bozorgnia 2000; Atkinson and Boore 2003; Bozorgnia and Campbell 2004). A comprehensive worldwide summary of strong-motion attenuation relationships was prepared by Douglas (2004). The next generation of attenuation relationships for shallow crustal earthquakes was developed in Pacific Earthquake Engineering Research Center (PEER) considering different site conditions (Abrahamson and Silva 2008; Boore and Atkinson 2008; Campbell and Bozorgnia 2008; Chiou and Youngs 2008; Idriss 2008). At recent years NGA WEST program (Next Generation of Ground-Motion Attenuation Models for the Western United States) and its enhancement NGA WEST 2 were developed at PEER in partnership with U.S. Geological Survey and southern California Earthquake Center to develop new ground motion relations (Boore et al. 2014; Campbell and Bozorgnia 2014; Idriss 2014).

Over the last 20 years, new variables have been added to ground motion prediction equation and made them more complex (Laurie et al. 2011). The majority of these models are empirical and with a large number of parameters and dependencies. Another drawback of these models is that they are developed based on a predicted linear or nonlinear equation, with the hypothesis of normality of residuals for testing the developed model (Thomas et al. 2016). The attenuation relationships suggested to be used world widely were developed based on data from different earthquake events on Earth, especially from those with high earthquake magnitude in different local areas. So, specific characteristics of data in local areas were not considered. On the other hand, the numbers of data in local areas are very low and their deviations and discrepancies are very high, which results in high standard deviations in the local attenuation relationships. Therefore, attenuation relationships are highly uncertain due to computational uncertainties and uncertainties of the input parameters. In this case, fuzzy logic method has several benefits as it uses the natural trends of the inspection data and applies flexible interference rules (Sun et al. 2002). Another advantage of fuzzy model is that it can be easily improved and updated by new data of new earthquakes, while this flexibility on both modeling process and updating data does

not exist in the regression methods. In fuzzy logic modeling, nonlinear functions of arbitrary complexity can also be modeled and there is no need to insist on a predefined equation.

For the application of fuzzy logic theory in dealing with the earthquake input motions, some research studies were conducted on the signal processing and classification of acceleration time histories; classification of earthquake strong ground motion records was performed by Alimoradi et al. (2005) using fuzzy pattern recognition. They showed that Fuzzy logic approach was a promising analytical tool in the classification of design ground motion records. An algorithm based on fuzzy logic techniques was introduced by Tsiftzis et al. (2006) for the classification of signals of acceleration time histories according to the damage that they cause in buildings. Mierlus-Mazilu and Majercsik (2010) used fuzzy classification method for classification of signals of ground acceleration time histories to define seismic damage potential of ground motions. Jorjashvili et al. (2012) discovered the role of the uncertainty of the data in the seismic hazard analysis through Fuzzy set theory; they found that when there are insufficient data for hazard assessment, site classification based on fuzzy set theory shows values of standard deviations less than those obtained using the classical way.

The only study on attenuation relationship via Fuzzy logic approach was conducted by Ahumada et al. (2015) for peak ground acceleration on different site conditions, which was based on selected data from PEER database. They showed that the epicentral distance and soil type have major effects on the attenuation characteristics; in comparison with other selected empirical attenuation relationships, their model had higher efficiency; however, they predicted only peak ground horizontal acceleration (PGHA) from data of 15 different earthquake events on Earth. A neuro-fuzzy approach was also proposed by Thomas et al. (2016) for predicting the peak ground acceleration (PGA), peak ground velocity (PGV) and peak ground displacement (PGD) using the database released by PEER; they compared their results with two other hybrid models of artificial neural network and genetic programming and showed that their method had comparatively higher accuracy and lesser computation time. They compared their results with prediction values from three attenuation relationships, but they did not discuss the trends of the output in connection with the input parameters; therefore, they did not investigate whether their method captured principles of earthquake ground motion or not.

The objective of this study is to predict both peak ground horizontal acceleration (PGHA) and peak ground vertical acceleration (PGVA) using fuzzy logic approach to deal with the uncertainties. The data from 28 different earthquakes in Iranian plateau are used with the independent input parameters including records of earthquakes with moment magnitude between 4.0 and 7.4, different site conditions of soft

soil, soil and rock, and source-to-site distance from 7 to 80 km. These are three input parameters of the fuzzy model, and PGHA and PGVA are its output. After evaluation of the fuzzy logic model based on training and testing data, the results are compared with those from attenuation relationships developed by other researchers for worldwide and local area. The comparison shows improved performance and better efficiency of the developed fuzzy model. In addition, trends of both PGHA and PGVA in connection with input motion parameters are discussed which show that the developed fuzzy model exhibits the important characteristics of the earthquake input motion.

Fuzzy logic theory

In order to handle problems with imprecise data and ambiguity, the fuzzy logic technique was introduced and pioneered by (Zadeh 1965). The method is about the relative importance of precision (MATLAB 2000). It comes from fuzzy set theory dealing with reasoning that is approximate rather than just conclude from classical predicate logic. Fuzzy logic is a mathematical and an organized method of handling inherent inexact concepts and can be used as a new method or way for expressing probability. It can apply to different kinds of uncertainty. Probability works with chances of that happening but fuzzy logic deals with imprecision. This technique can be applied to a nonlinear and complex set of data. There are five different steps that process fuzzy interface:

- Fuzzification of input data; the first step is to take the inputs and determine the degree to which they belong to each of the appropriate fuzzy sets via membership functions. In this study, these membership functions are defined for magnitudes and source-to-site distance and site conditions.
- Application of the fuzzy operator; based on expert judgment, some logical rules are defined to correlate outputs to the inputs.
- Implication from the antecedent to the consequent; in this step proper weighting is assigned to each rule and implication method is applied, resulted in a fuzzy set represented by a membership function.
- Aggregation of the consequents across the rules; the rules are combined in order to make a decision, which is a single fuzzy set for the output.
- Defuzzification; here some methods are used to build a single number from the resultant fuzzy set (MATLAB 2000)

Each input parameter has a degree on membership functions in a fuzzy set. Higher probability of an input has a higher degree of membership. If a degree of an element is zero, it

shows that the element is not a member of fuzzy set. For the case of this study, initial Gaussian membership functions for earthquake magnitude, source-to-site distance and site conditions are shown in Fig. 1. These inputs are ambiguous and imprecise and are the common input parameters of attenuation relationships. Because of the nature of fuzzy logic theory and its intrinsic feature in dealing with uncertainty, fuzzy logic modeling is used in this study as a technique to predict both PGHA and PGVA of a probable earthquake.

Strong-motion database

The ground motion at a site depends on the source-to-site distance, energy released by the earthquake, local geologic conditions and rupture mechanism (Wadia-fascetti and Gunes 2000). For the definition of source-to-site distance, Joyner and Boore (1981) found that the correct distance for estimating ground motion parameters is the distance between the site and exact point of earthquake occurrence (hypocentral distance). Determination of this distance in the past earthquakes and its prediction for future earthquakes is very difficult. Therefore, the distance used at this modeling is epicentral distance that is the distance from the site to epicenter of an earthquake.

For the energy of an earthquake, moment magnitude (M_w) is used as another input parameter, which was used in many equations (Boore et al. 1993; Kobayashi et al. 2000; Lawson and Krawinkler 1994; Sadigh et al. 1997). For the local geologic or site conditions, shear-wave velocity in the top 30 m (V_{s30}) of the site was considered as the input for the site condition, which was used in the recent developed attenuation relationships (Abrahamson and Silva 2008; Boore and Atkinson 2008; Campbell and Bozorgnia 2008; Chiou and Youngs 2008). Depending on the information of the site condition, categories such as “very soft soil,” “soft soil,” “stiff soil” and “rock” sites can be considered in the attenuation relationships. For instance, researchers such as Khademi (2002) and Amiri et al. (2007) used only general soil and rock conditions. However, in this study to be consistent with other attenuations relationships such as Boore et al. (1997), Ambraseys et al. (2005a, b) and Zare and Sabzali (2006), three different site conditions of rock, stiff soil and soft soil were considered with the definition described in Table 1.

Due to the lack of information in the recorded earthquake data, parameters such as rupture mechanism, hanging wall effect and alluvium depth were not considered in the developed fuzzy model. This keeps the model simple and consistent with other comparable attenuation relationships such as those developed by Boore et al. (1997), Khademi (2002), Zare and Sabzali (2006) and Amiri et al. (2007).

At this study, 316 acceleration records from 28 earthquake events in Iranian plateau were accessed from Road,

Fig. 1 Initial membership functions for **a** magnitude of earthquake, **b** source-to-site distance and **c** site condition

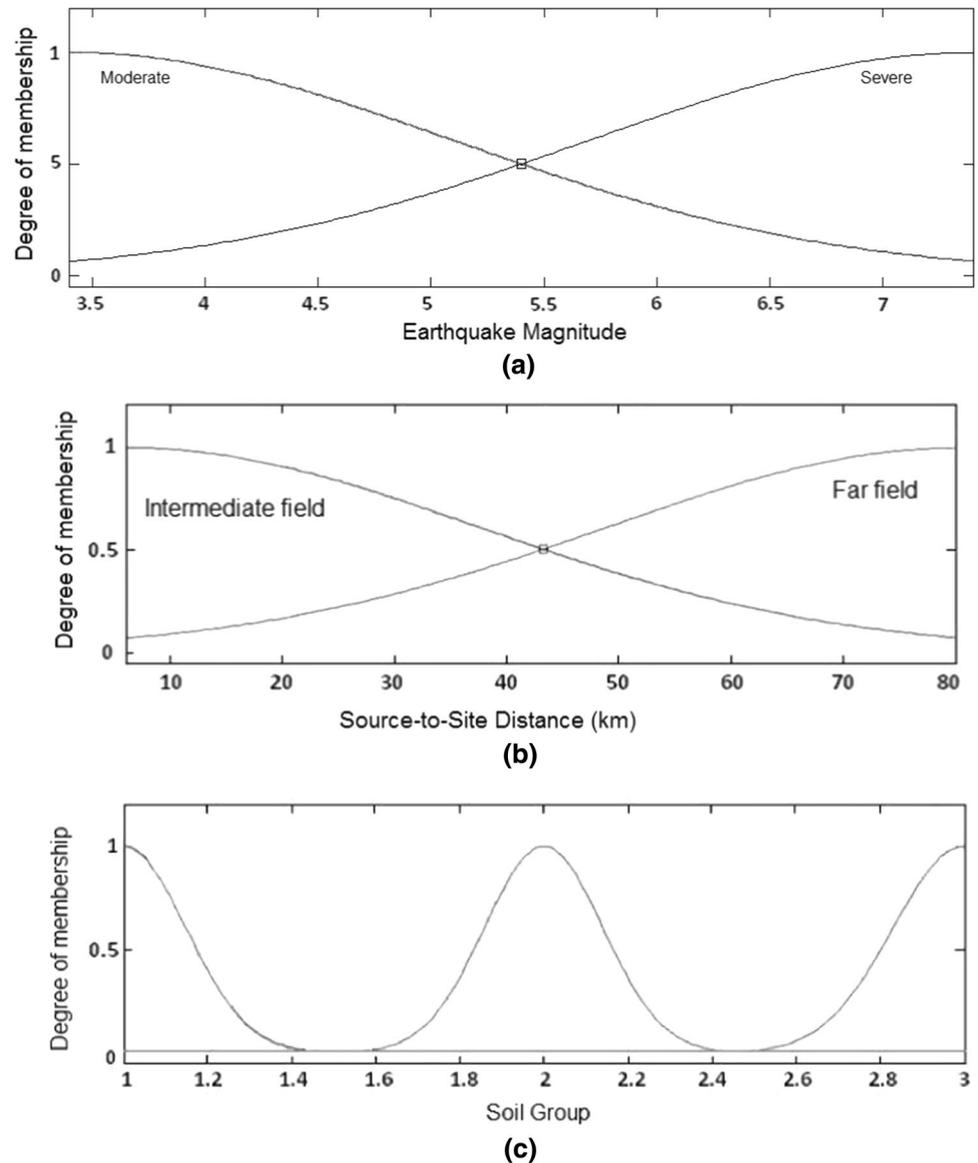


Table 1 Definition of the site condition in this study

Soil group	Site condition	$V_{s(30)}$ (m/s)
1	Rock	> 750
2	Stiff soil	350–750
3	Soft soil	< 350

Housing and Urban Development Center (2012). These data were collected from seismographs stations across the area. Each of the data involves the information of coordination of the station, date and time of earthquake, epicenter and magnitude of earthquake and acceleration record at two horizontal and one vertical direction. For the input of the model, source-to-site distance between 7 and 80 km and earthquake magnitude from 4 to 7.4 Richter were chosen.

For the output of the model, both horizontal and vertical accelerations were used. For the case of horizontal acceleration, average of two components of horizontal accelerations was used, which were in the range of 0.01–0.67 g. Because of the large number of data used for the modeling, some of these data are shown in Table 2 for examples. This includes data in the ranges that has been used in this study.

Modeling and fuzzy logic rules

Fuzzy logic toolbox in MATLAB program was used for fuzzy modeling (MATLAB 2000), in which moment magnitude, source-to-site distance and soil conditions were chosen as input parameters, and PGHA and PGVA were outputs. At first, three overlapped membership functions were used

Table 2 Example of data used for modeling

Nos.	Record	Station	Date	M_w	R (km)	Site	PGHA(1) m/s^2	PGHA(2) m/s^2	PGVA m/s^2
1	1012	Kiasar	11/05/1974	4.5	64	2	0.390	0.300	0.157
2	1006-1	Bandarabbas	03/07/1975	6.1	48	2	0.863	1.287	0.417
3	1006-2	Bandarabbas	03/07/1975	5.2	40	2	0.167	0.260	0.125
4	1007	Minab	03/07/1975	6.1	80	1	0.217	0.168	0.094
5	1008	Gheshm Island	03/07/1975	6.1	56	1	0.148	0.121	0.137
6	1013	Tonkabon	03/13/1975	4.4	34	3	0.428	0.192	0.102
7	1009	Minab	04/12/1975	4.8	48	1	0.357	0.154	0.077
8	1026	Shiraz	06/02/1975	4.1	60	3	0.505	0.251	0.286
9	1014-4	Hajiabad	10/08/1975	5.4	32	3	0.669	0.796	0.471
10	1024	Maraveh-Tappeh	12/27/1975	4.6	6	3	3.270	3.595	1.533
11	1034-1	Maku	02/27/1976	3.9	26	2	0.432	0.537	0.316
12	1034-2	Maku	04/02/1976	4.6	12	2	1.997	1.022	0.845
13	1040-3	Naghan-1	09/05/1976	4.7	88	1	0.437	0.495	0.452
14	1043	Ghaen	11/07/1976	6.4	10	1	1.150	1.570	1.700
15	1047-8	Vendik	11/07/1976	6.4	11	2	5.117	4.982	1.167
16	1047-9	Vendik	11/07/1976	4.8	9	2	1.750	1.559	0.674

for magnitude and source-to-site distance, but none of the simulation was good enough, so with the process of trial and error on the number of membership functions, better simulation was conducted with two-overlapped membership function for both magnitude and source-to-site distance. Membership functions for magnitude of earthquake (M) are called moderate and severe and for epicentral distance (D) are intermediate field and far field.

Soil condition (S) was classified into three different groups without any effect on each other, and because of this reason, there is no overlap of the membership functions. Figure 1 shows initial Gaussian membership functions for the magnitude of earthquake, source-to-site distance and site condition, respectively. It should be mentioned that due to the lack of data with epicentral distance less than 7 km, no membership function was considered for the near field earthquakes. In fact, the fuzzy logic model is proposed for the estimation of earthquake acceleration with the source-to-site distance between 7 and 80 km. This distance contains intermediate and far field distances.

Fuzzy rules for a sample data are defined as follows:

R : IF M is X_1 and D is X_2 and S is X_3
then PGHA is Y_1 and PGVA is Y_2

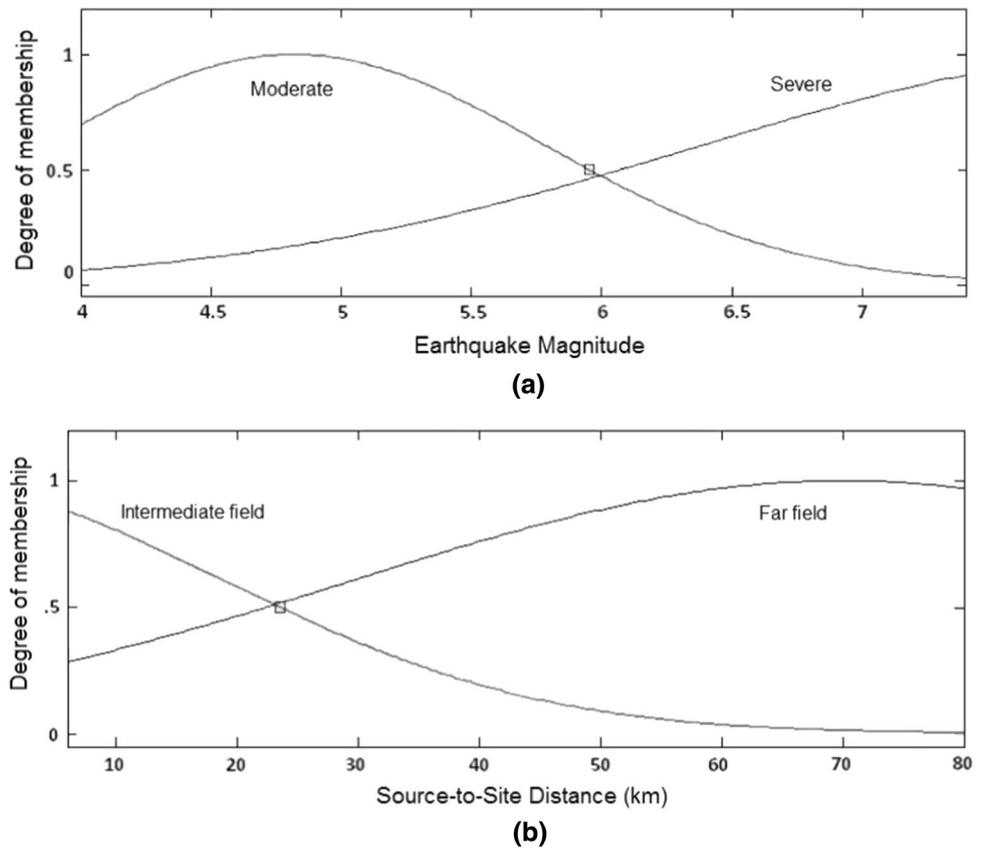
X_1 , X_2 and X_3 represent membership functions for magnitude of earthquake, source-to-site distance and site condition, respectively. Y_1 and Y_2 are resultant peak horizontal and vertical accelerations.

These rules may be triggered in different strengths based on the input data of the earthquakes and some of them may not be triggered. With the combination of the triggered rules,

PGHA and PGVA were achieved through use of two fuzzy interface systems: Mamdani-type interface (Mamdani and Assilian 1975) and Sugeno interface system (Sugeno 1985). After several trial and errors on the output and with the focus on reducing the error in the predicted values, Sugeno-type interface was used for the final layout of the fuzzy logic model. For the defuzzification of the output, the weighted averages of all functions were applied.

From the total data of 316 acceleration records, 240 random data were used for training the model and remaining data were used for testing. To train the model to reach an acceptable acceleration, more than 8000 training were done to reach a constant error reduction of 5%. During training the model, membership functions of the magnitude and source-to-site distance were changed accordingly with the aim of lowering the ultimate error in such a way that the best match occurred in the outputs. Figure 2 shows the optimized membership functions for magnitude and source-to-site distance after training the model. Comparing the membership function in Figs. 1a and 2a, it can be seen that the peak of the membership function for moderate function was changed from $M = 4$ to $M = 5$, which shows that in the selected database, data with magnitude around 5 are more effective than those with magnitude 4 in prediction of the accelerations. For the membership function of source-to-site distance, this comparison shows that the membership function for the far field and intermediate field were increased and decreased, respectively; membership function for far field was strengthened and for the intermediate field was weakened.

Fig. 2 Optimized membership functions for **a** magnitude of earthquake and **b** source-to-site distance



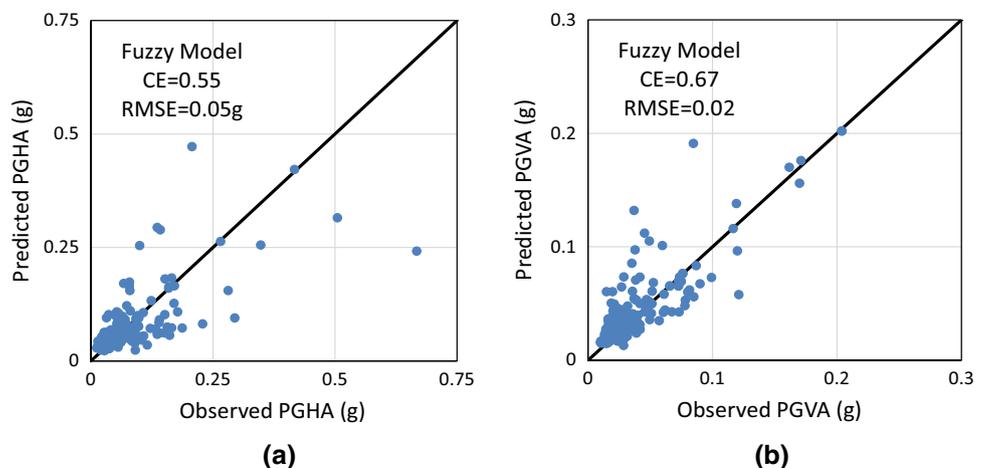
Evaluation of the model

All the input parameters have some uncertainty, but the most uncertainties occur in the site conditions and the source-to-site distance. Similar to the study by Ahumada et al. (2015), these uncertainties were not waived in this study, but a better and accurate approach was developed to address them.

In order to evaluate the performance of the model, predicted and measured values of PGHA and PGVA are shown

in Fig. 3. As can be seen, the data in both peak accelerations are distributed along the line $Y=X$, which shows that predicted accelerations are close to observed accelerations. In order to evaluate the performance of the model quantitatively, coefficient of efficiency (CE) and global error in the model as root mean square error (RMSE) were used based on the following definitions:

Fig. 3 Comparison between observed and predicted earthquake accelerations, **a** PGHA and **b** PGVA



$$CE = 1 - \frac{\sum_{i=1}^N (a_{pi} - a_{oi})^2}{\sum_{i=1}^N (\bar{a}_o - a_{oi})^2} \quad (1)$$

$$RMSE = \left(\frac{1}{N} \cdot \sum_{i=1}^N (a_{pi} - a_{oi})^2 \right)^{0.5} \quad (2)$$

where a_{pi} , a_{oi} and \bar{a}_o are predicted acceleration, observed acceleration and mean observed acceleration for the observation i , respectively. N is the total number of data.

Both CE and RMSE are presented in Fig. 3; due to a high discrepancy in the data, which usually exists in the earthquake input motions, the coefficient of efficiencies in both PGHA and PGVA is a little low, with a better accuracy in the vertical acceleration.

In order to compare the efficiency of the model with other attenuation relationships, two worldwide and three regional relationships were used; for the former, attenuation relationships developed by Boore et al. (1997), Ambraseys et al. (2005a, b) were selected and for the latter, the relationships presented by Amiri et al. (2007); Khademi (2002), Zare and Sabzali (2006) were used. The reason for selection of the worldwide attenuation relationships is first, because data from Middle East and Iranian plateau were used in the development of the relationships and second due to use of similar local site conditions as those in this study.

The results of the acceleration prediction with these attenuation relationships are shown in Fig. 4 in comparison with the fuzzy logic simulation. Figure 4a and b shows that there are high discrepancy in the observed and predicted accelerations in all attenuation relationships similar to fuzzy logic modeling, but the coefficient of efficiency in fuzzy logic model is better than those based on attenuation relationships.

For PGHA greater than 0.4 g almost all methods underestimate the horizontal acceleration, even fuzzy logic modeling; the reason for the former is the insufficient data for high accelerations. Among the attenuation relationship methods, the method presented by Amiri et al. (2007) has better results and its CE and RMSE are the same as those in the fuzzy logic model. RMSE as an overall error in all data is almost in the range of 0.05–0.06 g.

For PGVA, fuzzy logic modeling has a better result than the methods based on attenuation relationships. Methods presented by Zare and Sabzali (2006) and Ambraseys et al. (2005b) overestimate and underestimate the vertical acceleration, respectively, with low coefficient of efficiency. But there is no strong bias in other methods as well as in the fuzzy logic model. The deviation from the 45° diagonal line in the fuzzy logic model is less than methods relying on attenuation relationships, which results in higher CE and lower RMSE in the fuzzy logic model. These confirm that

the efficiency and accuracy of the fuzzy logic model are better than those based on attenuation relationships.

Earthquake input motion characteristics in the model

An important contribution of the fuzzy system is its inherent capacity to capture nonlinear relationships (Ahumada et al. 2015) in the input–output set of data. As can be seen in the developed model, the complex relationship was set in the selected earthquake database. But does this model incorporate the fundamental aspects of attenuation relationships? Does the model capture the important characteristics of the earthquake motions? In this part, these important aspects are discussed for the developed fuzzy logic model to demonstrate how well the model is aligned with principles of seismic wave propagations.

An important aspect of modeling with intelligent systems is that if the selected data are appropriate, these models should show the engineering principles and mechanisms throughout the whole system without enforcing by additional rules. In order to control these specific aspects in this study, the trends of the output in connection with changes in the inputs are investigated. For instance, after training and testing the model and achieving an acceptable fuzzy model in the previous section, different values were entered in the model with different magnitudes and source-to-site distances on three different site conditions to see the changes in both PGHA and PGVA. Changes in the output with changes in the inputs were assessed to find their dependencies, which are shown in Fig. 5 for horizontal acceleration. It can be seen from these figures that with increasing the source-to-site distance of earthquake, the horizontal acceleration decreases, which is expected due to the geometrical spreading of the earthquake motions in the ground. Also in this figure, the increase in PGHA with the increase of earthquake magnitude is shown, which is compatible with the general concept of wave source generation.

The effect of site condition has also been captured by the fuzzy logic model. The results of the fuzzy logic model show that the peak accelerations in the stiff soil and soft soil conditions in all distances and earthquake magnitudes are higher than those in rock condition. Based on data from Mexico City and the San Francisco Bay area and several response analyses, Idriss (1990) showed that at low to moderate acceleration levels (less than about 0.4 g), peak acceleration at the surfaces of soil deposits are likely to be greater than on rock sites. This is in agreement with the results of the modeling in this paper, depicted in Fig. 5. In higher accelerations, opposite trends may happen in soft soils because of the dissipation of earthquake energy in soil layers due to their nonlinear behavior. However, the maximum predicted earthquake on

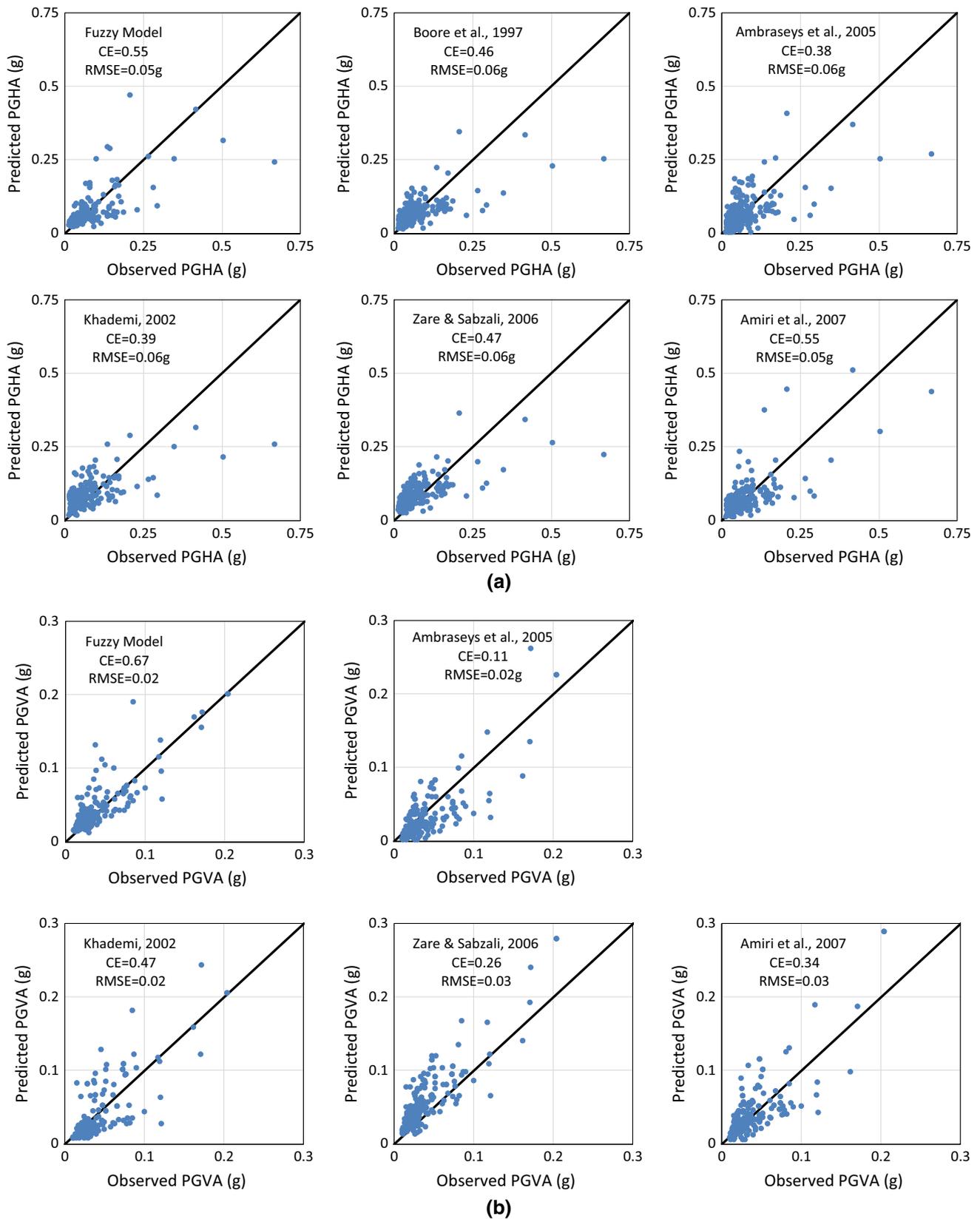


Fig. 4 Comparison between observed and predicted ground motion accelerations using fuzzy logic modeling and attenuation relationships **a** PGHA and **b** PGVA

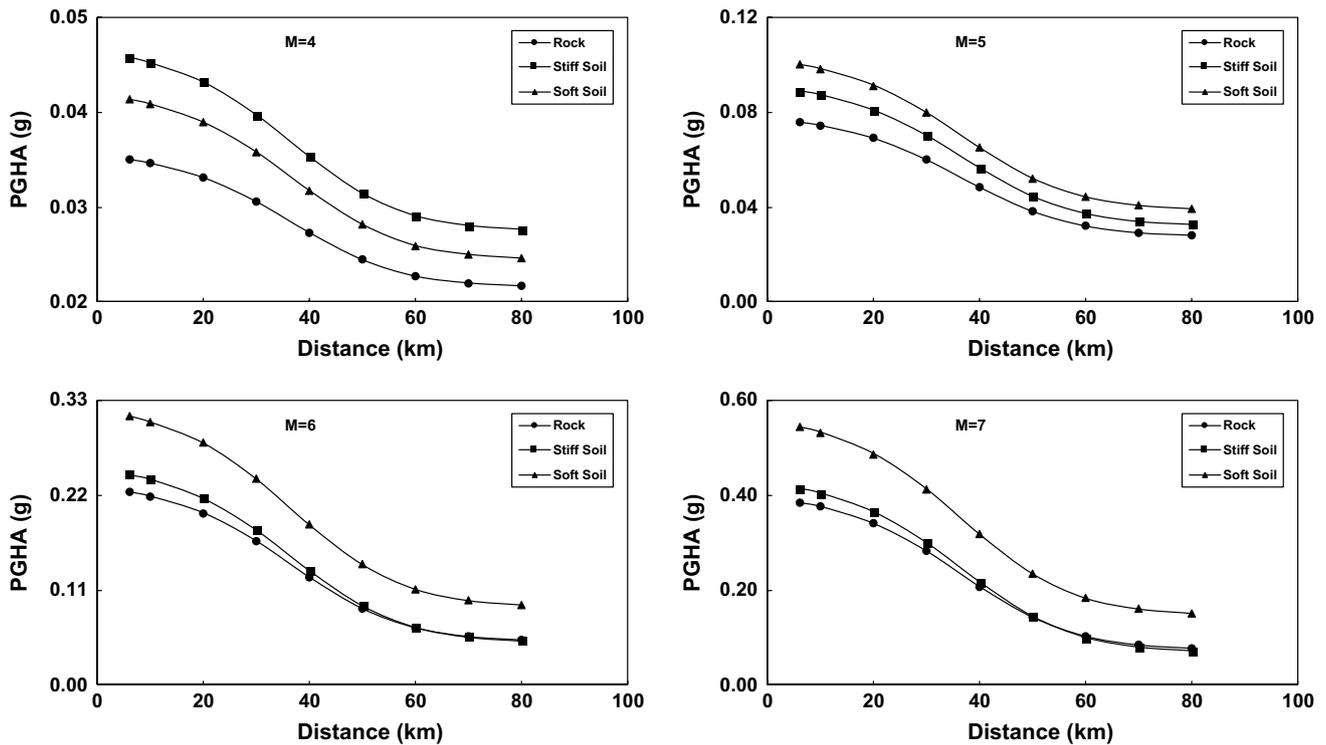


Fig. 5 Variation of peak ground horizontal acceleration with earthquake magnitude and source-to-site distance in different site conditions

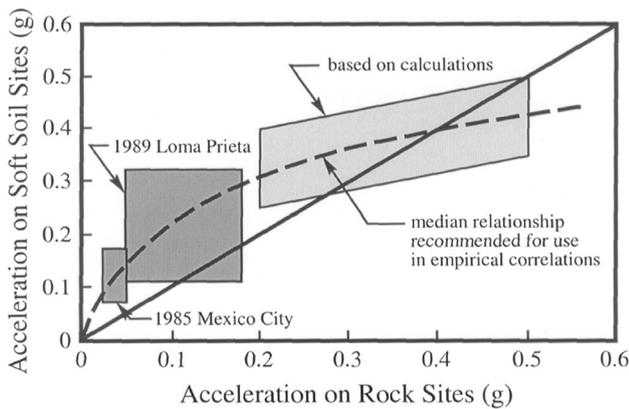


Fig. 6 Approximate relationship between peak accelerations on rock and soil sites (Idriss 1990)

rock sites in the fuzzy logic model even in the magnitude of 7 is around 0.4 g. Comparing with the results of the ground response analysis by Idriss, in these range of accelerations, both increase and decrease in peak ground acceleration on soils are possible (Fig. 6).

Changes in peak vertical acceleration with source-to-site distance for four values of earthquake magnitude and different site conditions are shown in Fig. 7. As can be seen on these graphs similar to horizontal accelerations,

vertical accelerations reduce by increasing the distance and decreasing the earthquake magnitude.

Similar to horizontal acceleration, site conditions affect the values of vertical acceleration. It can be seen from Fig. 7 that vertical accelerations in rock sites in all distances and in all earthquake magnitude are less than those in soil sites, which is in agreement with Idriss’s study mentioned before. This is the result of local site condition, which can influence all important characteristics of strong ground motion such as duration, frequency content and amplitude. The effect of local site condition can be illustrated in several ways: by theoretical ground response analyses, by measurement of ground surface motions from different sites with different subsurface conditions and by measurement of actual surface and subsurface motions at the same site, but in general softer soil will amplify low-frequency (long-period) bedrock motions more than stiffer soil and the reverse would be observed for high-frequency motions (Kramer 1996).

In order to see the change in the acceleration due to the change in the earthquake magnitude, PGHA and PGVA for the source-to-site distance of 35 km are presented in Fig. 8. The trend shows that with an increase in the earthquake magnitude, both PGHA and PGVA increase; the former in an S-shaped curve and the latter with increasing rate. This picture also shows the increase in the peak accelerations in the soil sites in comparison with rock sites.

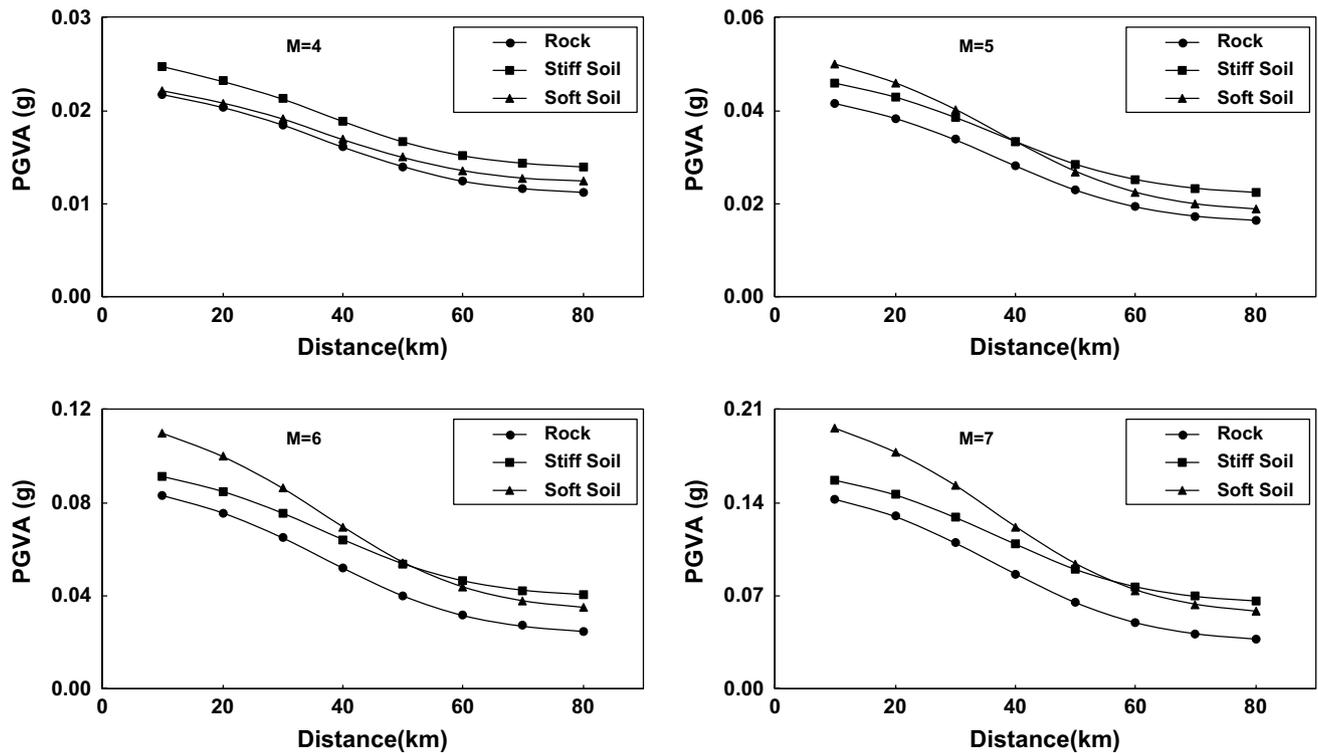


Fig. 7 Variation of peak ground vertical acceleration with earthquake magnitude and source-to-site distance in different site conditions

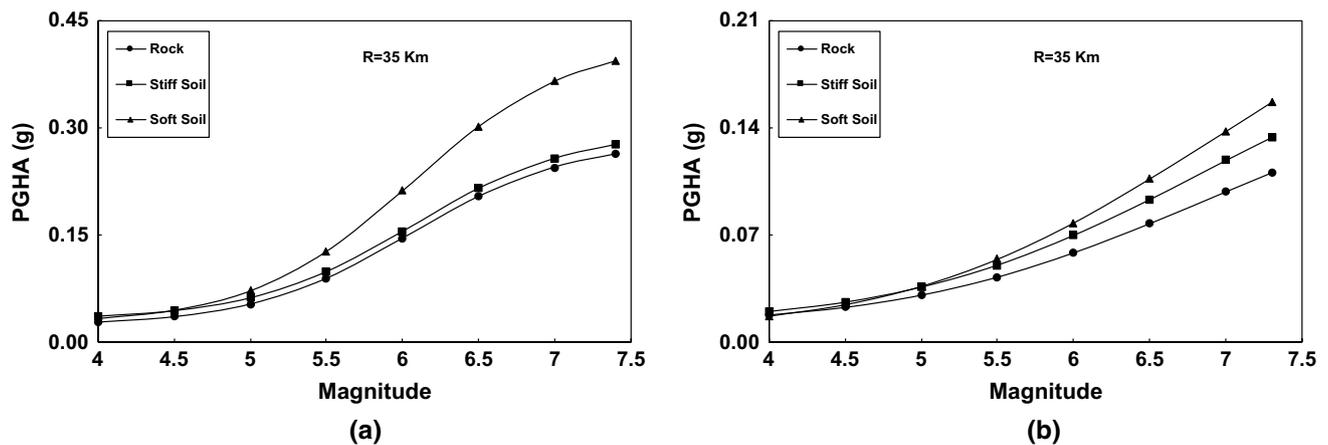


Fig. 8 Changes in horizontal and vertical acceleration at the distance of 35 km

Changes in horizontal and vertical acceleration at two different magnitudes of 4 and 6 for the rock site condition are shown in Fig. 9. Reduction in the acceleration in both PGHA and PGVA is shown in these curves. Results show that differences between horizontal and vertical acceleration in higher magnitudes are more considerable than lower magnitudes, especially when distance is lower than 40 km. In the lower magnitude ($M=4$), difference between PGHA and PGVA remains constant up to almost 40 km.

The above discussion shows that important characteristics of earthquake motion are modeled via fuzzy logic model; reduction of peak ground acceleration with an increase in the source-to-site distance and decrease in the magnitude of the earthquake and effect of site conditions on the magnification of input motion. Therefore, the developed fuzzy logic model captures these characteristics without enforcing them via additional equations; enforcing is inherited by the input data.

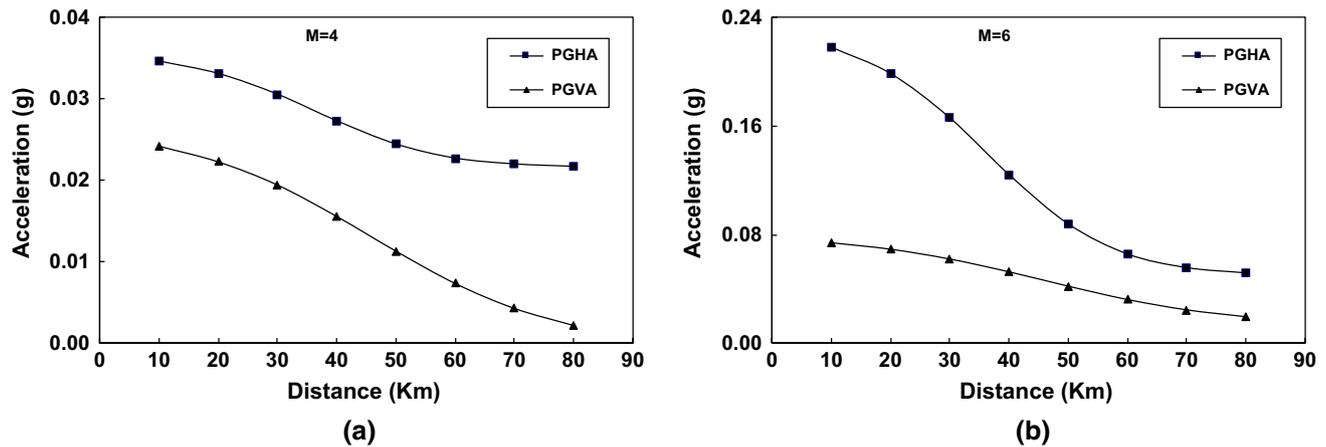


Fig. 9 Changes in the PGHA and PGVA in rock site condition at magnitudes of 4 and 6

Comparison with other attenuation relationships

In Fig. 4, peak ground accelerations from the fuzzy logic model are compared with those from different attenuation relationships. In this section, the results of the fuzzy logic model are compared with these methods considering the concept of attenuation relationships; for instance, the simulation was done with a constant magnitude of 5 and 7 for the local site conditions of soil and rock with various source-to-site distances. In Fig. 10, curves of PGHA and PGVA versus distance from fuzzy logic modeling are compared with those from the study by other researchers. Figure 10a shows that the developed fuzzy logic model has a tendency in presenting higher PGHA at the high earthquake magnitudes in the site-to-source distance less than 40 km, regardless of the site condition; the results of the fuzzy model in this condition in long distances are almost in the range of other relationships. For PGHA in lower magnitudes, the tendency of the model is lowering the acceleration in the shortest distance. These alterations of the fuzzy logic from other methods are most probably due to the lack of regional acceleration data in high magnitudes and low distances.

Figure 10b shows that attenuation relationship for PGVA from fuzzy logic modeling is almost in the range of acceleration from traditional attenuation relationships, except for the data at high magnitudes and low distances, in which fuzzy logic gives an underestimated result. Again this is due to insufficient acceleration data in these ranges of inputs.

The above discussion shows that when there are no sufficient data in some areas of the strong database, the results of modeling will be overestimated or underestimated and care should be taken to apply the method for practical purposes. One way to reduce the lack of data in the input motion could be the use of data from other areas on Earth, which has the

same seismotectonic characteristics and site conditions as the selected regional area.

As can be seen in Fig. 10, different ground motion models give a different value for the peak ground accelerations and the differences between peak accelerations are high in both worldwide and regional attenuation relationships. Because of these uncertainties in these models, logic tree method is recommended as a framework for the explicit treatment of model uncertainty (Kramer 1996). Instead in the fuzzy logic model, the attenuation relationship is developed based on previous data in the regional area and regional ground motion characteristics are inherited in the model.

Discussion

In traditional approaches for development of attenuation relationships, with regression analysis on all data, a formula is obtained based on theoretical and empirical assumptions. But in a fuzzy system each of the input data has its own weight and validity on the output. Because of the dynamic nature, fuzzy logic modeling is more flexible in predicting the acceleration. Furthermore, on the fuzzy logic model for peak ground acceleration, the uncertainty of data is considered automatically, which changes from one data to another data. Therefore, each output has its own special and separate uncertainty, and because of this fact, a complex relationship between accelerations and input parameters can be established.

The results of this study showed that the efficiency of the fuzzy logic model is higher than the regional and worldwide attenuation relations for prediction of acceleration (Fig. 4). The efficiency in PGVA is even higher. This is probably because the effect of individual data is considered naturally in the fuzzy logic model. Along with the fuzzy logic model, the accuracy of the prediction by regional attenuation

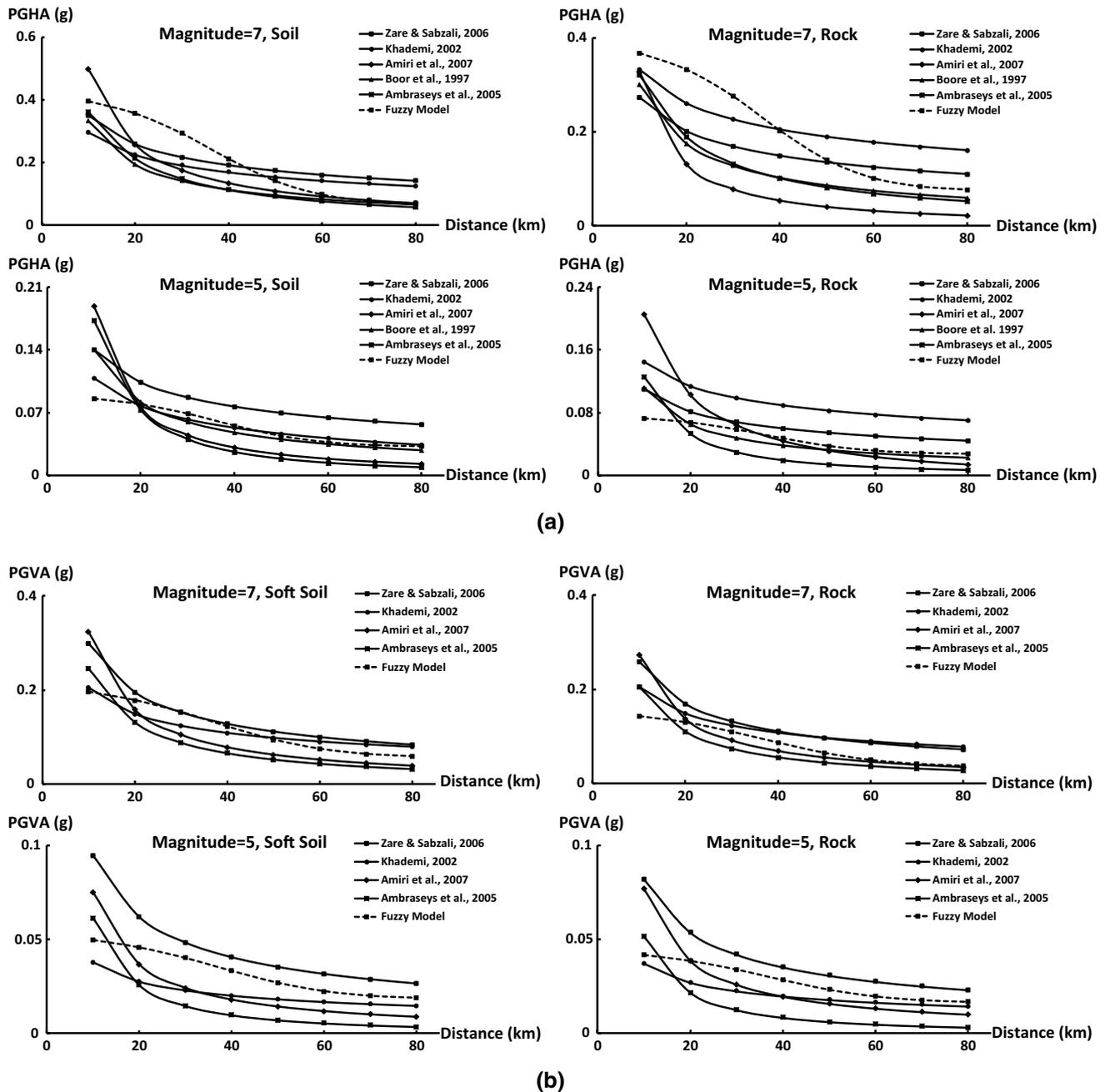


Fig. 10 Curves of peak ground accelerations versus distance for magnitude 5 and 7 in soil and rock sites, **a** PGHA and **b** PGVA

relationships is generally higher than worldwide attenuation. This is because in the formula for the local attenuation relations such as those developed by Khademi (2002), Zare and Sabzali (2006) and Amiri et al. (2007), the data from Iranian Plateau were used with the same site conditions. However, in other attenuation relationships, data of various seismotectonic environments was employed. In the relationship developed by Ambraseys et al. (2005a, b), data from Europe and the Middle East were used; although this covers Iranian plateau, the discrepancy in its results is

higher than regional attenuation relationships and fuzzy model. The accuracy of the results from the attenuation relation developed by Boore et al. (1997) is a little better, while data from western North American earthquakes were used in its development. This means that in addition to the worldwide attenuation relations, regional attenuation relations such as the one developed in this study should be applied.

One issue with the implementation of the fuzzy logic theory is the number of data selected for the prediction. In this study for estimation of ground motion acceleration in

Iranian Plateau, the number of data was limited to the number of records from earthquake events in the region. In fact, the developed model for attenuation relationships works only in this region and cannot be used in other area on Earth. Distribution of the input data is another shortcoming of the fuzzy logic models for earthquake input motions, which is inherited in the earthquake data. More data are available for earthquake input motions with lower magnitudes due to its high frequency of occurrence. In the current paper, 20% of the data have magnitude greater than 5.5. This means that accuracy of the model for earthquake events more than 5.5 magnitude is lower than the rest of them. This is because in the fuzzy logic model more data have been used for training within the range of 4–5.5 magnitude. For the worldwide attenuations, quite a few earthquake acceleration histories are available (e.g., in PEER earthquake database) due to the number of earthquake recorded across the world. However, the number of ground motion records is increasing due to recording new earthquakes. The benefit of the developed model is the fact that new data can be added to update the model. In fact, this is the flexibility of the fuzzy logic model, which benefits users to advance the attenuation model.

Conclusions

In this study, a fuzzy logic model was developed to predict peak ground horizontal and vertical earthquake accelerations inside Iranian plateau. Input parameters were earthquake magnitude between 4 and 7.4, source-to-site distance from 7 to 80 km and local site conditions of rock, stiff soil and soft soil. Comparison between predicted and observed accelerations showed an acceptable level of acceleration prediction for both PGHA and PGVA, especially in the far field. In the developed fuzzy logic model, the uncertainty of earthquake ground motion records was accurately simplified by implementing only a few of motion parameters.

In order to evaluate the accuracy of the fuzzy logic model in presenting basic characteristics of earthquake wave propagation, a parametric study was conducted; it was shown that trends of the PGHA and PGVA with changes in earthquake magnitude, source-to-site distance and local site conditions are meaningful with respect to ground motion characteristics. These were not implemented in the fuzzy logic model as additional rules, but overall rule were achieved from the input–output data. These basic characteristics are:

- With an increase in the distance and a decrease in the magnitude, both horizontal and vertical accelerations decrease, which is in agreement with the concept of geometric spreading and anelastic attenuation.
- In the ranges of data used in this study, the peak ground acceleration in soil sites are more than rock sites.

The comparison between the attention relations showed that the attenuation curves from the fuzzy logic model were in the range of regional attenuation relations for large distances. This is probably because the data from the same seismotectonic environment have been used. The attenuation relations for PGHA developed by the fuzzy logic model were also comparable with those developed by worldwide formula. However, PGVA from worldwide formula (Ambraseys et al. 2005a, b) was lower than the fuzzy logic model and other regional attenuation relations. A possible explanation for this might be that other seismotectonic environments were used in the worldwide formula. Attenuation curves from the fuzzy logic model for the near field acceleration deviate from other relations, which is probably due to the lack of data in the near field.

Overall, the results of the study confirm that new methods based on fuzzy logic theory can be used for improving and developing attenuation relations and predicting both PGHA and PGVA for engineering applications. However, various physics-based GMPEs suffer from a satisfactory number of records of large earthquakes ($M > 6.5$), especially in the near field ($R < 10$ km). The method applied in this study can give more reliable and robust results, provided that the data set used exhibits satisfactory distribution in magnitude, distance and soil conditions.

Acknowledgements We thank the Road, Housing and Urban Development Research Centre, Tehran, Iran, for providing us with the strong-motion database. The authors also gratefully acknowledge the support from Ali Beitollahi, Head of Engineering Seismology Department at this research center, and Mehdi Zare from International Institute of Earthquake Engineering and Seismology (IIEES) for their critical and helpful comments, which have led to significant improvement of the article. The authors also thank Ataturk University for using MATLAB at its computer laboratory.

References

- Abrahamson NA, Silva WJ (2008) Summary of the Abrahamson and Silva NGA ground-motion relations. *Earthq Spectra* 24:67–97. <https://doi.org/10.1193/1.2924360>
- Abrahamson NA, Silva WJ, Kamai R (2014) Summary of the ASK14 ground motion relation for active crustal regions. *Earthq Spectra* 30(3):1025–1055. <https://doi.org/10.1193/070913EQS198M>
- Ahumada A, Altunkaynak A, Ayoub A (2015) Fuzzy logic-based attenuation relationships of strong motion earthquake records. *Expert Syst Appl* 42(3):1287–1297. <https://doi.org/10.1016/j.eswa.2014.09.035>
- Alimoradi A, Pezeshk S, Naeim F, Frigui H (2005) Fuzzy pattern classification of strong ground motion records. *J Earthq Eng* 9(3):307–332. <https://doi.org/10.1080/13632460509350544>
- Ambraseys N, Douglas J (2000) Reappraisal of the effect of vertical ground motions on response. In: ESEE Report 00–4. Department of Civil and Environmental Engineering, Imperial College, London, <http://www.esee.cv.ic.ac.uk/reports.htm>
- Ambraseys NN, Simpson KA (1996) Prediction of vertical response spectra in Europe. *Earthq Eng Struct Dyn* 25(4):401–412. <https://doi.org/10.1002/eqe.143>

- [://doi.org/10.1002/\(SICI\)1096-9845\(199604\)25:4%3c401:AID-EQE551%3e3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-9845(199604)25:4%3c401:AID-EQE551%3e3.0.CO;2-B)
- Ambraseys NN, Simpson KA, Bommer JJ (1996) Prediction of horizontal response spectra in Europe. *Earthq Eng Struct Dyn* 25(4):371–400. [https://doi.org/10.1002/\(SICI\)1096-9845\(199604\)25:4%3c371:AID-EQE550%3e3.0.CO;2-A](https://doi.org/10.1002/(SICI)1096-9845(199604)25:4%3c371:AID-EQE550%3e3.0.CO;2-A)
- Ambraseys NN, Douglas J, Sarma SK, Smit PM (2005a) Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the middle east: horizontal peak ground acceleration and spectral acceleration. *Bull Earthq Eng* 3:1–53. <https://doi.org/10.1007/s10518-005-0183-0>
- Ambraseys NN, Douglas J, Sarma SK, Smit PM (2005b) Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the middle east: vertical peak ground acceleration and spectral acceleration. *Bull Earthq Eng* 3:55–73. <https://doi.org/10.1007/s10518-005-0186-x>
- Amiri GG, Mahdavian A, Manouchehri-Dana F (2007) Attenuation relationships for Iran. *J Earthq Eng* 11(4):469–492. <https://doi.org/10.1080/13632460601034049>
- Anderson JG (1991) Strong motion seismology. *Rev Geophys* 29:700–720
- Atkinson GM, Boore DM (2003) Empirical ground-motion relations for subduction zone earthquakes and their application to Cascadia and other regions. *Bull Seismol Soc Am* 93(4):1703–1729. <https://doi.org/10.1785/0120080108>
- Bommer JJ, Elnashai AS, Chlimentzas GO, Lee D (1998) Review and development of response spectra for displacement based seismic design. In: ESEE Report 98-3. Department of Civil Engineering, Imperial College, London
- Boore DM (1983) Strong-motion seismology. *Rev Geophys Space Phys* 21(6):1308–1318
- Boore DM, Atkinson GM (2008) Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s. *Earthq Spectra* 24:99–138. <https://doi.org/10.1193/1.2830434>
- Boore DM, Joyner WB, Fumal TE (1993) Estimation of response spectra and peak accelerations from western North American earthquakes: an interim report. In: Open-file report. US Geological Survey, pp 93–509
- Boore DM, Joyner WB, Fumal TE (1997) Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: a summary of recent work. *Seismol Res Lett* 68:128–153
- Boore DM, Stewart JP, Seyhan E, Atkinson GM (2014) NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthq Spectra* 30:1057–1085. <https://doi.org/10.1193/070113EQS184M>
- Bozorgnia Y, Campbell KW (2004) The vertical-to-horizontal response spectral ratio and tentative procedures for developing simplified V/H and the vertical design spectra. *J Earthq Eng* 8(2):175–207. <https://doi.org/10.1080/13632460409350486>
- Bozorgnia Y, Campbell KW, Niazi M (2000) Observed spectral characteristics of vertical ground motion recorded during worldwide earthquakes from 1957 to 1995. In: Proceedings of twelfth world conference on earthquake engineering, paper no. 2671
- Campbell KW, Bozorgnia Y (2000) New empirical models for predicting near-source horizontal, vertical, and V/H response spectra: implications for design. In: Proceedings of the sixth international conference on seismic zonation
- Campbell KW, Bozorgnia Y (2003) Updated near-source ground-motion (attenuation) relations for the horizontal and vertical components of peak ground acceleration and acceleration response spectra. *Bull Seismol Soc Am* 93(1):314–331. <https://doi.org/10.1785/0120020029>
- Campbell KW, Bozorgnia Y (2008) NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s. *Earthq Spectra* 24:139–171. <https://doi.org/10.1193/1.2857546>
- Campbell KW, Bozorgnia Y (2014) NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthq Spectra* 30:1087–1115. <https://doi.org/10.1193/062913EQS175M>
- Chiou BJS, Youngs RR (2008) An NGA model for the average horizontal component of peak ground motion and response spectra. *Earthq Spectra* 24:173–215. <https://doi.org/10.1193/1.2894832>
- Chiou BJS, Youngs RR (2014) Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthq Spectra* 30:1117–1153. <https://doi.org/10.1193/072813EQS219M>
- Douglas J (2001) A critical reappraisal of some problems in engineering seismology. Ph.D. Thesis, University of London
- Douglas J (2002) Earthquake ground motion estimation using strong-motion records: a review of equation for the estimation of peak ground acceleration and response spectral ordinate. *Earth Sci Rev* 61:43–104. [https://doi.org/10.1016/S0012-8252\(02\)00112-5](https://doi.org/10.1016/S0012-8252(02)00112-5)
- Douglas J (2004) Ground motion estimation equations 1964–2003: reissue of ESEE report 01-1: ‘A comprehensive worldwide summary of strong-motion attenuation relationships for peak ground acceleration and spectral ordinates (1969 to 2000)’ with corrections and additions. In: Technical report 04-001-SM. Department of Civil and Environmental Engineering; Imperial College of Science, Technology, and Medicine; London
- Idriss IM (1990) Response of soft soil sites during earthquakes. In: Duncan JM (ed) Proceedings, H. Bolton seed memorial symposium, vol 2. BiTech Published, Vancouver, pp 273–289
- Idriss IM (2008) An NGA empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthq Spectra* 24:217–242. <https://doi.org/10.1193/1.2924362>
- Idriss IM (2014) An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthq Spectra* 30:1155–1177. <https://doi.org/10.1193/070613EQS195M>
- Jorjashvili N, Yokoi T, Javakhishvili Z (2012) Assessment of uncertainties related to seismic hazard using fuzzy analysis. *Nat Hazards* 60(2):501–515. <https://doi.org/10.1007/s11069-011-0026-z>
- Joyner WB (1987) Strong-motion seismology. *Rev Geophys* 25(6):1149–1160. <https://doi.org/10.1029/RG025i006p01149>
- Joyner WB, Boore DM (1981) Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake. *Bull Seismol Soc Am* 71(6):2011–2038
- Joyner WB, Boore DM (1988) Measurement, characterization, and prediction of strong ground motion. In: Proceedings of earthquake engineering and soil dynamics II, geotechnical division ASCE, 43–102
- Khademi MH (2002) Attenuation of peak and spectral accelerations in the Persian plateau. In: Proceedings of twelfth European conference on earthquake engineering, Paper reference 330
- Kobayashi S, Takahashi T, Matsuzaki S, Mori M, Fukushima Y, Zhao JX, Somerville PG (2000) A spectral attenuation model for Japan using digital strong motion records of JMA87 type. In: Proceedings of twelfth world conference on earthquake engineering, paper no. 2786
- Kramer SL (1996) Geotechnical earthquake engineering. Prentice-Hall, Inc, Upper Saddle River
- Laurie G, Baise, Eric M (2011) Thompson, incorporating site effects in ground motion prediction equations. In: Final technical report. USGS Award Number G11AP20033

- Lawson RS, Krawinkler H (1994) Cumulative damage potential of seismic ground motion. In: Proceedings of tenth European conference on earthquake engineering, vol 2. pp: 1079–1086
- Mamdani EH, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Man-Mach Stud* 7(1):1–13
- MATLAB (2000) Fuzzy logic toolbox for use with MATLAB: user's guide. In: Fourth printing, revised for MATLAB 6.0. The Math Works Inc
- Mierlus-Mazilu I, Majercsik L (2010) Efficient methods for the classification of seismic damage potential of ground motions. *Sci J Ser Math Modell Civ Eng* 1:51–60
- Ozmen B, Babsbug Erkan BB (2014) Probabilistic earthquake hazard assessment for Ankara and its environs. *Turk J Earth Sci* 23:462–474. <https://doi.org/10.3906/yer-1302-6>
- Road, Housing and Urban Development Research Center (2012) Iran strong motion network. <http://www.bhrc.ac.ir/>
- Sadigh K, Chang CY, Egan JA, Makdisi F, Youngs RR (1997) Attenuation relationships for shallow crustal earthquakes based on California strong motion data. *Seismol Res Lett* 68(1):180–189
- Shoushtari AV, Adnan AB, Zare M (2016) On the selection of ground-motion attenuation relations for seismic hazard assessment of the Peninsular Malaysia region due to distant Sumatran subduction intraslab earthquakes. *Soil Dyn Earthq Eng* 82:123–137. <https://doi.org/10.1016/j.soildyn.2015.11.012>
- Sugeno M (1985) Industrial applications of fuzzy control. Elsevier Science Pub. Co, New York
- Sun SS, Sung DC, Yong RK (2002) Empirical evaluation of a fuzzy logic-based software quality prediction model. <http://dl.acm.org/citation.cfm?id=765833>. Retrieved on 15/05/2016
- Thomas S, Pillai GN, Pal K, Jagtap P (2016) Prediction of ground motion parameters using randomized ANFIS (RANFIS). *Appl Soft Comput* 40:624–634. <https://doi.org/10.1016/j.asoc.2015.12.013>
- Tsiftzis I, Andreadis I, Elenas A (2006) Fuzzy system for seismic signal classification. *IEE Proc Vis Image Signal Process* 153(2):109–114. <https://doi.org/10.1049/ip-vis:20050068>
- Wadia-fascetti S, Gunes B (2000) Earthquake response spectra models incorporating fuzzy logic with statistics. *Comput Aided Civ Infrastruct Eng* 15(2):134–146. <https://doi.org/10.1111/0885-9507.00178>
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–353
- Zare M, Sabzali S (2006) Spectral attenuation of strong motions in Iran. In: Third international symposium on the effects of surface geology on seismic motion grenoble, France



Magnetotelluric sounding evidence of development of nappes in the Tuolai Sag, Yin-E Basin

Zhang Penghui^{1,2,3} · Fang Hui^{1,2,3} · Zhang Xiaobo^{1,2,3} · Zhang Yaoyang^{1,2,3} · Xu Haihong⁴ · Peng Yan^{1,2,3} · Yuan Yongzhen^{1,2,3}

Received: 25 June 2018 / Accepted: 16 October 2019 / Published online: 20 January 2020
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2020

Abstract

To explore the tectonic framework and features of stratigraphic distribution in the Tuolai Sag, Yin-E Basin, a 47-km-long magnetotelluric (MT) sounding measurement was performed around Well MAZD1 in the sag. During field data acquisition, a remote reference technique was used to ensure data quality, with apparent resistivity and phase curves of all measuring points obtained using the Fourier transform, power spectrum selection, robust estimation and other methods. After MT data processing, dimensionality analysis and the degree of two-dimensional deviation indicated that the study area had the dimensionality needed for two-dimensional inversion. The major electrical axis in the sag was determined, using a multi-point–multifrequency point statistical imaging technique, to be in the WNW direction. Within the constraints of the resistivity log data for Well MAZD1, inversion results for TE and TM models were compared, after which the TM model, which corresponded well to geological conditions, was selected for conducting the nonlinear conjugate gradient method inversion and a reliable resistivity model was finally obtained. Based on regional petrophysical properties, resistivity logging, and near-well bathymetric data, the electrical characteristics of different formations within the sag were obtained and a set of low-resistance clastic rock identified in the lower Carboniferous strata. Based on an integrated analysis of the regional surface geology, tectonic setting, and depositional environment, and within the constraints of gravity to fit with electrical structure, a tectonic framework of two subsags sandwiched by an uplift is proposed for the Tuolai Sag. The scale of the northern subsag is large, with development of pre-Carboniferous nappe as well as of Carboniferous–Permian strata within the lower part of the nappe. The southern subsag is small and filled mainly with Carboniferous–Permian strata.

Keywords Magnetotellurics · Yin-E Basin · Electrical conductivity · Nappe

Introduction

The Yingen-Ejinaqi Basin (Yin-E Basin) is a medium–large sedimentary basin in China that covers approximately 123,000 km². To date, little onshore oil and gas exploration has been conducted in this region. The basin, which

is situated primarily in the western part of the Inner Mongolia autonomous region, is delineated by the Sino–Mongolian border to the north and by Gansu Province to the south. Mesozoic and Paleozoic Strata of great thickness are found deposited in the basin. Early oil and gas exploration focused mainly on the Mesozoic, but no major discoveries were made. The Paleozoic stratum, seen as the metamorphic sedimentary basement of the Mesozoic basin (Geng et al. 2016), has attracted little attention. As a result, oil and gas exploration here is still in its early phases. In recent decade, with the advancement of the “Oil and Gas Resources Investigation of New Area and New Strata of North China” initiated by the China Geological Survey, a series of geological surveys has been conducted in the Yin-E Basin targeting the Carboniferous–Permian layer. Hydrocarbon flows discovered in the Guazihu Sag and the Hari Sag offer promising

✉ Fang Hui
fanghui@igge.cn

¹ National Research Center of Geo-exploration Technology, Langfang 065000, China

² Key Laboratory of Electromagnetic Detection Technology, Ministry of Land and Resources, Langfang 065000, China

³ Institute of Geophysical and Geochemical Exploration, CAGS, Langfang 065000, China

⁴ Xi’an Geological Survey Center, Xi’an 710000, China

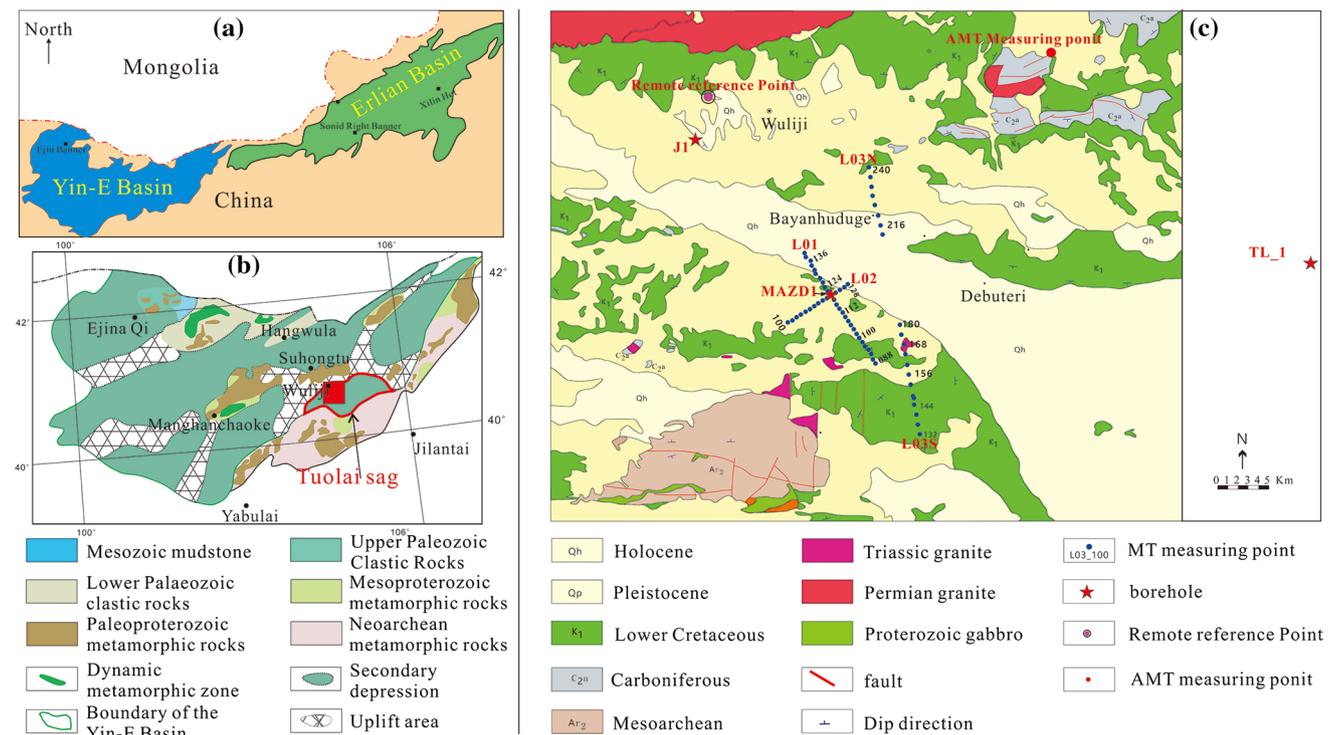
prospects for exploration and suggest the potential presence of resources in the basin (Lu et al. 2011, 2014, 2017).

Drilling in the upper Permian was restricted to a small number of boreholes entirely within the Yin-E Basin, with seismic data on Paleozoic strata limiting the complexities of tectonic evolution and the widespread presence of volcanic rocks, which restricted research into oil and gas exploration and even the resource potential of the basin. To promote oil and gas exploration in the Paleozoic strata in the Yin-E Basin, magnetotelluric (MT) sounding measurement was carried out around Well MAZD1 in the Tuolai Sag. The magnetotelluric method uses electromagnetic sounding of passive sources to infer the distribution of underground resistivity by measuring changes in electric and magnetic fields with time on the surface (Simpson and Bahr 2005). Its efficacy is not reduced by the presence of highly resistive volcanic rocks, making it an advantageous technique for use in geological surveys of oil and gas resources within underlying basins in development areas that are overlaid with volcanic rocks. Analysis of regional geoelectrical features was carried out using the resulting MT data, with stratigraphic structure, fault structure, basin basement, and other characteristics of the sag analyzed along with drilling, surface geology, petrophysical, and other data that provided a reference for the exploration of oil and gas resources in the Carboniferous–Permian stratum in the Yin-E Basin.

Geological background

The Yin-E Basin is located at the intersection of the Kazakhstan Block, the Tarim Block, the North China Block, and the Siberian Block. Before the Devonian, it was similar to the Erlian Basin and the Hailar Basin, and all were part of the Central Asia–Mongolia Ocean. In the Late Devonian, with the convergence of the blocks, the ancient Asian land was formed by the consumption of the ancient ocean, and structural development entered the intraplate tectonic deformation stage. The region passed through evolutionary stages as a Carboniferous–Permian intraplate rift basin and a Mesozoic–Cenozoic inland basin (Lu et al. 2011, 2012, 2014, 2017), forming a complex Carboniferous–Permian and Mesozoic imposed basin (Fig. 1).

During the Carboniferous–Permian rifting stage, the lower Carboniferous–middle Permian was dominated by continuous deposition in the area, and a parallel unconformity was present locally, with a sequence missing (Lu et al. 2017). The unconformity was a formation of clastic and volcanic rocks (intermediate-acidic volcanic rocks; pyroclastic rocks) dominated by littoral neritic facies intercalated locally with carbonate. During the Late Permian, influenced by local uplift, the seawater gradually retreated and continental volcanic rock sedimentation predominated



in the vast areas in the northern and southern portions of the basin. Marine sedimentary formations predominated in the eastern portion of the basin. Affected by uplifting and erosion in the Late Hercynian epoch and the Indo–Chinese epoch, the residual thickness of Carboniferous–Permian varies significantly within the basin, locally exceeding 3000 m. Several sets of dark mudstones of shallow marine shelf facies having good organic geochemical indices were developed in the basin and can be productive source rocks. During the evolutionary stage of the Mesozoic inland basin, the Triassic and Jurassic developed locally. During the Cretaceous, which represented the peak of basin evolution, a stably distributed clastic assemblage of fluvial and lacustrine facies developed with a thickness of up to 2000–3500 m (Wei et al. 2006). Strata outcropping within and around the Tuolai Sag are dominated by Cretaceous and Carboniferous outcrops locally. After drilling of the Cretaceous in existing boreholes in the sag (Well J1), the Permian sandy conglomerate was drilled directly, with Well Tuolai 1 (TL1) in the eastern study area entering the Permian at 940 m. Triassic and Jurassic strata are absent in the depression.

Regional rock geophysical characteristics

Differences in resistivity between different rocks provide the basis for the magnetotelluric sounding method, with analysis of the physical properties of regional rock guiding geological interpretation of the electrical structure profile. Under normal circumstances, the strata of a normal sedimentary sequence in a sag undergo aging; as diagenesis continues, porosity and water content gradually decrease and resistivity gradually increases. Initial magnetotelluric resistivity measurements in the Yin-E Basin display similar features (Liu et al. 2011, 2013). However, when abnormally conductive materials are developed in a formation, trends in normal resistivity's variation with depth change, so that layers having abnormally low or high resistivity can be used as the principal indicator for strata identification and division. For example, high-resistive volcanic rocks, or intrusive rocks, exhibit abnormally high resistivity, which is a universal feature in basins or areas where volcanic structures have developed (Danda et al. 2017). If coal measure mudstone strata having high shale content are interbedded with sandstone and limestone, they will exhibit low-resistivity features, as has been confirmed during oil and gas exploration in small and medium basins found in volcanic rock-covered areas of the Peripheral Songliao Basin (Fang et al. 2013). Accordingly, magnetotelluric sounding can be used to explore the distribution of mudstone strata with low-resistivity features as well as the basin's structural framework.

To examine the features and structure of strata in the Tuolai Sag, data on the first magnetotelluric resistivity values of different outcrop areas in the sag were compiled for analysis, revealing that resistivity increased gradually with age in the stratigraphic chronology (Liu et al. 2011). Research into the electrical characteristics of the entire formation and divisions of marker beds is essential for the geological survey of oil and gas within the basin. Three drilling wells (MAZD1, J1, and TL1) were bored around the study area, and the Cretaceous, Permian, Carboniferous, and pre-Carboniferous strata were drilled, although the Triassic–Jurassic was lost. The Cretaceous strata drilled by the three wells are generally developed in the region and consist primarily of interbedded sandstone and mudstone. Downhole resistivity logs are characterized chiefly by low resistivity, and near-well MT bathymetric results for Well MAZD1 exhibit similar resistivity features in the Cretaceous interval (Fig. 2). The Permian was drilled in wells J1 and TL1, the former of which is composed of coarse sandy conglomerate about 780 m thick and the latter of which consists mainly of metasandstone of 240 m thick. The formation exhibits relatively high resistivity features similar to the upper pre-Carboniferous but lower overall. The Carboniferous was drilled in Well TL1, and its electrical features are similar to those of the Permian, having internal high–low–high-layered features. The outcrop area of the formation is greater on the north side of the sag; it is exposed sporadically within the sag and consists mainly of feldspar sandstone, conglomerate, limestone, and like. Surficial geological surveys have established that thick clastic rocks are developed beneath the volcanic rocks in the lower Permian in the Yin-E Basin (Lu et al. 2017). Normally, clastic rocks rich in carbonaceous components have low resistivity. Electrical structure data obtained from one-dimensional OCCAM inversion of the TM model of one audio magnetotellurics (AMT) measuring point near the exposed Carboniferous limestone suggests the presence of a low-resistivity layer from 200 to 600 m. The upper high-resistivity layer is presumed to be limestone, whereas the lower low-resistivity layer might correspond to the electrical response of the clastic rocks. Overall, the pre-Carboniferous strata exhibit the highest resistivity, and the Carboniferous–Permian strata are characterized by high-resistivity and low-resistivity interbeds, whereas the Cretaceous strata display the lowest resistivity.

Method

MT data acquisition

To control the strata and tectonic units near the boreholes and in the Tuolai Sag, three survey lines were deployed: two lines perpendicular to each other deployed around Well

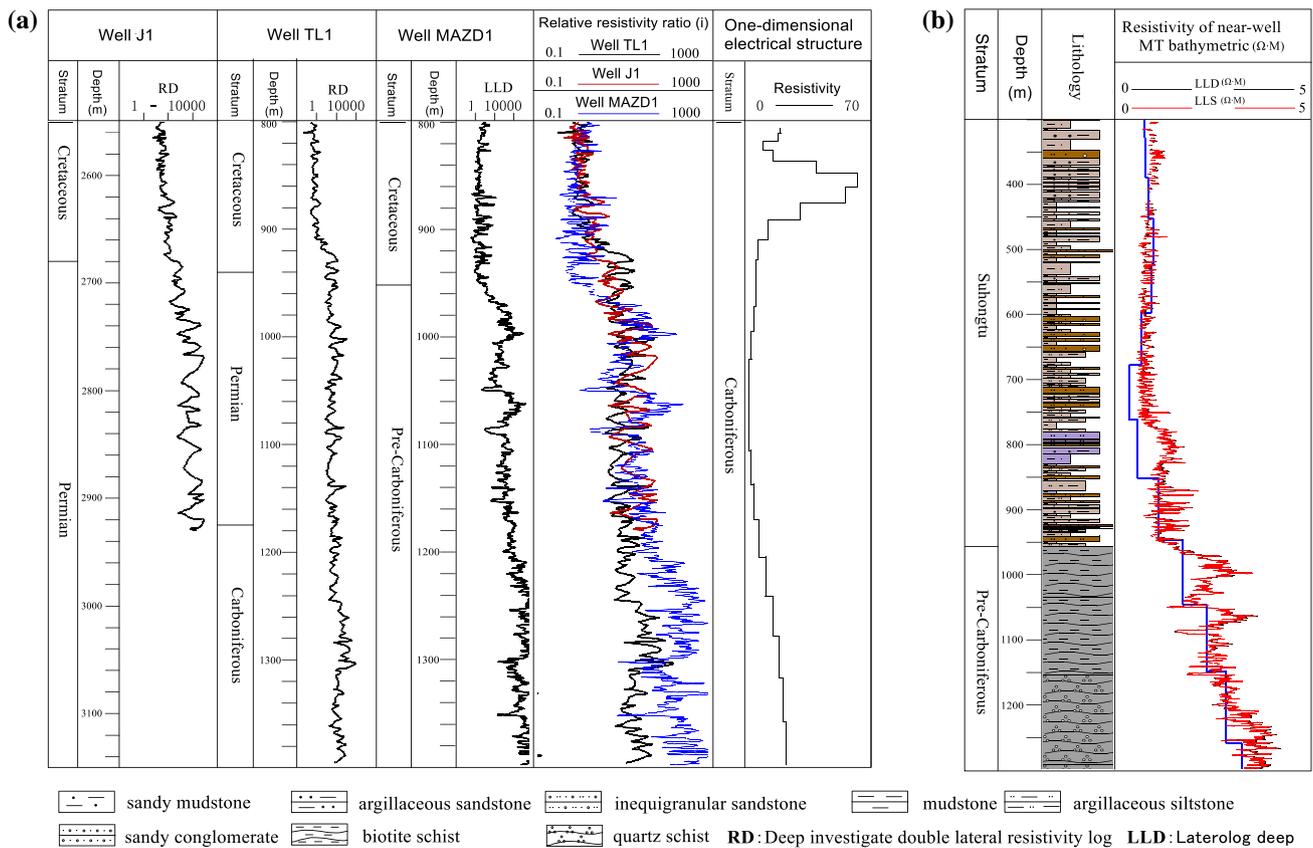


Fig. 2 Comparisons of resistivity logging results between different boreholes and near-well bathymetric data. *Note:* **a** Resistivity logging curve of wells J1, TL1, and MAZD1. The relative resistivity ratio in Column 4, used to analyze the electrical characteristics of regional strata, refers to the ratio of the value of resistivity log to the average resistivity of Cretaceous formation in a well, reflecting variations in

the resistivity log methods and instrument used in three wells. Column 5 (the rightmost side of [a]) shows the one-dimensional electrical structure of the TM model for the AMT point in the Carboniferous outcrop area. **b** Comparisons between the resistivity log curves of Well MAZD1 and one-dimensional OCCAM inversion results of TM model of the MT measuring point

MAZD1 and a third located 3 km east of Well MAZD1. Data acquisition in the field was made using a Phoenix V5 magnetotelluric sounding instrument of Canadian origin. It was laid in a standard plus shape, with the Ex and Hx directions coinciding with the extension direction of the survey line and the Ey and Hy directions perpendicular to the extension direction of the survey line. The designated magnetotelluric point interval was 500 m, and the part point interval was 1 km. The effective bandwidth recording frequency was 0.01 to 320 Hz, the electric dipole distance was 100 m and the acquisition time for each station was guaranteed to exceed 8 h.

Electromagnetic interference was expected because the work area was adjacent to roads and high-voltage lines. To ensure the quality of the raw data, a remote reference technique was using in the acquisition process, based on the minimal little change in magnetic field signal over a certain range at the same time, with the electromagnetic noise between the measuring point and the reference point deemed irrelevant, so that interference could be suppressed

(Shalivahan and Bhattacharya 2002; Kappler 2012). In the northwest, 20 km from MT Line 01, Point 140 (Point L01_140), a long-term observation point, was set up in a plains area free of electromagnetic noise. The apparent resistivity and phase curves of MT Line 01_118, obtained using remote the reference method (B1, B2; Fig. 3), were compared with its original apparent resistivity and phase curves (A1, A2), indicating that the technique is able to suppress electromagnetic interference.

The SSMT2000 software provided developed by the Canadian Phoenix Co. was used for field data preprocessing, and the time domain data were converted to frequency domain data. The power spectrum selection, far reference, and “robust” estimation processing techniques (Egbert 1997) were used to obtain the apparent resistivity and phase curves of the measuring points, which can intuitively reflect the electrical distribution characteristics of the underground medium. The apparent resistivity and impedance phase curves of typical observation measuring points in the study area are shown in Fig. 4: Points L01_132 and L02_114 were

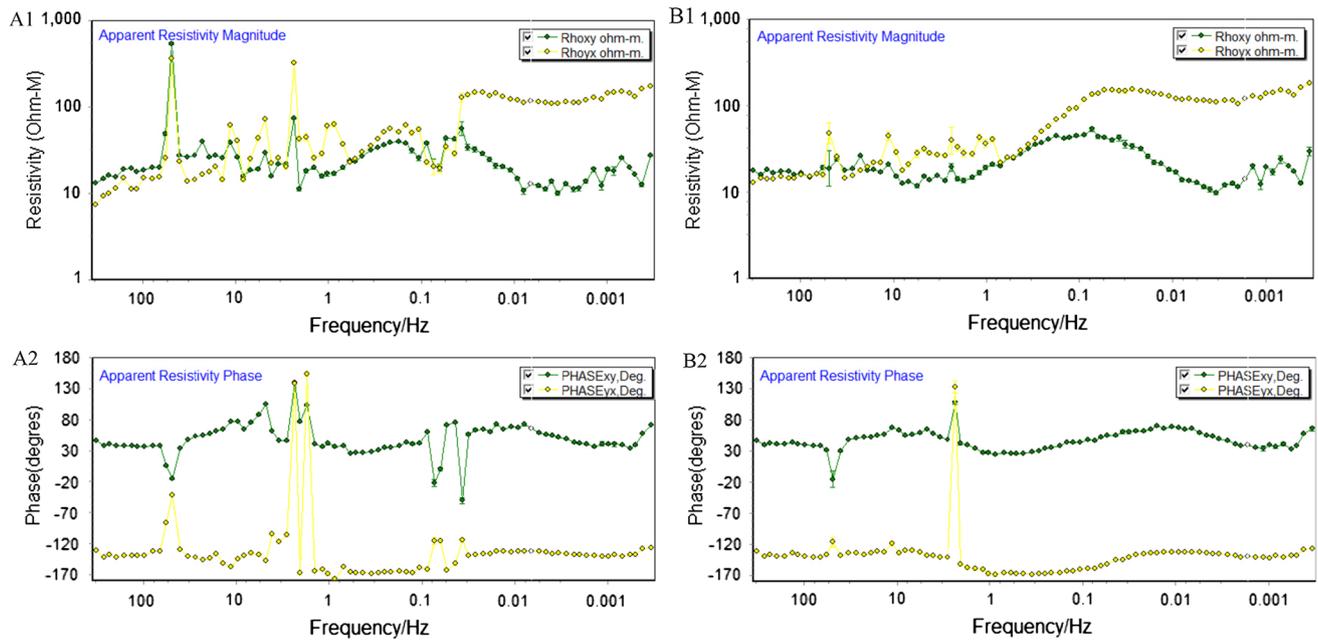


Fig. 3 Comparison diagrams of original apparent resistivity and phase curves (A1, A2), as well as curves obtained using the remote reference method (B1, B2) for MT Point L01_118. *Note:* The point is

450 m away from the 220 V high-voltage line. A1 and B1 are apparent resistivity curves, and A2 and B2 are phase curves

located in the Quaternary covered area and Point L03_156 in the Cretaceous stratum outcrop area. The original apparent resistivity and phase curves indicated that the shallow portion had one-dimensional characteristics. The electrical structures of points L01_132 and L02_114 were similar, with the deep portion having two-dimensional characteristics. However, the electrical structure of Point L03_156 was apparently different from that of the other two points, and its deep structure tended to be one-dimensional, reflecting large differences in structure between the north and south sides of the study area.

Dimension analysis

The two-dimensional deviation skewness (S) of magnetotelluric sounding impedance can be used to distinguish the dimensional characteristics of underground media. Generally, the smaller the S value is, the more two-dimensional the performance of underground media, generally speaking, the underground structure is a two-dimensional feature when the value of S is less than 0.3–0.4, but is a three-dimensional feature at the corresponding depth when the value of S exceeds 0.5.

Based on the Bahr tensor impedance decomposition technique (Bahr 1991), skewness calculations for Line 01 (L01) and Line 02 (L02) were carried out and their dimensional properties then analyzed in the study area. Figure 5 shows variations in Bahr skewness for all measuring points in L01

and L03. The S value was essentially less than 0.3, indicating that the shallow portions (frequency over 1 Hz) of the L01 and L03S sections had very good one-dimensional properties, and that the area possessed the basic geological conditions for two-dimensional electrical inversion.

Goelectric trend analysis

The presence of a specific goelectric trend in the study area is needed to obtain accurate data about the goelectric structure along the survey line (Chen et al. 2014). The structural trend of the underground medium is complex and changeable, showing non-one-dimensional features. During field measurement, the measuring spindle often fails to coincide with the electrical spindle. Accordingly, to accurately identify the stratum's electrical features, electrical information is transferred from the major measurement axis to the major electrical axis. Decomposition of the magnetotelluric impedance tensor is a widely used technique for orienting of electrical axes, as seen in the GB method (Groom and Bailey 1989), Bahr method (Bahr 1988), and CBB decomposition method of the phase tensor (Caldwell et al. 2004; Bibby et al. 2005). A conjugate impedance method (CCZ, based on the first letters of the inventors' surnames) has been proposed by Cai and Chen (2010a) based on the specific relationship between observed impedance and regional impedance, which is unaffected by local distortion. The observed impedance tensor is transformed into conjugate

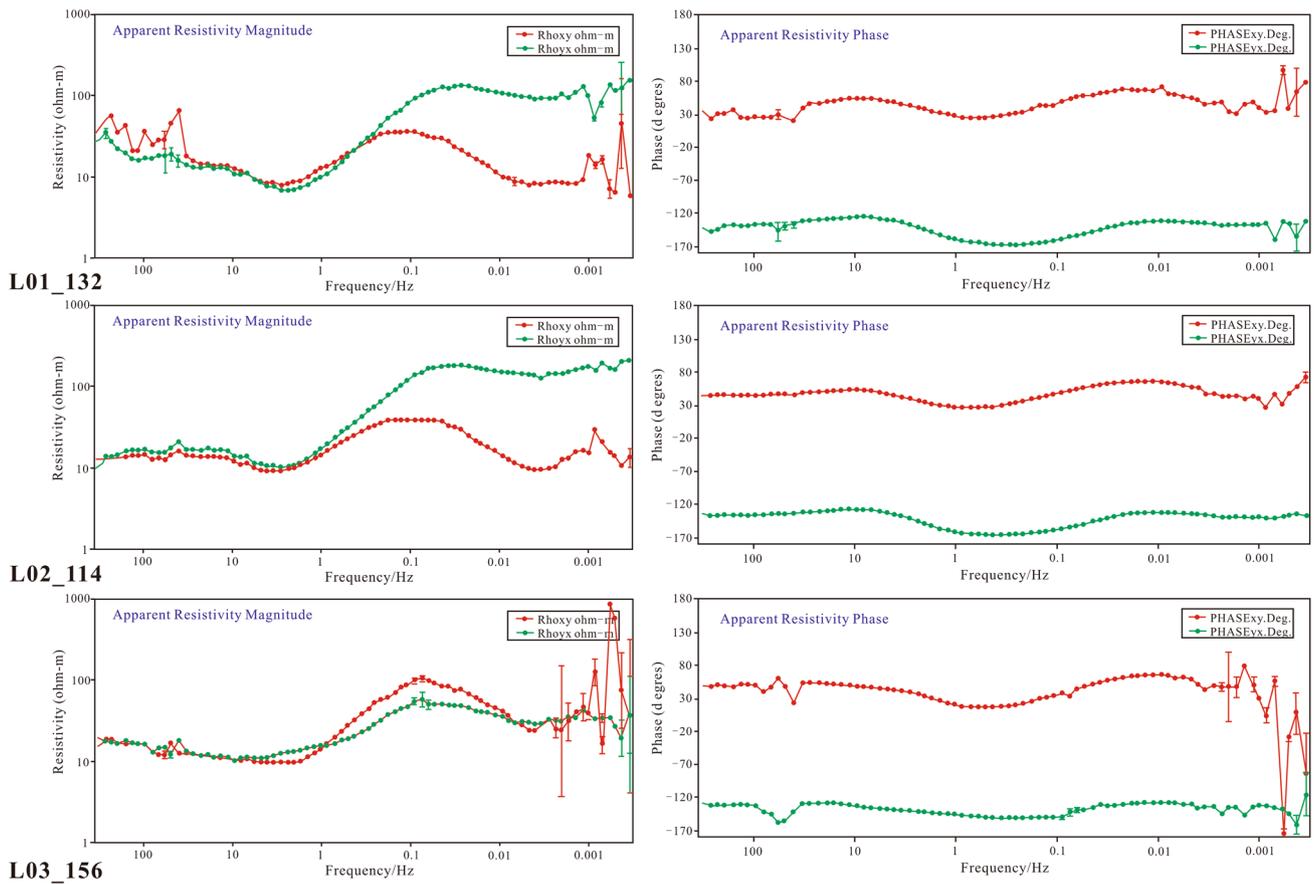


Fig. 4 Apparent resistivity and phase curves of typical measuring points (after remote reference processing)

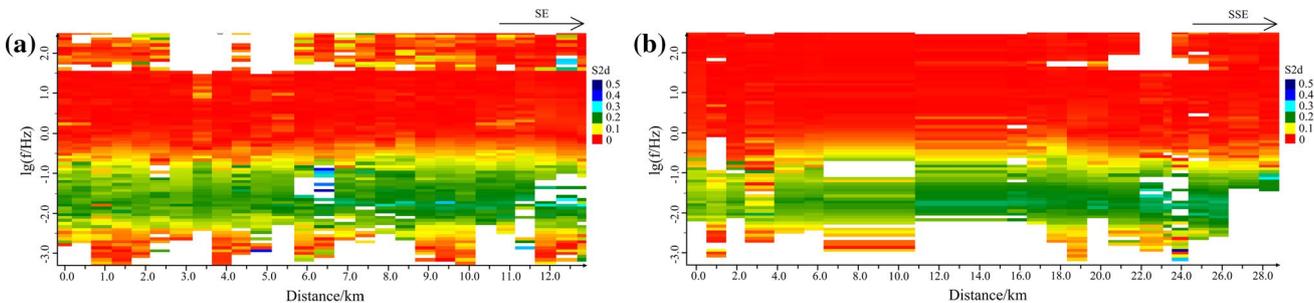


Fig. 5 Profiles of two-dimensional skewness of Line 01 (a) and Line 03 (b). i: The white part of the figure is the frequency point with poor-quality data removed during calculation

impedance according to formula (1), and based on that the series of important parameters of original impedance tensor are obtained:

$$Z^{-1}Z^* = (CZ_r)^{-1}(CZ_r)^* = Z_r^{-1}C^{-1}C^*Z_r^* = Z_r^{-1}C^{-1}CZ_r^* = Z_r^{-1}Z_r^* \tag{1}$$

Note: Z refers to the observational magnetotelluric impedance tensor, Z_r the regional impedance tensor, C the

distortion matrix, Z^{-1} the inverse of Z , Z^* the conjugation of Z , and C^* the conjugation of C .

Single-frequency impedance decomposition is unstable as a result of environmental electromagnetic noise and structural dimensionality, but Chen et al. (2014) proposed a multipoint–multifrequency statistical imaging analysis technique based on the conjugate impedance method (CZZ). Taking the approach of clarifying the phase data

of a single-frequency point, the tensor decomposition of a fixed spindle is carried out with increasing step size at a certain angle (in the range 0–90°), and the phase data are then fitted with the data obtained by the previous free decomposition, with the best-fitting azimuth angle is selected as the decomposition result of multipoint–multifrequency. In this paper, using aforementioned method, the single-frequency conjugate impedance tensor decomposition and multipoint–multifrequency statistical analysis of magnetotelluric sounding data were carried out in the study area.

Figure 6 shows the electrical spindle orientation of a single measuring point and “rose diagram” drawn based on multipoint–multifrequency statistics for three survey lines. The azimuth angles of the electrical spindle for lines L01, L02, and L03 were 34° or 124°, 32° or 122°, and 38° or 128°, respectively. The orientation of the electrical spindle determined by the impedance tensor had 90° of ambiguity, which can be defined by the direction of the regionally dominant tectonic line. Given the regional 1:250,000 Bouguer gravity anomaly, the structural trend of the sag was generally near EW and the main fault strike direction was WNW. Accordingly, the spindle trend of the geoelectric was considered to be WNW.

After orientation of the electrical spindle, the impedance tensor of the measured point was rotated to match the geoelectric direction to determine the corresponding polarization mode. Apparent resistivity (ρ_{xy}) and impedance phase (φ_{xy}) were TE polarization mode data parallel to the tectonic trend. What’s more, apparent resistivity (ρ_{yx}) and impedance phase (φ_{yx}) were TM polarization mode data.

MT inversion

The well-established nonlinear conjugate gradient method (NLCG) developed by Rodi and Mackie (2001) was used. This method has the advantages of rapid inversion calculation, minimal memory capacity requirement, stable inversion results, and the like. (Ye et al. 2007). Furthermore, the method is a local optimization method and is susceptible to the initial resistivity model (Ye et al. 2013). Accordingly, the resistivity model selected for inversion calculation should be as close as possible to the de facto geological model (Zhao et al. 2016). Considering that the sedimentary strata in the sag usually have the characteristics of “horizontal continuity, and vertical stratification,” one-dimensional Occam inversion results of polarization TM were selected as the initial model. In this model, the macro-contour information

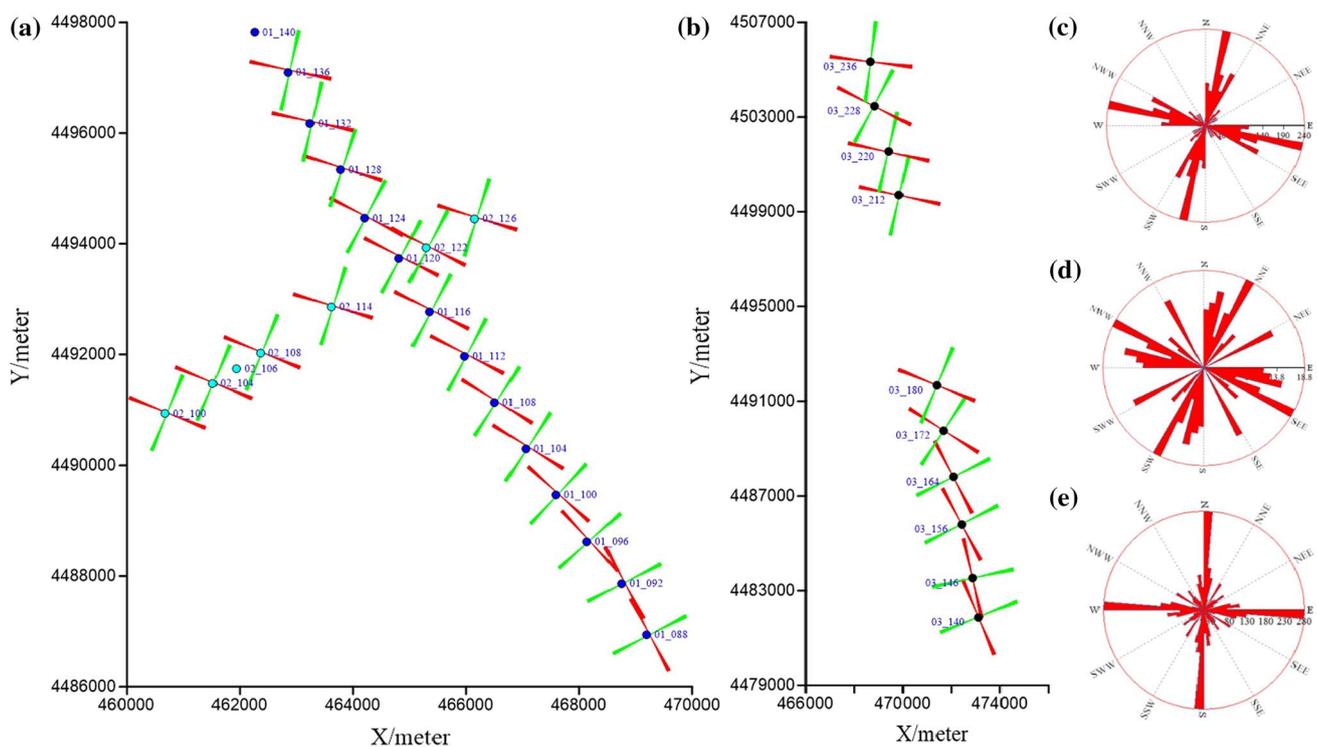


Fig. 6 Rose diagrams drawn by single measuring point and multipoint–multifrequency point statistics of electrical spindle of three survey lines. *Note:* **a** The electrical spindle azimuth of a single measuring point for lines L01 and L02. **b** The electrical spindle azimuth of a

single measuring point for Line L03. **c–e** Rose diagrams drawn using multipoint–multifrequency point statistics for the electrical spindle of lines L01, L02, and L03, respectively

of the layered medium model is essentially reserved in the one-dimensional inversion results and has little effect on the smoothing effect of the deep low-resistivity anomalous body.

It is of great importance that the polarization model data be selected for inversion, owing to the substantial differences between TE and TM model inversion results when observation data were rotated to tectonic strike for inversion (Chen et al. 2008). Previous studies have shown that TM mode response is less affected by three-dimensional effect than TE mode, ideally even to the point of allowing reconstruction of the original model information (Banks et al. 1996; Berdichevsky et al. 1998; Cai and Chen 2010b; Dong et al. 2012). Especially in the complex model, under complex model conditions, the inversion results obtained from the TM polarization model are more credible than those obtained using other models (Chen et al. 2014). To determine which model could reflect the actual electrical structure underground, TE, TE + TM, and TM data for L01 and L03S line were inverted, then compared with the log data of Well MAZD1. As shown in Fig. 7, TE and TE + TM model inversion results identified the underlying formations under 0.9 km as high-resistive layers, with no electrical interface inside. TM model inversion, however, additionally revealed discontinuous low-resistivity layers beneath the high-resistivity layers. Regional geology and drilling results showed that Carboniferous–Permian strata were developed in the lower Cretaceous of the Tuolai Sag, with thick low-resistivity clastic rocks beneath the Permian volcanic rocks and with Carboniferous strata also having the characteristics of low resistance and high resistance, so that the inversion results of the TM model were more reasonable and credible. For this reason, TM model data were ultimately selected for inversion, and the two-dimensional electrical profile obtained by inversion of the L01 line was compared with the resistivity logging of Well MAZD1 (LLD: deep laterolog, LLS: shallow laterolog; Fig. 7) to choose the optimal inversion parameters.

Additionally, the regularization factor (τ) is an important parameter for use in adjusting the roughness of the model and the weight of the fitting error (Hansen 1992). Inversion analysis was performed by selecting several τ values (1, 5, 10, 30, 50, 100, and 300). Taking model roughness as the transverse axis and the root-mean-square error (RMS) as the longitudinal axis when drawing the L-curve (Fig. 8), $\tau = 10$ was located at the curve's inflection point, so that reflecting the reversion result could guarantee the smoothness of the model while taking into account the relationship with the original data (Farquharson and Oldenburg 2004). Accordingly, a value of 10 at the inflection point was chosen as the model's τ value.

Ultimately, the following settings were chosen: 60 frequency points included in the inversion; frequency band 320–0.01 Hz, regularization factor 10; error floor for apparent resistivity and apparent phase 10%; and

root-mean-square error (RMS) of L01, L02, L03S, and L03 N of 1.81, 1.76, 1.92, and 2.07, respectively. The inversion results are shown in Fig. 9.

Electrical structural characteristics

The shallow portion of the three electrical structure profiles has a strong one-dimensional property, with characteristics of “vertical stratification” and low resistivity as a whole. The relatively low-resistivity formations above 960 m in L01 and L02 profiles, as determined by drilling Well MAZD1, are the Cretaceous Kuquan Formation and the Suhongtu Formation. Two sets of low-resistivity layers are developed at the top and bottom of Suhongtu Formation, with corresponding depths of 212–415 m and 600–850 m. These low-resistivity layers are mainly sandy mudstone interbedded with shaly sand, with intercalated thin mudstone. The north and south sections of survey line 03 (L03 N and L03S, respectively) have different electrical structural characteristics, so that the buried depth of the bottom interface of the low-resistivity layer in L03 N is relatively large and changes gently, perhaps reflecting the location of the line in the secondary depression. Whereas the bottom boundary depth of the low-resistivity layer of the L03S line first increased at first and then decreased along the line, indicating the development of a small depression between 140 and 164 points on the L03S line, the thickness of the Cretaceous in the study area changed only slightly on the whole, with no tectonic damaging present in the later period.

High resistors exhibiting significant north–south difference are developed under the higher set of above low-resistivity layers. The southern portion of the study area shows a large set of high-resistance block structures below sea level, whose burial depth gradually becomes shallower to the south. Considering that a large area of middle Archean metamorphic rocks having high resistivity is exposed about 9 km to southwestern side, the underlying high-resistivity block of L03S can be inferred to comprise Proterozoic metamorphic rocks. The transversely continuous high-resistivity layers are developed below the low-resistivity layer of the other three survey lines, which decrease in thickness from north to south but are stable from east to west, suggesting that the high-resistivity layer may be related to the north–south-trending tectonic movement. The depth of 960–2520 m for Well MAZD1 at Point L01_120 is revealed to be a set of pre-Carboniferous mylonites having shallow dynamic metamorphism, and it can be inferred that the high-resistivity layer is widely developed in the northern portion of the study area.

Notably, in the electrical structure profile, a set of low-resistivity layers with medium continuous development, stable thickness, and stepped distribution is developed under the high-resistivity layer of the three survey lines L01, L02, and L03 N. Any inferences about the geological

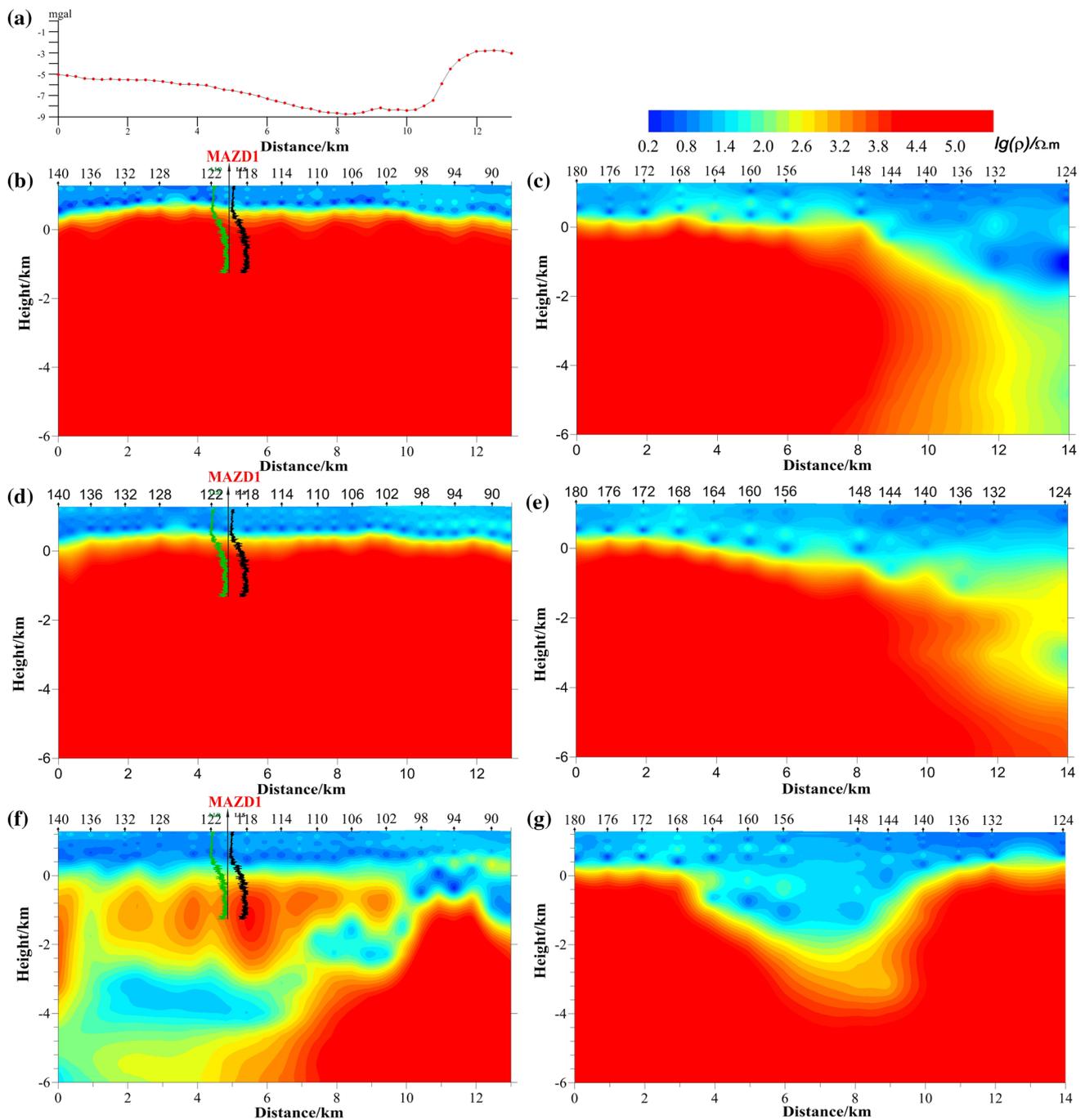


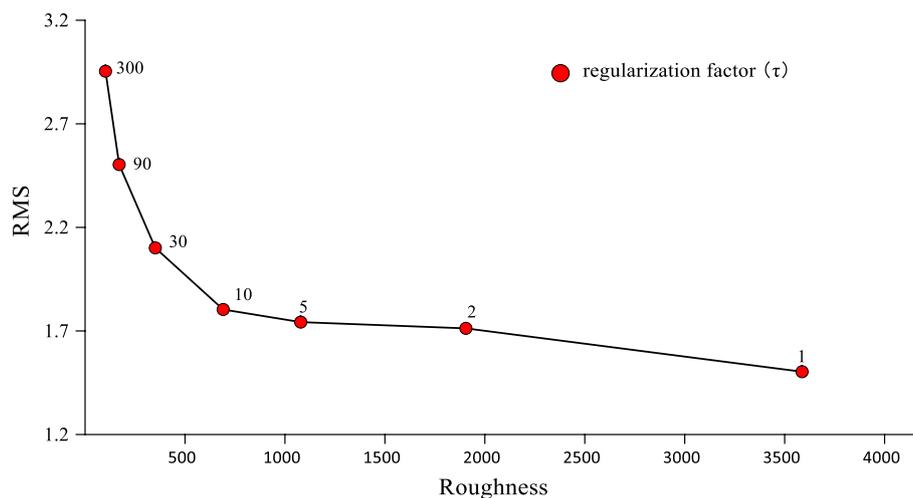
Fig. 7 Comparison of TE and TM inversion results of L01 and L03S lines. *Note:* **a** Residual gravity line curve of Line L01. **b** TE model inversion results for Line L01. **c** TE model inversion results for Line

L03; **d** TE+TM model inversion results for Line L01. **e** TE+TM model inversion results for Line L03S. **f** TM model inversion results for Line L01. **g** TM model inversion results for Line L03

implication of this low-resistivity layer must take into account regional geology and the rock's physical properties. As noted in Chapter 2 above, Well TL1, drilled in the eastern part of study area encountered Permian and Carboniferous strata underlying the Cretaceous, and Carboniferous strata in large scale are exposed on the north side of the Tuolai Sag. Electrical logging and near-well

sounding have demonstrated that the clastic rocks of the Cretaceous are characterized by low resistivity. Being thick in the north and thin in the south of the overlying high-resistivity body, the low-resistivity layer could be an electrical response of the Carboniferous strata. Accordingly, the overlying pre-Carboniferous high-resistivity layer is considered to be nappe.

Fig. 8 L-shape of RMS values and roughness for MT measures point of L01 when τ was changed



Geological–geophysical discussion

Pre-Carboniferous high-resistivity body may be a nappe

The area study shows that the Yin-E Basin entered the evolutionary stage of intracontinental rifting and taphrogeny in the Late Devonian. During this period, large-scale neritic shelf deposits or carbonate platform deposits were developed, forming several kilometers of continuous deposition of Carboniferous–Permian clastic/carbonate/volcanic rock formations, and the area experienced no Carboniferous–Permian hiatus (Lu et al. 2017). The strata experienced multiple structural transformations, such as the Late Hercynian, Indosinian, Yanshanian, and Himalayan periods, during which the Carboniferous–Permian in the Yin-E Basin suffered varying degrees of erosion. Currently, residual Carboniferous–Permian is distributed in most sags in the Yin-E Basin (Gillespie et al. 2016; Tian et al. 2016). In the Tuolai Sag, the Carboniferous–Permian is very thick, as verified in Well J1 on the northwest side (15 km to the study area) and Well TL1 on the east side of the study area, and the Carboniferous–Permian Amushan Formation is widely exposed in the northern boundary of the depression.

Carboniferous–Permian remains in the Tuolai Sag, but Well MAZD1 passed through Cretaceous to directly drill pre-Carboniferous, perhaps as the result tectonic evolution. In the electrical section, a low-resistivity layer, laterally continuous, monoclinically stepped, and distributed from north to south, is developed in the lower portion of the pre-Carboniferous, whose burial depth gradually decreases in the southern part of the section. The regional geological survey shows that the Carboniferous in the southern part of the depression is exposed in a nearly EW trend. The surface location is close to the southern uplift area of the low-resistance layer in the L01 line. As already stated, the view

that the low-resistance layer belongs to the Carboniferous is also supported by the results of near-well sounding and electrical logging. Second, using a preliminary geological model based on the electrical structure of the L01 line and constrained fitting with application of residual gravity, it can be found that if the L01 section is a normal sedimentary sequence, then the northern portion is a gravitationally anomalous low-value area—a finding not consistent with the measured gravity data. If the lower resistivity layer is considered to be Carboniferous–Permian, and the pre-Carboniferous is fitted as a nappe from west to south, then the inversion data fit the measured data well, indicating that the pre-Carboniferous high-resistance layer should belong to an allochthonous nappe (Fig. 10). In the Jurassic–Cretaceous, as a result of intensive thrust napping, a series of large-thrust nappe structures was formed in the area, aligned from north to south.

In addition, the history of regional tectonic evolution indicates that the stress environment was dominated by alternate compression and differential uplifting during tectonic evolution after Carboniferous–Permian sedimentation. In the Late Triassic, with the closure of the southern Tethys Ocean and continent–continent collision, the study area was in an intensively stress-filled environment. A tight east–west fold was widely developed in the Carboniferous–Permian in the basin. In the Jurassic–Cretaceous, as a result of intensive thrust napping, a series of large-thrust nappe structures were formed in the area, aligned from north to south. For instance, the Mujishan–Hurenwuzhuer area in the northern part of the study area, the “klippe” east–west distribution zone, and the large continental margin thrust nappe system were developed in the Bayinhot Basin in the south (Zhang et al. 2008), indicating the presence of a stress environment needed for thrust nappe in the regional structure. Considering the continuous distribution of Cretaceous strata, it can be speculated that thrusting of high-resistive bodies may have occurred

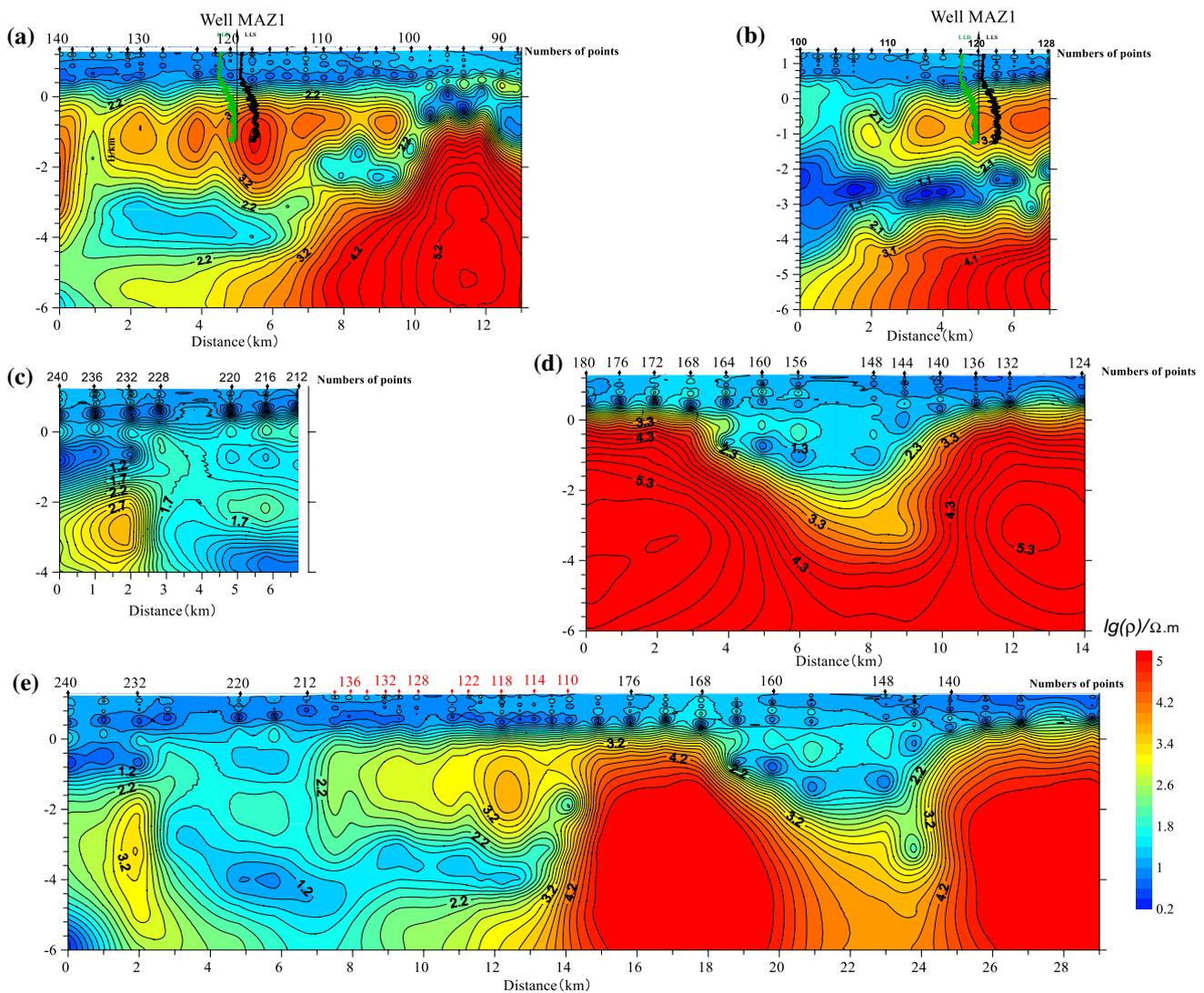


Fig. 9 Magnetotelluric sounding two-dimensional inverted electrical section. *Note:* **a** Two-dimensional electric section of survey line L01. Well MAZD1 is located 60 m south of Point 120. The green line is the deep lateral resistivity log curve (LLD), and the black line is the shallow lateral resistivity log curve (LLS). **b** Two-dimensional electrical section of survey line L02. The green and black lines are the same as shown in survey line L01. **c** Two-dimensional electrical

section of survey line L03 N. **d** Two-dimensional electrical section of survey line L03S. **e** To understand the overall structure of the sag, several points in Line L01 (from L01_110 to L01_138, the red arrow marked in the figure) are projected in the missing part of Line L03 (desert area). The positions of these measuring points on Line L03 are processed according to their projection positions

in the Late Triassic. Accordingly, it can be speculated on the basis of the comprehensive geological and geophysical data that the pre-Carboniferous high-resistive body is thrust nappe from the Late Triassic and that the corresponding underlying low-resistivity layer is Carboniferous–Permian.

Tectonic framework and strata distribution

Synthesizing two nearly NS-trending magnetotelluric electric sections, the study area overall shows a tectonic framework in which two subsags are sandwiched by an uplift (Fig. 11). The southern subsag is small, having a

north–south axis about 6 km long; the northern sag is large both in scale and in burial depth, and the sag boundary is not controlled by the north end of the survey line. The basin basement exhibits high resistivity features and fluctuates considerably in the central northern part, with stepped uplift from north to south. The burial depth of the southern part is small and is essentially stable. The northern sag can be divided into three structural layers longitudinally. Graben sags with faults in both sides are developed in the Cretaceous, with the faults manifesting as normal faults. Electrical anomalies are shown as vertical displacement with dense resistivity contours. Carboniferous is developed on

Fig. 10 Gravity fitting and comprehensive interpretation based on magnetotelluric two-dimensional electrical section (initially geological model based on the electrical structure of Line L01). *Note:* **a** Measured and fitted curves of residual gravity anomaly. **b** Density model based on residual gravity fitting. **c** Comprehensive interpretation map based on electrical structure model and density model

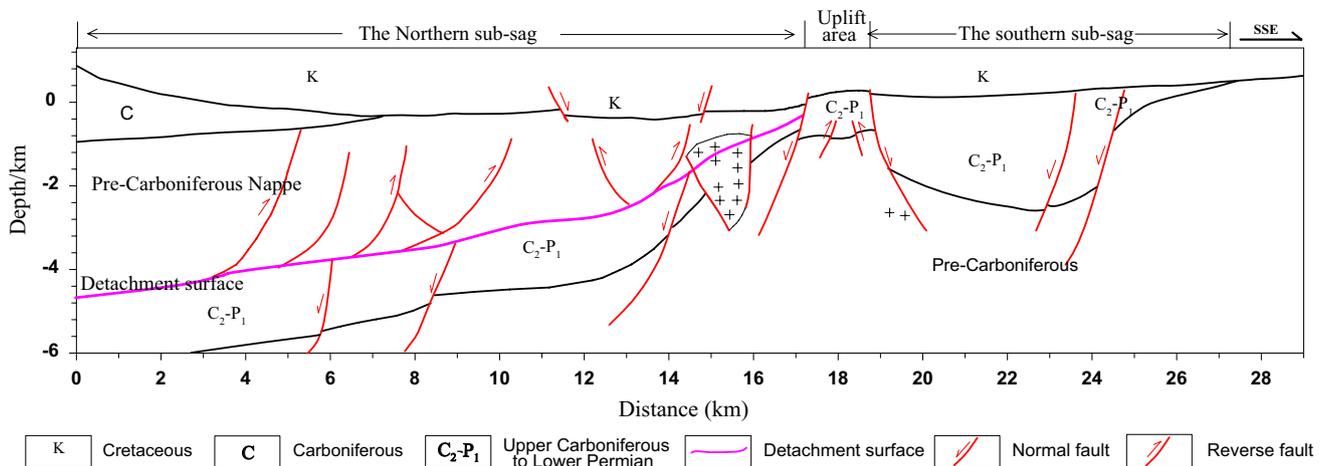
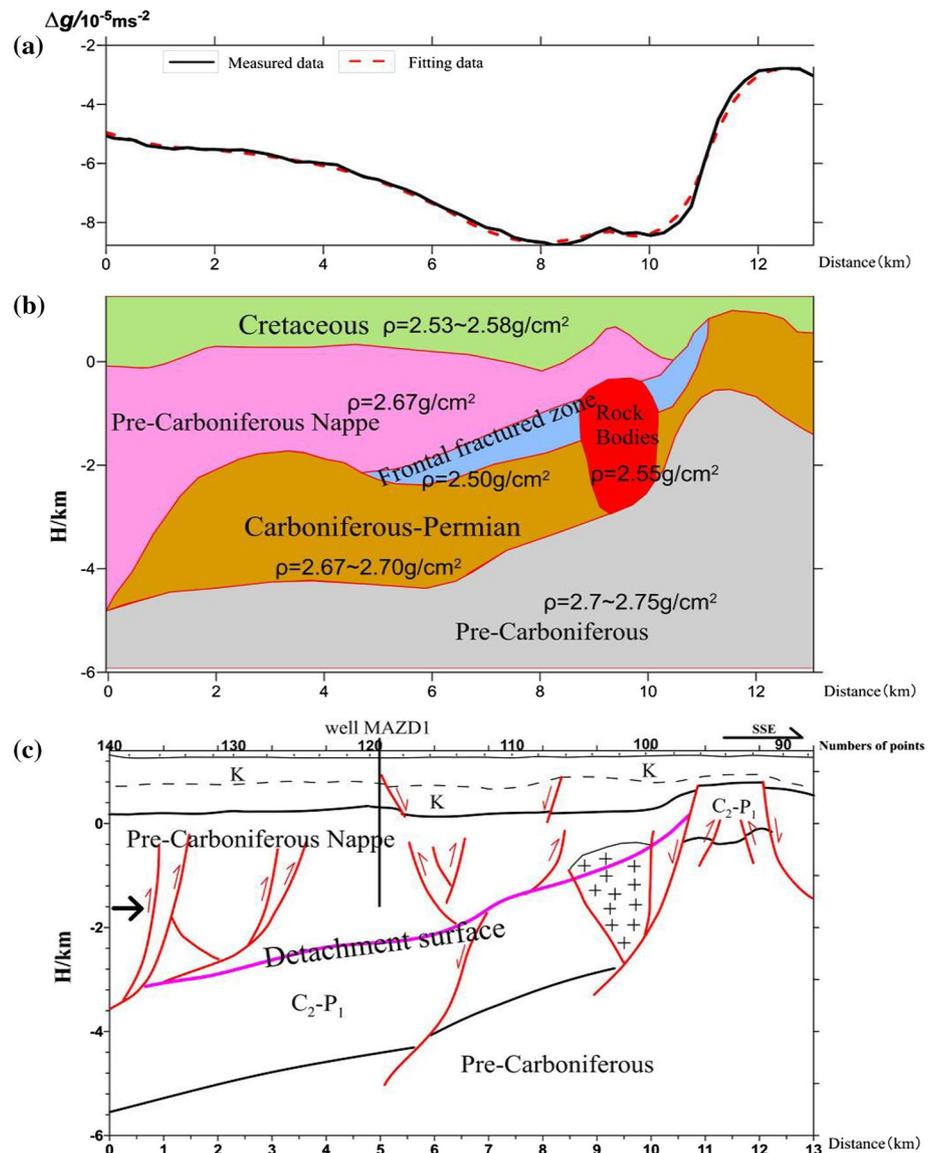


Fig. 11 Sketch map of the overall tectonic framework of the Tuolai Sag

the northern side of the survey line, presumably normal sedimentary strata overlying the nappe. The central part is pre-Carboniferous nappe consisting of quartz schist, which diminishes in thickness and displays thrust features from north to south. A number of nearly parallel thrust faults are developed in the nappe, of listric shape, steep in the upper part but gentle in the lower part, with alternately steep and gentle steps. The fault does not enter the overlying Cretaceous, and the lower portion becomes gradually more gentle, merging into the thrust detachment surface at the bottom of the nappe at depth. The detachment surface is shallow in the south and deep in the north and is mainly N-dipping. The burial depth is small at the southern boundary of the northern subsag (about 1.2 km), corresponding to the southern uplift. The lower part is Carboniferous–Permian, with a series of internal parallel normal faults. The burial depth gradually increases from south to north, and it is assumed to be a faulted structure. The southern sag manifests a single fault with a trough, and the burial depth of the basement is about 3 km, with boundary faults developed on the north side. Normal faults from the opposing dip to the boundary faults are developed in the gentle slope, together controlling the development and sedimentation of the fault trough.

The Cretaceous strata are continuously distributed horizontally in the study area. The thickness distribution is relatively stable overall (about 900 m) and is relatively thin in front of the nappe. Triassic–Jurassic is generally absent in the sag, and the formation does not outcrop in the peripheral surficial geology survey. The Carboniferous–Permian is buried beneath the detachment surface of the nappe in the northern subsag. The burial depth and thickness gradually decrease from north to south, with the thinnest portion appearing east of the margin of the northern subsag. The southern subsag is mainly filled with Carboniferous–Permian.

Conclusions

- (1) The shallow part of the study area has strong one-dimensional characteristics, and the deep part has mainly a two-dimensional structure. The direction of the major electrical axis in the sag is WNW and the inversion results of TM model data correspond more nearly to actual geological conditions.
- (2) The high-resistivity body is a nappe within the sag, exhibiting the following tectonic characteristics: thick in the north and thin in the south, deep in the north and shallow in the south. The nappe front is located on the north side of the central uplift zone. Multiple listric reverse faults are developed within the nappe, and a detachment surface is developed at the bottom. The low-resistivity body distributed continuously and

transversely in the footwall of the nappe is Carboniferous–Permian strata.

- (3) The Tuolai Sag shows a tectonic framework of two subsags sandwiched by an uplift. The southern subsag presents with small size and is in the form of a single fault with a trough. Boundary faults are developed in the north of the sag and are mainly filled with Carboniferous–Permian strata. The northern subsag is in large scale and has a complex structure. The Cretaceous fault depression, the pre-Carboniferous nappe, and the Carboniferous–Permian faulted structure developed from top to bottom, respectively.

Acknowledgements This research was jointly supported by the China Geological Survey (Grant Nos. DD20160164, DD20190030, JYYWF20180902). We truly appreciate the assistance of Dashuang He and Liang Chen from the Institute of Geophysical and Geochemical Exploration for the field work and data acquisition. We also would like to acknowledge our indebtedness to the editor and two reviewers for their kind work and comments.

Compliance with ethical standards

Conflict of interest We declare that we have no financial and personal relationships with a third party whose interests could be positively or negatively influenced by the article's content. On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Bahr K (1988) Interpretation of the magnetotelluric impedance tensor: regional induction and local telluric distortion. *J Geophys* 62(2):119–127
- Bahr K (1991) Geological noise in magnetotelluric data: a classification of distortion types. *Phys Earth Planet Inter* 66(1–2):0–38
- Banks RJ, Livelybrooks D, Jones PC, Longstaff R (1996) Causes of high crustal conductivity beneath the Iapetus suture zone in Great Britain. *Geophys J Int* 124(2):433–455. <https://doi.org/10.1111/j.1365-246X.1996.tb07031.x>
- Berdichevsky MN, Dmitriev VI, Pozdnjakova EE (1998) On two-dimensional interpretation of magnetotelluric soundings. *Geophys J Int* 133(3):585–606. <https://doi.org/10.1046/j.1365-246X.1998.01333.x>
- Bibby HM, Caldwell TG, Brown C (2005) Determinable and non-determinable parameters of galvanic distortion in magnetotellurics. *Geophys J Int* 163(3):915–930. <https://doi.org/10.1111/j.1365-246X.2005.02779.x>
- Cai JT, Chen XB (2010) Refined techniques for data processing and two-dimensional inversion in magnetotelluricII: which data polarization mode should be used in 2D inversion. *Chin J Geophys* 53(11):2703–2714. <https://doi.org/10.3969/j.issn.0001-5733.2010.11.018>
- Cai JT, Chen XB, Zhao GZ (2010) Refined techniques for data processing and two-dimensional inversion in magnetotelluricI: tensor decomposition and dimensionality analysis. *Chin J Geophys* 53(10):2516–2526. <https://doi.org/10.3969/j.issn.0001-5733.2010.10.025>

- Caldwell TG, Bibby HM, Brown C (2004) The magnetotelluric phase tensor. *Geophys J Int* 158(2):457–469. <https://doi.org/10.1111/j.1365-246X.2004.02281.x>
- Chen XB, Zhao GZ, Ma X (2008) Preliminary discussion on selecting rotation direction in 2-D MT inversion. *Oil Geophys Prospect* 43(1):113–118,128
- Chen XB, Cai JT, Wang LF, Ye T (2014) Refined techniques for magnetotelluric data processing and two-dimensional inversion(IV): statistical image method based on multi-site, multi-frequency tensor decomposition. *Chin J Geophys* 57(6):1946–1957. <https://doi.org/10.6038/cjg20140625>
- Danda N, Rao CK, Kumar A (2017) Geoelectric structure of northern Cambay rift basin from magnetotelluric data. *Earth Planets Space* 69(1):140. <https://doi.org/10.1186/s40623-017-0725-0>
- Dong H, Wei WB, Ye GF, Jin S, Jing TE, Zhang LT, Zhang F, Xie CL (2012) Study of two dimensional magnetotelluric inversions of complex three dimensional structures. *Chin J Geophys* 55(12):4003–4014. <https://doi.org/10.6038/j.issn.0001-5733.2012.12.012>
- Egbert GD (1997) Robust multiple-station magnetotelluric data processing. *Geophys J Int* 130(2):475–496
- Fang H, Zhong Q, Chen SW (2013) The discovery of two conductive layers in Tuquan basin, Inner Mongolia, and its geological significance. *Geol Bull China* 32(8):1253–1259
- Farquharson CG, Oldenburg DW (2004) A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems. *Geophys J R Astron Soc* 156(3):411–425. <https://doi.org/10.1111/j.1365-246X.2004.02190.x>
- Geng YS, Shen QH, Du LL, Song HX (2016) Regional metamorphism and continental growth and assembly in China. *Acta Petrol Sin* 32(9):2579–2608
- Gillespie J, Glorie S, Xiao WJ, Zhang ZY, Collins AS, Evans N, McInnes B (2016) Mesozoic reactivation of the Beishan, southern Central Asian Orogenic Belt: insights from low-temperature thermochronology. *Gondwana Res* 43:107–122. <https://doi.org/10.1016/j.gr.2015.10.004>
- Groom RW, Bailey RC (1989) Decomposition of magnetotelluric impedance tensors in the presence of local three-dimensional galvanic distortion. *J Geophys Res* 94(B2):1913–1925. <https://doi.org/10.1029/jb094ib02p01913>
- Hansen PC (1992) Analysis of discrete Ill-posed problems by means of the L-curve. *SIAM Rev* 34(4):561–580. <https://doi.org/10.2307/2132628>
- Kappler KN (2012) A data variance technique for automated despiking of magnetotelluric data with a remote reference. *Geophys Prospect* 60(1):179–191
- Liu JL, Shen AB, Chen XL (2011) Application of magnetotelluric sounding for Carboniferous–Permian petroleum geological survey in Yingen-Ejin Banner basin, western Inner Mongolia. *Geol Bull China* 30(6):993–1000. <https://doi.org/10.1007/s11589-011-0776-4>
- Liu JL, Li XZ, Zhang Q (2013) The application of gravity-magnetic-magnetotelluric joint inversion to the quantitative interpretation of yin’e basin. *Geophys Geochem Explor* 37(5):853–858. <https://doi.org/10.11720/j.issn.1000-8918.2013.5.18>
- Lu JC, Chen GC, Wei XY, Li YH, Wei JS (2011) Carboniferous–Permian sedimentary formation and hydrocarbon generation conditions in Ejin Banner and its vicinities, western Inner Mongolia: a study of Carboniferous–Permian petroleum geological conditions (part 1). *Geol Bull China* 30(6):811–826
- Lu JC, Chen GC, Li YH, Wei XY, Wei JS, Jiang T, Shi JZ, Dang B, Zhao XM, Liu JL, Yang GY, Chen JF, Bu JJ, Han W, Li W (2012) Carboniferous–Permian geological conditions and resources perspective in Yingen-Ejin banner basin and its vicinities. Geological Publishing House, Beijing, pp 1–439
- Lu JC, Chen GC, Li YH, Wei XY, Zhao XM (2014) Main progress and achievements of the Permo-Carboniferous petroleum prospective survey in Yine Basin and its surrounding areas. *Geol Surv China* 1(02):35–44
- Lu JC, Zhang HA, Niu YZ, Liu HC, Chen QT, Wei JS (2017) Carboniferous–Permian petroleum conditions and exploration breakthrough in the Yingen-Ejin Basin in Inner Mongolia. *Geol China* 44(1):13–32
- Rodi W, Mackie RL (2001) Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion. *Geophysics* 66(1):174–187. <https://doi.org/10.1190/1.1444893>
- Shalivahan N, Bhattacharya BB (2002) How remote can the far remote reference site for magnetotelluric measurements be? *J Geophys Res* 107(B6):2105
- Simpson F, Bahr K (2005) Practical magnetotellurics. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511614095>
- Tian ZH, Xiao WJ, Zhang ZY, Lin X (2016) Fission-track constrains on superposed folding in the Beishan orogenic belt, southernmost Altaiids. *Geosci Front* 7(2):181–196. <https://doi.org/10.1016/j.gsf.2015.11.007>
- Wei PS, Zhang HQ, Chen QL (2006) Petroleum geological characteristics and exploration prospects of Yingen-Ejin Banner Basin and its vicinities. Petroleum Industry Press, Beijing, pp 1–345
- Ye GF, Sheng J, Wei WB, Unsworth M, Unsworth M (2007) Research of conductive structure of crust and upper mantle beneath the South-Central Tibetan Plateau. *Earth Sci J China Univ Geosci* 18(4):491–498. [https://doi.org/10.1016/S1002-0705\(08\)60014-X](https://doi.org/10.1016/S1002-0705(08)60014-X)
- Ye T, Chen XB, Yan LJ (2013) Refined techniques for data processing and two-dimensional inversion in magnetotelluric(III): using the impressing Method to construct starting model of 2D magnetotelluric inversion. *Geophysics* 56(10):3596–3606. <https://doi.org/10.6038/cjg20131034>
- Zhang JS, He ZX, Fei AQ, Li TB, Huang XN (2008) Epicontinental mega thrust and nappe system at north segment of the western rim of the Ordos Block. *Sci Geol Sin* 43(2):251–281
- Zhao WJ, Zhao HQ, Guo XD, Sun ZR (2016) Magnetotelluric sounding in the Jingxin Basin, Jilin, China. *Prog Geophys* 31(4):1542–1549



Assessing the effects of 1D assumption violation in vertical electrical sounding (VES) data processing and interpretation

Seyyed Reza Mashhadi¹ · Hamidreza Ramazi¹ · Mohammad Asgarpour Shahreza¹

Received: 15 October 2018 / Accepted: 25 October 2019 / Published online: 4 November 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

This research aimed to discover the possible effects of 1D assumption violations on VES data interpretations. In order to do so, 1D inversion results of logarithmically spaced and linearly spaced VES measurements are compared with their relevant 2D inverted models. Some real case studies are also examined by 1D and 2D inversions to test the results. It is found that linearly spaced VES measurements are not really suitable for 1D inversion in the case of 1D assumption violations and logarithmically spaced VES can better handle these problematic features. In the case of semi-infinite horizontal layers and also small surface resistivity inhomogeneities, logarithmically spaced VES datasets mostly provide a reliable 1.5D model while linearly spaced VES datasets suffer from remarkable artifacts. In the case of vertical structures, both linearly spaced and logarithmically spaced VES techniques fail. In this case (i.e., a vertical dike), artifacts in the form of “extra layer” appear in those VES stations that are adjacent to the dike. However, for VES stations on the dike structure, no extra layer appears in the 1D inversion result. It must be emphasized that 1D violating features are not improbable in many geological situations so they must be considered in mind when processing and interpreting the geophysical VES data.

Keywords Resistivity · Vertical electrical sounding (VES) · 1D assumption · Inversion · 2D/3D inhomogeneity

Introduction

Resistivity surveying is a routine geophysical method for exploration, environmental and engineering studies. It was often applied in the form of sounding and profiling in the past (Reynolds 2011). However, after the great developments in the past two decades, this method is mostly used in the form of 2D, 3D and even 4D surveys (Loke 2019; Loke et al. 2013). These days, 2D and 3D resistivity surveys are routinely undertaken due to the availability of inversion softwares, multichannel measurement systems and new data acquisition techniques, i.e., roll-along technique (Dahlin et al. 2002).

Despite the advances in resistivity imaging, vertical electrical sounding (VES) measurements are still a common tool for many investigations, especially for studying groundwater aquifers. Many researchers have used VES data alone, or in conjunction with other data, to estimate the hydraulic parameters and aquifer characteristics (Tizro et al. 2012; Sattar et al. 2014; Pedromo et al. 2014; Asfahani 2016; Topolewska et al. 2016), to study sea-water intrusions (Song et al. 2007; Hodlur et al. 2006), to delineate groundwater pollutions (Sundararajan et al. 2012; Mota et al. 2004; Porsani et al. 2004), to assess the piping phenomena in landslide (Sajinkumar et al. 2015), and to map deep sediments trapping oil reserves (Veeraiah & Babu 2014). Today, with multichannel measurement systems, the sounding is performed for the subsequent pair of electrodes, which is called continuous vertical electrical sounding (CVES). The CVES data are usually treated as a 2D/3D dataset where 2D/3D inversion is undertaken, i.e., Danielsen and Dahlin (2009). However, 1D inversion is sometimes performed to estimate the true resistivity and thickness of subsurface structures, i.e., Sundararajan et al. (2012).

In 1D inversion of resistivity data, there are two fundamental problems: the principle of equivalence and the

✉ Seyyed Reza Mashhadi
Reza.mashhadi93@aut.ac.ir; Reza.mashhadi93@yahoo.com

Hamidreza Ramazi
Ramazi@aut.ac.ir

Mohammad Asgarpour Shahreza
mohamadasgarpour69@gmail.com

¹ Department of Mining and Metallurgical Engineering, Amirkabir University of Technology (Tehran Polytechnic), 424 Hafez Ave, 15875-4413 Tehran, Islamic Republic of Iran

principle of suppression (Reynolds 2011). These principles are added to the main geophysical inverse problems, making the 1D inversion more complicated than 2D and 3D inversions. Theoretically, the results of 2D inversion are better and more realistic than 1D inversion. The 3D inversion is also preferable in comparison to 2D/1D (Loke 2019). While the pros and cons of 2D and 3D inversions have been greatly assessed in the past two decades, 1D inversion was not really examined and there are a few pieces of researches on this topic. Some examples of researches on improving the 1D resistivity inversion are Zohdy (1989), Gupta et al. (1997), Gyulai and Ormos (1999), Singh et al. (2005), Kumar et al. (2007) and Gyulai et al. (2010).

In this paper, it tried to understand the possible effects of 1D assumption violations on VES data processing and interpretation. The 1D and 2D inversion results of the linearly spaced and logarithmically spaced VES measurements are compared and analyzed in detail. It is very important to have a clue about the possible inaccuracies or incapacities of 1D inversion results in the case of 1D violating structures. That is because “these structures are considered to be totally absent when designing a VES survey.” However, this assumption is intrinsically questionable in many cases, and several violating structures could be present in the study area, i.e., different inhomogeneities within a layer, edge of horizontal layers (sedimentary interlayers), vertical resistivity interfaces created by faulting, etc. To reach the main goal of the research, we have followed these steps: (1) some synthetic models are prepared that represent the 1D violating features, (2) several VES stations are selected from each model’s dataset, (3) the VES data is processed by 1D and 2D inversion routines and (4) finally, 1.5D and 2D models are provided together in order to make a suitable comparison between the real subsurface features and geophysical responses. In the following sections, these stages are discussed in detail.

Theoretical background

The “sounding” and “profiling” terms are well known in geophysics. From the 1920s to the late 1980s, resistivity surveying was performed using these two techniques (Loke et al. 2013). In resistivity profiling method, the distances between the electrodes were kept fixed and the four electrodes were moved along the survey line to assess the resistivity variability in the desired direction. In resistivity sounding technique, while the center point of the electrode array remains fixed, the spacing between the electrodes is increased to obtain more information about the deeper parts of the subsurface (Reynolds 2011; Loke 2019).

Vertical electrical sounding (VES) is a 1D electrical surveying technique. It assumes that the subsurface consists of

several horizontal layers with different electrical resistivities. This assumption means that the subsurface resistivity changes only with depth, but does not change in the horizontal directions (Loke 2019). Resistivity sounding/VES can be performed by using different electrode arrays but Schlumberger, Wenner and dipole–dipole have been used more frequently. Among the mentioned arrays, Schlumberger is preferable because of its high signal-to-noise ratio (SNR) and also easier data acquisition procedure (Reynolds 2011).

To interpret the sounding data, apparent resistivity values are plotted on a log–log graph (sounding curve or field curve) where x- and y-axes represent current electrode half-separation (or the outer active electrode spacing) and the apparent resistivity values, respectively (Reynolds 2011; Loke 2019). In the early stages of VES method development, standard or pre-computed curves were used to quantitatively interpret the data. Manual curve-fitting process was a laborious and time-consuming task. Years later, VES data inversion softwares became available that was based on semi-automatic or automatic methods (Loke et al. 2013).

Consider a three-layer model where the middle layer is conductive (relative to the upper and lower layer). This situation makes the current lines to converge and pass through the conductive layer. The longitudinal conductance (ratio of resistivity to thickness) for this layer is constant, and as long as the thickness and resistivity are changed (within limits) so as to maintain that ratio, there will be no appreciable change in the resulting apparent resistivity curve. In this case, all the pairs of h^2/ρ are electrically equivalent and no single pair of values is preferable to any other. The same effect happens when there is a resistive layer between two more conductive layers, where the current lines tend not to pass through the second layer. This problem is addressed as the principle of equivalence in VES data inversion (Reynolds 2011). Another significant issue is the principle of suppression. When more than 2 or 3 layers exist in the subsurface, the effect of the middle layers would not be visible on the curve (Reynolds 2011). In some cases, when the layers are too small, this problem could be related to the decrease in resolution with depth in geoelectric (Loke 2019), which makes these layers naturally undetectable. In VES curve interpretation, an additional form of equivalence must be taken into account which is the equivalence between an isotropic layer and an anisotropic layer (Reynolds 2011). This problem (electrical anisotropy) is a really important characteristic which can severely affect the response of subsurface materials. As Greenhalgh et al. (2010) have shown, the anisotropy has a severe effect on the sensitivity function (it greatly differs from the isotropic one, which is the basis of 2D geoelectrical inversion these days). Here, 2D inverse modeling of resistivity data is not discussed. Comprehensive notes about 2D and 3D resistivity inversion are presented by Loke (2019).

Electrical resistivity modeling

Synthetic modeling

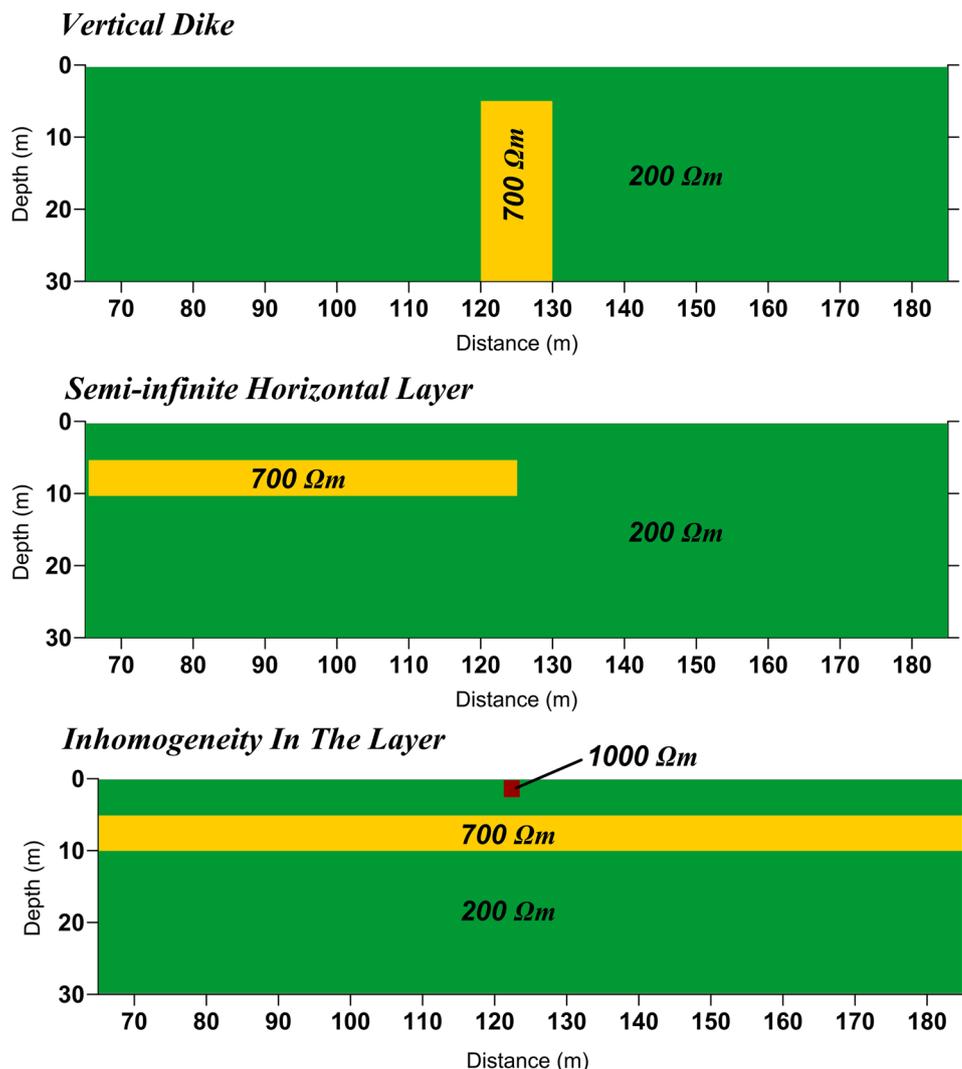
To assess the possible effects of 1D assumption violations on VES data interpretations, we have examined some simple synthetic models (Fig. 1). These models include

1. *Vertical dike* This model is a vertical dike that its top part is located at the depth of 5 m. Its horizontal boundaries are at $X=120$ m and $X=130$ m. From the viewpoint of electrical property, it is a resistive feature ($700 \Omega\text{m}$) in a more conductive background ($200 \Omega\text{m}$).
2. *Semi-infinite horizontal layer* It is a semi-infinite horizontal layer where its edge is located at $x=125$ m. This model was created to see whether the edge of a layer could affect 1D inversion results or not.

3. *Inhomogeneity in the layer* This model is a simple three-layer model with a 2.5 m by 2.5 m high resistivity inhomogeneity at the surface ($\rho=1000 \Omega\text{m}$, the boundaries are at $X=121.25$ m and $X=123.75$ m, depth of the boundaries are 0 and 2.5 m).

The electrical responses of the synthetic models have been calculated by using the RES2DMODE software (a forward modeling program for resistivity data). This software provides the theoretical response that would be observed in the field if there is no noise (RES2DMOD Manual). At the next stage, we have extracted the desired data for each scenario (logarithmically spaced and linearly spaced configurations) for each of the sounding stations. Finally, the extracted data were subjected to 1D and 2D inversion schemes to compare the results. The 1D and 2D inversions are done by means of IX1D and RES2DINV softwares, respectively. Surfer software is used for graphical editing.

Fig. 1 The synthetic models used to assess the possible effects of 1D assumption violations on VES data interpretations



Before starting the main discussions, some points are noteworthy:

1. The effects of 1D assumption violations on VES data interpretations are assessed by means of simple 2D models. Despite the simplicity of the synthetic models, some of the common effects have been clearly characterized.
2. Although all types of arrays can be used for sounding measurements, only the Schlumberger array is discussed since it is more frequently utilized. Some discussions are based on the sensitivity concept (Loke 2019), so the sensitivity section of Schlumberger array for a homogenous half space is presented in Fig. 2.
3. The dataset is provided by using a 2D resistivity forward modeling program. In other words, it is assumed that the electrodes are spread only perpendicular to the strike of 2D violating structures. This is not the best scenario since the VES electrodes might not be positioned exactly perpendicular to the strike of the mentioned features (especially if we are unaware of the geometrical characteristics of the undesirable features, i.e., their shape, strike, dip, etc.). In fact, the VES measurements might be performed in arbitrary directions with respect to the 1D violating structures. Despite this fact, the analysis could provide good insights into the problem.
4. Inevitably, avoiding all noises at all levels is impossible in the field. However, it would be theoretically better to examine the datasets in the ideal situation. So, synthetic datasets are examined without adding any noise.
5. The selected resistivity contrast is high in all models. Also, there are abrupt interfaces in the models. This criterion represents an ideal situation for a resistivity survey. On the other hand, different models with different dimensions and different resistivity contrasts were created and tested by both 1D and 2D inversions. However, some of them are presented here and only some indicative examples are shown.
6. The resistivity distribution is ideal for any 2D inversion method. Although the inversion results were almost identical for both l_1 -norm (blocky) and l_2 -norm (smooth) inversions, the result of the smooth inversion is presented. In fact, gradual resistivity variations are more common with respect to abrupt changes.
7. Bentley and Gharibi (2004) have used too many iterations in the 2D inverse modeling of noise-free synthetic datasets. However, no more than five iterations were adopted in this study since it was meant to avoid data overfitting which can disrupt the estimated resistivity values (Hauck and Kneisel 2008).
8. In the sounding curves, the scale of both x- and y-axes must be the same (especially if we are going to do the curve-fitting manually!). To be able to see the differences more clearly, the log–log axes of the sounding curves do not have the same scales in the following sections.

Linearly spaced versus logarithmically spaced VES measurements

In the traditional VES technique, the current electrode spacing was increased logarithmically to achieve a higher current penetration depth (Barker 1989). Furthermore, after performing several apparent resistivity readings, the potential electrode spacing was also increased to prevent high geometric factors which may result in noisy recordings (Rucker and Glaser 2015). On the other hand, in the CVES technique, the current electrode spacing was increased linearly while the potential electrode spacing would be kept fixed (Danielsen and Dahlin 2009). It is vital to discriminate between these two scenarios of VES measurements since they have different characteristics which can affect the credibility of the inverted models provided by different 1D/2D algorithms. Consequently, throughout this research, we have taken these facts into account. The current and potential electrode spacings for the used datasets are presented in Table 1.

From the viewpoint of field operation, logarithmically spaced intervals were more convenient with the old-fashioned one-channel DC resistivity meters. Today, multicore cable multichannel measurement systems are available for electrical surveying from different manufacturers around the

Fig. 2 Sensitivity section of Schlumberger array for a homogenous half space (in the case that $AB = 1$ m and $MN = 0.2$ m). It is calculated based on the 3D sensitivity function

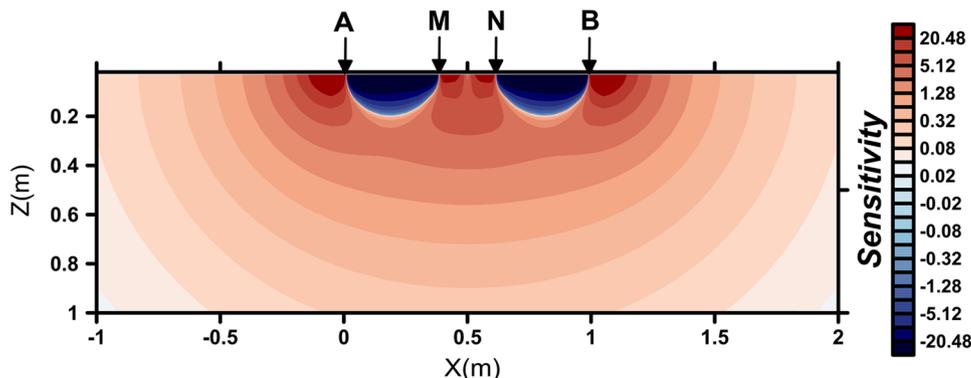


Table 1 The current electrode spacing and potential electrode spacing of the used datasets

VES measurement scenario			
Logarithmically spaced		Linearly spaced	
AB	MN	AB	MN
7.5	2.5	22.5	7.5
12.5	2.5	37.5	7.5
22.5	2.5	52.5	7.5
37.5	2.5	67.5	7.5
22.5	7.5	82.5	7.5
37.5	7.5	97.5	7.5
52.5	7.5	112.5	7.5
82.5	7.5	127.5	7.5
112.5	7.5	142.5	7.5
157.5	7.5	157.5	7.5

world. Commercial manufacturers mainly provide multicore cables with linearly spaced takeouts. So, VES surveys might be performed by linearly spaced rather than logarithmically spaced intervals these days, especially for small-scale investigations.

Inverse modeling of synthetic data

We have inverted the synthetic datasets by 1D and 2D inversion algorithms provided in IX1D and Res2dinv inversion softwares, respectively. Note that linearly spaced VES dataset is inverted by different unit electrode spacings (2.5 m and 7.5 m). It was meant to ensure about the final inverted models since unit electrode spacing (UES) can have extreme effects on the inversed models. As Loke (2019) has examined, it can have different negative or positive effects on the final electrical model. Selecting a small UES may improve the model by decreasing the distortions. On the other hand, it might prevent the deeper structures to be imaged in the inversion process. The original UES of the used linearly spaced VES measurements is 7.5 m. The original UES of logarithmically spaced VES dataset is 2.5 m, so the resolution of these cases would be different. That is why the UES = 2.5 m is also used for linearly spaced VES measurements to have a better comparison. Note that the same data processing procedures and also similar inversion parameters are used throughout this research. Furthermore, an ideal dataset for 2D ERT with the UES = 2.5 m is also analyzed to show the results of the 2D survey (a CVES dataset with UES = 2.5 m, having two more VES stations between S_i and S_{i+1} with linearly spaced Schlumberger array readings). This is done to enhance the capabilities and resolution of 2D data acquisition scenario.

1D inversion without any bias would face extreme challenges since there are a lot of model parameters that would lead to an acceptable rms error (as a consequence of the

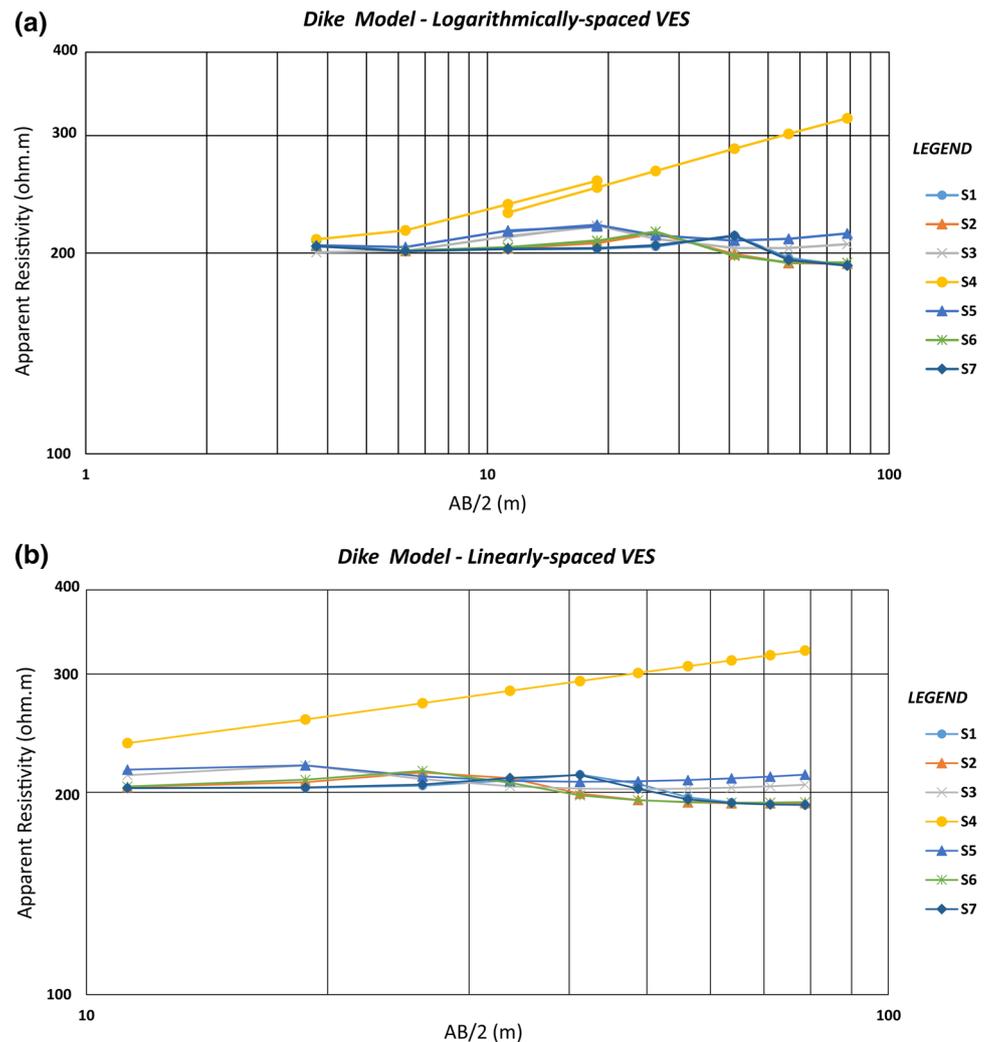
principle of equivalence or generally the nonuniqueness in geophysical inverse theory). In fact, this is a common problem in VES data interpretation, and for a better thickness and resistivity estimations, primary information about the range of resistivities and thicknesses should be available. Another reason for the low rms error is the data itself, which is free of noise. Curve overfitting (using too many iterations, just decreasing the rms error) must be avoided to prevent artifacts and/or data overinterpretation (Hauck and Kneisel 2008). However, in many cases, overfitting would not be clearly definite for the interpreter. According to the performed analysis, curve overfitting in 1D inverse modeling approach would change the resistivity and thickness of the middle layers while it keeps the resistivity of the first and last layers almost fixed. This fact implies the great effect of surface layers in 1D resistivity inverse modeling. The importance of the resistivity of surface layers is also mentioned in 2D and 3D inversions, even if the surface layers are not noticeable as a geophysical target (Dahlin et al. 2002). Consequently, near-surface structures are really important in all of the electrical surveys. In the VES technique, knowledge about the resistivity of the surface layers would be crucial for good data processing since too many models could fit the measured data in a reasonable way. As a result, it might be a good idea to perform some resistivity measurements on representative samples taken from the first layers (whenever possible). It is also preferred to make sure about the apparent resistivity readings of the surface measurements by recording reciprocal measurements and/or performing several repeated readings.

Vertical dike model

The apparent resistivity curves of different VES stations on the dike model are presented in Fig. 3. For logarithmically spaced VES curves, a two-layer model over the dike can be interpreted. However, for other VES stations that are not on the dike structure, a two-layer model or three-layer model would be fitted. In fact, extreme variations exist that a one-layer model would be considered unrealistic. The same patterns and explanations can be mentioned for the linearly spaced VES curves. The results of 1D and 2D inversions are also shown in Fig. 4. Table 2 lists the rms errors of 1D inversion for linearly spaced and logarithmically spaced VES datasets. The rms error of 2D inversions is also provided in Table 3.

The 1D inversion results of logarithmically spaced and linearly spaced VES datasets are presented in Fig. 4a, b, respectively. It is very interesting that a suitable connection of the interpreted interfaces is impossible and a geologically reasonable model cannot be proposed for both VES measurement scenarios. Clearly, if there is any information about the existing dike, the VES measurement would not

Fig. 3 Apparent resistivity curves of the logarithmically spaced and linearly spaced VES measurements for the “vertical dike” model

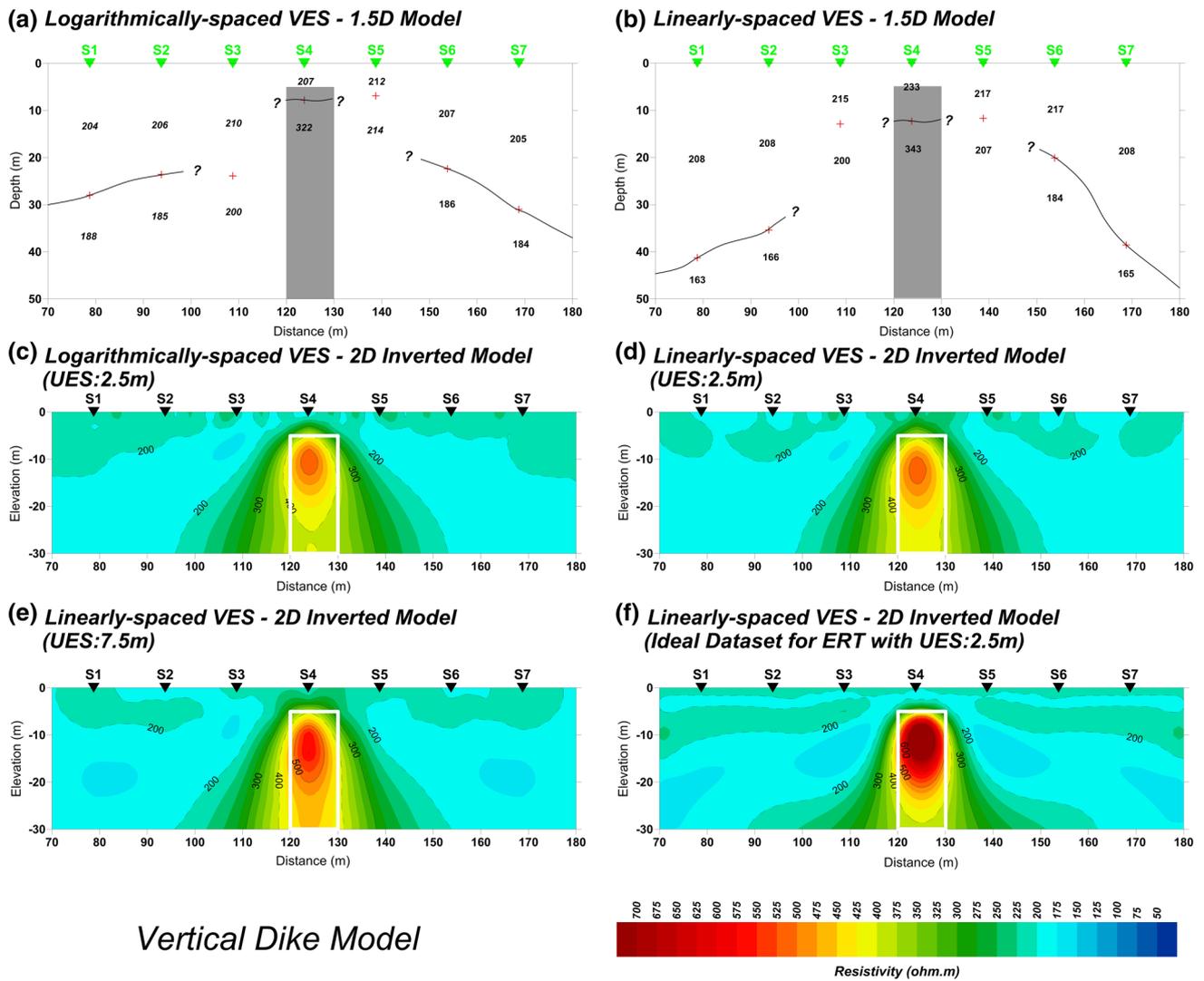


be performed. However, if it is not known, the 1D inversion results would be affected even if the VES station is not located on the dike structure. The artifacts in the form of “extra layer” can appear in the VES stations that are adjacent to the dike, which in turn, create an unrealistic layer in the 1.5D model. This point implies that vertical resistivity interfaces can severely affect the VES measurement results even if the interface is not located close to the VES station. In fact, if current electrodes reach the vertical interface, the negative effects might distort the final results.

Clearly, the effect of the vertical dike on VES data is not dispensable at all. The reason for these disturbances can be discussed by the sensitivity concept (refer to Fig. 2). As the high sensitivity parts in the measurement reach the dike, the apparent resistivity values are affected. Consider the VES station which is located beside the dike. When the current length (AB) is increasing in Schlumberger array, one of the current electrodes is getting closer to the resistive dike. At one point, where it is close enough, the high positive sensitivity of the mentioned electrode configuration enters the

resistive dike, leading to an increase in the measured voltage and consequently increasing the measured apparent resistivity. As the current electrode moves further, the resistive dike lies between the current and potential electrodes. In this situation, if the resistive dike is close enough to the surface, it enters the high negative sensitivity part, leading to a decrease in measured voltage and consequently decreasing the measured apparent resistivity. These phenomena lead to the mentioned disturbances and pose a negative effect on the 1D inversion result.

The 2D inversion result for logarithmically spaced VES measurement scenario is shown in Fig. 4c. The results of linearly spaced VES measurements are presented in Fig. 4d ($UES = 2.5$ m) and e ($UES = 7.5$ m). It is visible that linearly spaced models (Fig. 4d, e) have fewer surface irregularities in the form of surface high resistivity features with respect to logarithmically spaced 2D model (Fig. 4c). Note that lesser distortions appeared for $UES = 7.5$ m with respect to $UES = 2.5$ m in the linearly spaced 2D models, indicating the negative effects of smaller unit electrode spacing. The



Vertical Dike Model

Fig. 4 The results of 1D and 2D inversion algorithms for the “dike model”

Table 2 The rms error of different VES stations in the 1D inversion approach for the “vertical dike” model

	VES station	S1	S2	S3	S4	S5	S6	S7
rms error (%)	Logarithmically spaced VES	2.40	1.96	3.01	1.53	2.41	2.22	1.99
rms error (%)	Linearly spaced VES	2.09	2.07	1.49	0.72	1.31	1.96	2.06

Table 3 The rms error of the represented 2D inverted models of the “vertical dike”

Type of VES	Logarithmically spaced	Linearly spaced	Linearly spaced	Ideal dataset
Unit electrode spacing (UES) in meters	2.5	2.5	7.5	2.5
rms error (%)	0.37	0.38	0.40	0.27

resistivity values are estimated more precisely in the linearly spaced VES with the UES = 7.5 m. The estimated resistivity values in the linearly spaced models are better than the logarithmically spaced model. The ideal dataset for an electrical resistivity tomography (ERT) survey is also prepared with the original UES = 2.5 m, and it is inverted by the 2D inversion algorithm (Fig. 4f). This model is better than any other 2D model provided here both in terms of distortions and estimated resistivity values. This has two reasons: (1) the number of datapoints in the complete dataset is higher leading to higher resolution and also (2) it does not suffer

from the negative effects resulting from the decreasing the original UES of the dataset. Note that the 2D inverted models are not suffering from remarkable artifacts.

It is notable that the resistivity values are underestimated by both 1D and 2D inversion routines. A similar phenomenon is reported by Chambers et al. (2001). They pinpointed that the resistivity of buried resistive features is usually underestimated. Furthermore, it is visible that the boundaries of the dike cannot be imaged precisely, especially in the deeper parts of the 2D model. Note that some inefficiencies of the 2D model are related to the array characteristics as Schlumberger array is mainly suitable for mapping horizontal resistivity variations (Loke 2019).

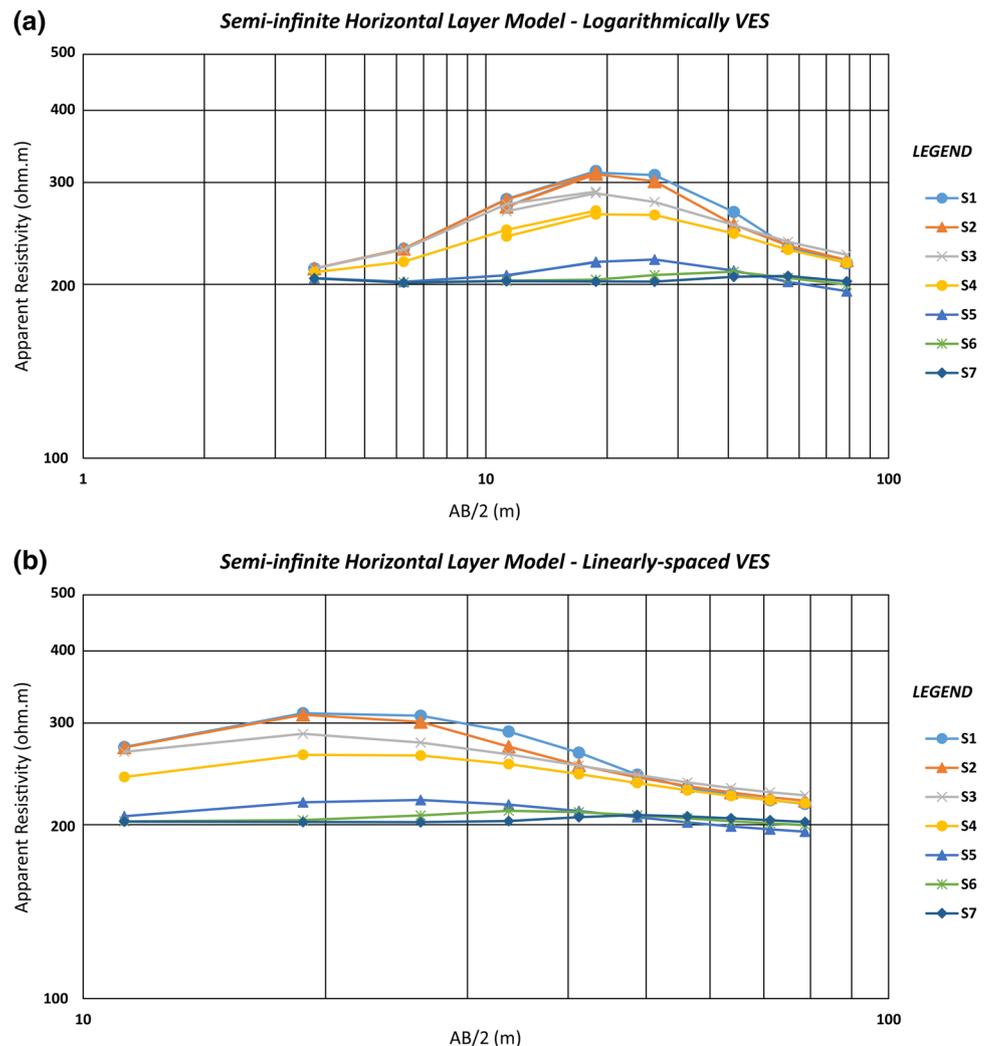
To conclude, the synthetic modeling reveals that vertical features could extremely affect 1D and the subsequent 1.5D model of VES datasets, especially when these structures have great dimensions and/or their resistivity contrast with respect to the background material is noticeable. Additionally, the dike structure itself is not expectable to be seen in

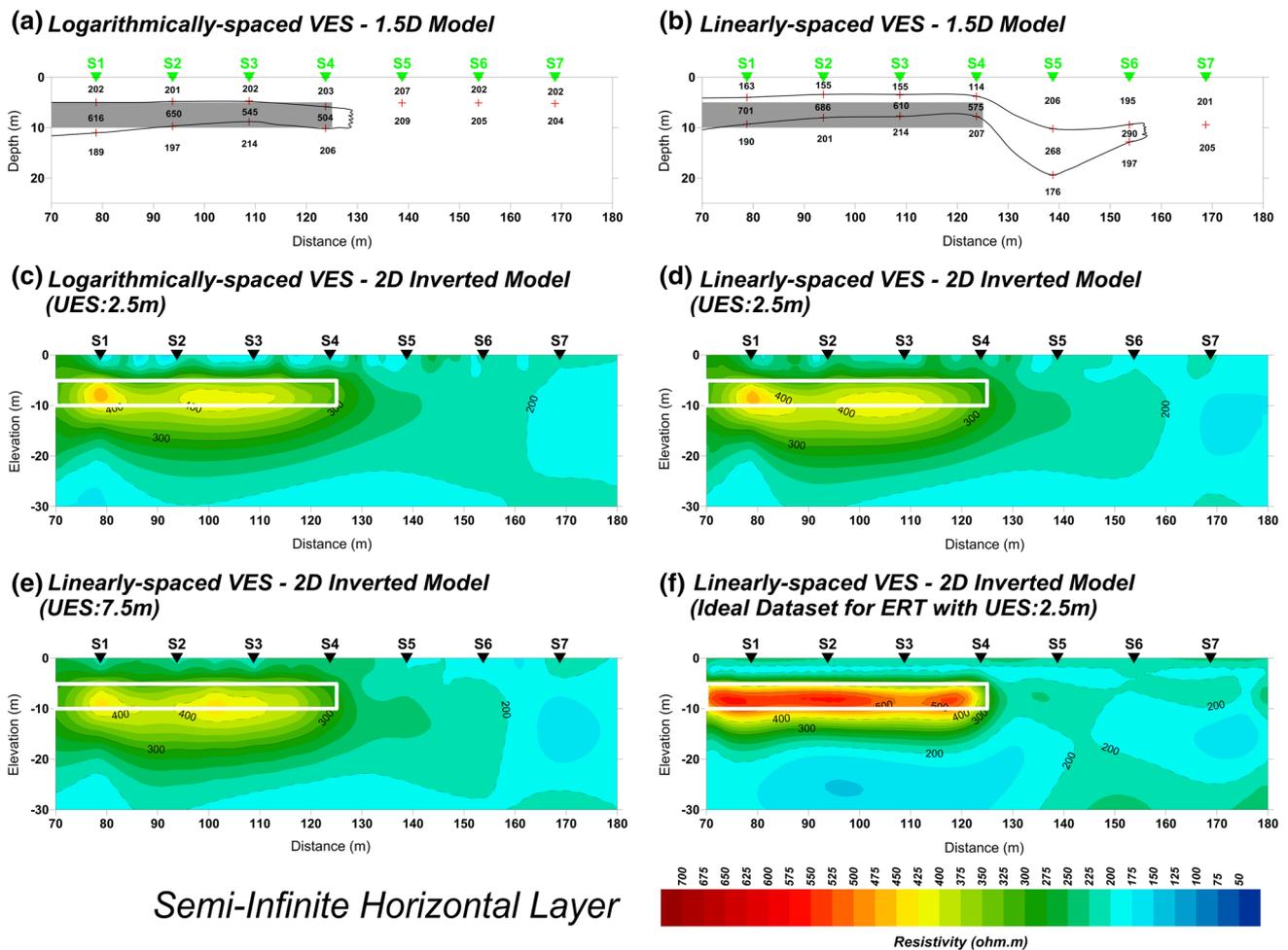
1.5D model since (i) dike represents a great violation of 1D assumption (vertical instead of horizontal interface) and VES technique is not usually applied in the environments having these type of structures and (ii) while VES stations are usually spaced tens or hundreds of meters apart, dike structures are usually small and improbable to be detected in these surveys. On the other hand, in the case of bigger dikes and/or small-scale VES surveys, the location of VES stations with respect to the dike center and the number of VES stations that have been affected by the dike are among the key factors in the detectability of the dike. Consequently, they are not distinguished in final 1.5D models most of the time.

Semi-infinite horizontal layer

The apparent resistivity curves of different VES stations on the semi-infinite horizontal layer are presented in Fig. 5. In this case, a transitional state between two-layer and three-layer models can be seen in both the logarithmically spaced

Fig. 5 Apparent resistivity curves of the logarithmically spaced and linearly spaced VES measurements for the “semi-infinite horizontal layer” model





Semi-Infinite Horizontal Layer

Fig. 6 The results of 1D and 2D inversion algorithms for the “semi-infinite horizontal layer”

Table 4 The rms error of different VES stations in the 1D inversion approach for the “semi-infinite horizontal layer” model

	VES station	S1	S2	S3	S4	S5	S6	S7
rms error (%)	Logarithmically spaced VES	1.26	0.97	0.62	0.63	3.93	1.35	0.84
rms error (%)	Linearly spaced VES	1.35	1.47	0.86	0.4	0.94	0.96	0.86

Table 5 The rms error of the represented 2D inverted models of the “semi-infinite horizontal layer”

Type of VES	Logarithmically spaced	Linearly spaced	Ideal dataset	
Unit electrode spacing (UES) in meters	2.5	2.5 7.5	2.5	
rms error (%)	0.50	0.65 0.74	0.32	

and linearly spaced VES scenarios (it is more clearly visible in the logarithmically spaced curves). The results of 1D and 2D inversions are also shown in Fig. 6. Table 4 lists the rms

errors of 1D inversion for linearly spaced and logarithmically spaced VES datasets. The rms error of 2D inversions is provided in Table 5.

According to Fig. 6a, b, the results of logarithmically spaced VES dataset are better than linearly spaced VES measurements in the presence of this structure. Although the logarithmically spaced 1.5D model is a bit inaccurate, the result does not suffer from artifacts. The linearly spaced 1.5 model is not realistic adjacent to the layer’s edge. The edge of the layer is not correctly characterized by linearly spaced VES dataset. In addition to this artifact, the depth of the second layer is not estimated as exact as the logarithmically spaced 1.5D model. On the other hand,

the estimated resistivity values for the linearly spaced 1.5D model are closer to the true resistivity values. Note that the estimated thicknesses are not defined correctly in both measurement scenarios. If linearly spaced VES is utilized, the appearance of the artifact will be an undistinguishable phenomenon in the 1.5D model. Hence, it seems that logarithmically spaced VES method can handle the finite horizontal layers and their equivalent edge effects on the VES curves will not affect the final 1.5D model. So, these kinds of structures (i.e., interlayers in the sedimentary basins) are not really problematic for logarithmically spaced VES surveys. On the other hand, if linearly spaced VES method is going to be implemented, the possible distortions must be taken into account.

The 2D inversion result for logarithmically spaced VES measurement scenario is shown in Fig. 6c. The results of linearly spaced VES measurements are also presented in Fig. 6d (UES = 2.5 m) and e (UES = 7.5 m). Although linearly spaced and logarithmically spaced 2D models (Fig. 6c–e) are not really different from each other, the linearly spaced VES dataset with UES = 7.5 m is a bit better. The logarithmically spaced and linearly spaced VES with UES = 2.5 m have some surface high or low anomalous point features. These distortions do not seem to be problematic, i.e., preventing real subsurface structures to be imaged or making the interpretations challenging. The resistivity values in the 2D models of VES datasets are all underestimated. The background resistivity value (200 Ω m) is estimated very well in all of the 2D models. The ideal dataset for an ERT survey is prepared with the original UES = 2.5 m, and its inverted model is shown in Fig. 6f. Clearly, this 2D model is better than the previously mentioned 2D models with no surface artifacts. Furthermore, the resistivity of the semi-infinite layer is more accurately estimated. Note that it is still lower than the real resistivity values of the synthetic model. Similar reasons for the accuracy and precision of this 2D model can be expressed, i.e., the higher resolution of the dataset. On the whole, all of the 2D models do not have remarkable artifacts. The same underestimation of the buried resistive feature is also notable.

To add up the consequences related to 1D inversion of VES datasets, it should be concluded that several problems and artifacts could result in the presence of similar violating features. These types of structures are commonly found in some geological settings (i.e., sedimentary basins, some type of faults, etc.), so care must be taken seriously. Consequently, it is better to use logarithmically spaced VES rather than linearly spaced VES technique if 1D inversion is going to be performed. If linearly spaced VES method is used, possible errors about the edge's location must be kept in mind. Furthermore, it should be noted that the depth/thickness estimations in the 1.5D model will not be

as accurate as the normal situation (ideal setting without any 1D assumption violation factors).

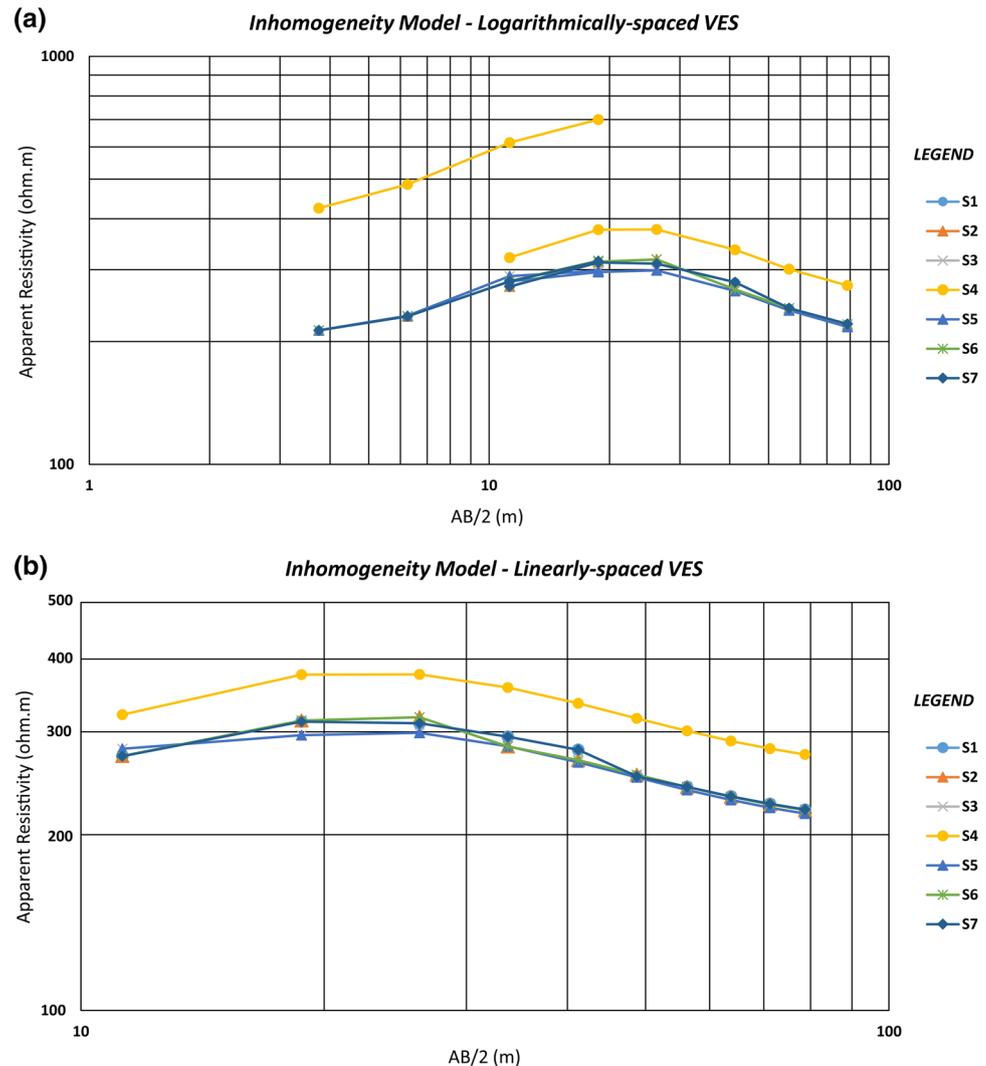
Inhomogeneity in the layer

The apparent resistivity curves of different VES stations on the “inhomogeneity in the layer” model are presented in Fig. 7. An important point is the considerable shift of the curve for S4 (which is over the inhomogeneity). This shift can tell us the great effect of lateral resistivity variations just below the VES station, so the effect of the inhomogeneity diminishes by considering a greater MN for the inversion process. Furthermore, an overall increase in apparent resistivity values can be seen in both logarithmically spaced and linearly spaced VES curves over station S4. In other cases (other VES stations), the effect of the inhomogeneity is a small deviation from the general trend of the three-layer VES curve. The results of 1D and 2D inversions are shown in Fig. 8. Table 6 lists the rms errors of 1D inversion for linearly spaced and logarithmically spaced VES datasets. The rms error of 2D inversions is provided in Table 7.

In Fig. 8a, b, 1.5D models of logarithmically spaced and linearly spaced VES datasets are presented. Although the depth of the third layer (also the thickness of the second layer) is not accurately estimated in the logarithmically spaced 1.5D model, this 1.5D model seems to be more reliable and more precise than the linearly spaced 1.5D model. The logarithmically spaced 1.5D model is relatively precise in determining the depth of the second layer. It is also more accurate in resistivity estimations with lower deviation from the real resistivity values.

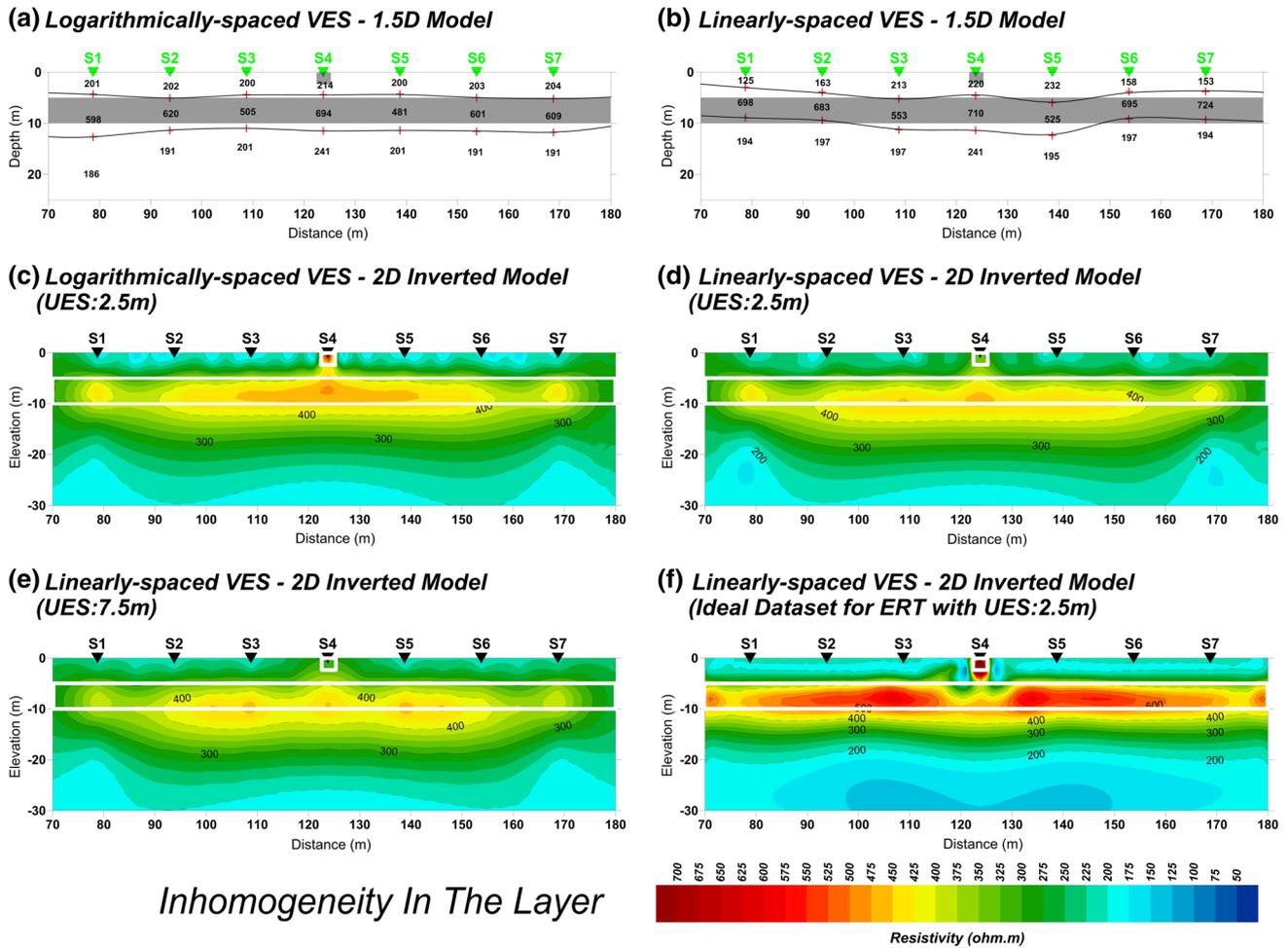
If the linearly spaced VES dataset involved $MN < 7.5$ m, it might detect an extra unrealistic layer just adjacent to the high resistivity inhomogeneity, i.e., at the VES station “S4” (see Fig. 9 that represents a complete dataset of linearly spaced VES data drawn with MN ranging from 0.5 to 12.5 m and AB ranging from 1.5 to 200 m.). When the inhomogeneity lies between the potential electrodes, it will be located in the highest positive zone of the sensitivity section. This point means that the presence of resistivity inhomogeneity will further enhance the apparent resistivity readings especially those with smaller current electrode separations. Consequently, the surface readings responsible for surface layers will be affected and the 1D inverse modeling will distinguish an “extra layer” in the VES curve. Figure 9 can clearly illustrate the appearance of “extra layer” artifact in 1D inversion results where the VES curve will change dramatically with respect to the normal situation (simple three-layer model without inhomogeneity). Note that the general trend of the curve changes when the inhomogeneity lies outside the potential electrodes without representing a new layer into the model. In this case, the

Fig. 7 Apparent resistivity curves of the logarithmically spaced and linearly spaced VES measurements for the “inhomogeneity in the layer” model



sensitivity (its sign and/or value) of different readings will not be the same in the location of inhomogeneity and the effects diminish by getting away from the potential electrodes. In this case, the affected measurements make distortion in the VES curve. So, the datapoints do not follow the general trend of the VES curve and usually look like an outlier in the case of extreme resistivity inhomogeneities. If these points appear on VES curves, they are often treated as noisy data. However, according to synthetic modeling results, these types of points are related to the resistivity inhomogeneities (better to be called as “2D/3D effect”). Interestingly, the same interpretational point is expressed by some researchers like Bentley and Gharibi (2004) where they introduce these outliers as “3D effects” in the sounding curves. They called it “3D effect” rather than “2D effect” because these points have increased the inaccuracy of 2D inversion results.

The 2D inversion result for logarithmically spaced VES measurement scenario is shown in Fig. 8c. The results of linearly spaced VES measurements are presented in Fig. 8d (UES = 2.5 m) and e (UES = 7.5 m). According to the 2D models, the inhomogeneity is characterized in the logarithmically spaced 2D model but it did not appear in the linearly spaced models because the resolution of the linearly spaced dataset cannot image this small inhomogeneity. In fact, the surface readings of the logarithmically spaced dataset are strongly affected by the inhomogeneity. If these surface measurements are removed from the dataset, the inhomogeneity will not be imaged by the 2D inverse modeling approach. In the linearly spaced 2D models, the inhomogeneity affected the accuracy and precision of the model both in terms of estimated resistivity and thickness of subsurface features. The result of the ideal dataset for ERT is presented in Fig. 8f. This model is better



Inhomogeneity In The Layer

Fig. 8 The results of 1D and 2D inversion algorithms for the “inhomogeneity in the layer”

Table 6 The rms error of different VES stations in the 1D inversion approach for the “inhomogeneity in the layer” model

	VES station	S1	S2	S3	S4	S5	S6	S7
rms error (%)	Logarithmically spaced VES	1.16	1.3	1.67	0.75	1.64	1.32	0.85
rms error (%)	Linearly spaced VES	1.01	1.41	0.99	0.7	0.78	1.4	1.02

Table 7 The rms error of the represented 2D inverted models of the “inhomogeneity in the layer”

Type of VES	Logarithmically spaced	Linearly spaced	7.5	Ideal dataset
Unit electrode spacing (UES) in meters	2.5	2.5	7.5	2.5
rms error (%)	0.60	0.89	0.99	1.75

than other 2D models provided by logarithmically spaced and linearly spaced datasets. Although the 2D inverted models are not suffering from remarkable artifacts, the inaccuracies can be severe in the case of high resistivity contrasts and also in the detection of the bottom of anomalous features.

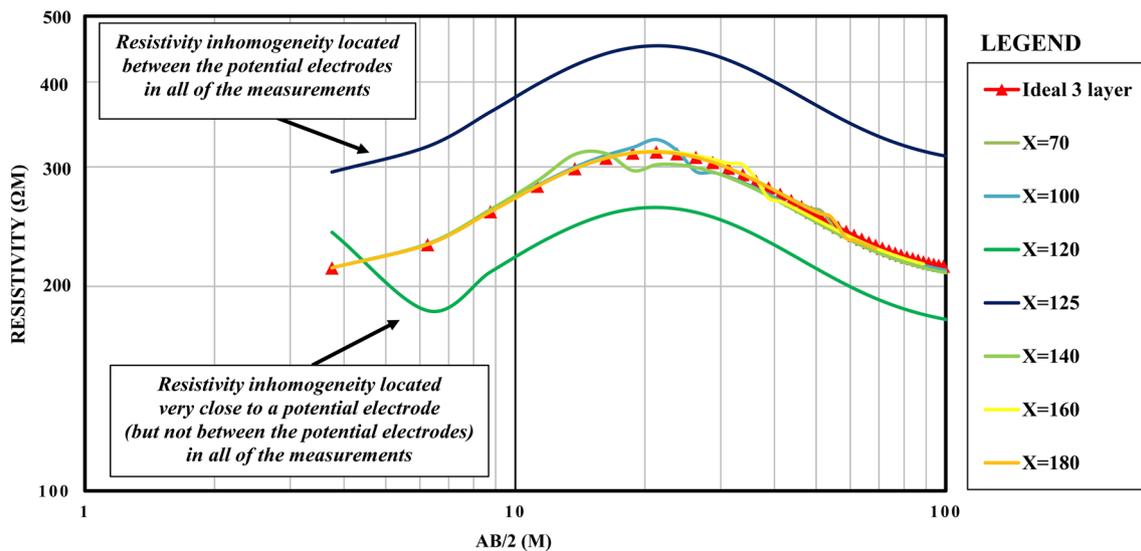


Fig. 9 A more accurate apparent resistivity curve for VES stations on the “inhomogeneity in the layer” model with much more datapoints with MN ranging from 0.5 to 12.5 m and AB ranging from 1.5 to 200 m. The effect of the inhomogeneity can be very different

according to its location with respect to the potential electrodes (in the Schlumberger array) and all of the effects can be discussed based on the sensitivity concept

Some real examples

In this section, several case studies are presented that were subjected to both 1D and 2D inversion routines. Two examples represent the situation where 1D violating structures are present. The other case study shows the results when dealing with a relatively ideal situation for VES survey, without problematic 1D violating features. Here, case studies are briefly discussed and more detailed information about geology, hydrology, etc., can be found in the mentioned references.

Bitumen exploration at Gilan-gharb

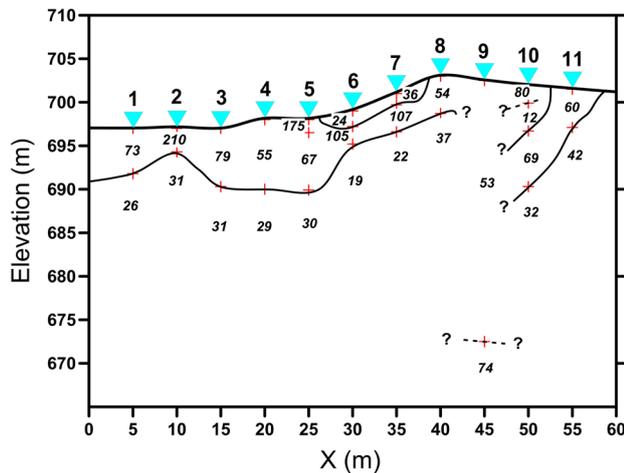
This example represents a profile from an exploration study at Gilan-gharb city, Kermanshah province, Iran. Basically, it was a 2D IP and resistivity survey to detect bitumen mineralization zones in a hydrogeophysically complicated area. Combined resistivity sounding and profiling (CRSP) array was used to do the investigations. Further information about this exploration study can be found at Mashhadi et al. (2017). As the CRSP array is used, a VES dataset will be available that is composed of linearly spaced measurements. A part of profile 1 at Gilan-gharb is subjected to 1D and 2D inversion methods (Fig. 10). In the 1.5D model, it is clearly visible that a reasonable connection between the detected interfaces is impossible. This ambiguity is the result of a vertical resistivity interface (bituminous zone), just like the effects observed in the synthetic models. Note that the 1.5D and 2D models follow a relatively similar pattern in other

parts of the profile where violating structures have a limited effect on the measurements. It is also notable that the resistivity of the groundwater saturated zone is estimated similarly in both 1D and 2D inversion routines (it is less than 30 Ω m due to high salinity).

Skarn copper exploration at Malayer

Skarn copper exploration was done by means of time-domain IP and resistivity tomography in Ghale-Alimoradkhan area, Malayer city, Hamedan province, Iran. It was measured by using the CRSP array, so it could be viewed as a VES dataset that is composed of linearly spaced measurements. This study aimed to delineate skarn alteration haloes within the Ghale-Alimoradkhan prospect (Mashhadi and Ramazi 2018). One of the profiles is selected to perform 1D and 2D inversions (Fig. 11). In Fig. 11, it could be seen that the surface horizontal structures (agricultural lands) could be characterized by 1.5D model (from $X = 0$ m to $X = 45$ m and also from $X = 185$ m to the end of the profile). Also, the surface horizontal feature in the middle of the profile is indicative of relatively sulfide-barren skarn units or altered limy units. However, as the structures tend to become more complex especially with the appearance of vertical or steeply dipping resistivity interfaces, the credibility of 1.5D model becomes questionable. Furthermore, a clear relationship between the 1D models cannot be proposed. This phenomenon also proves the synthetic modeling results. Note that the estimated resistivities for surface horizontal layers are in great agreement with both 1D and 2D inversion results.

Resistivity 1D Inversion Result (1.5D Model)



Resistivity 2D Inversion Result (2D Model)

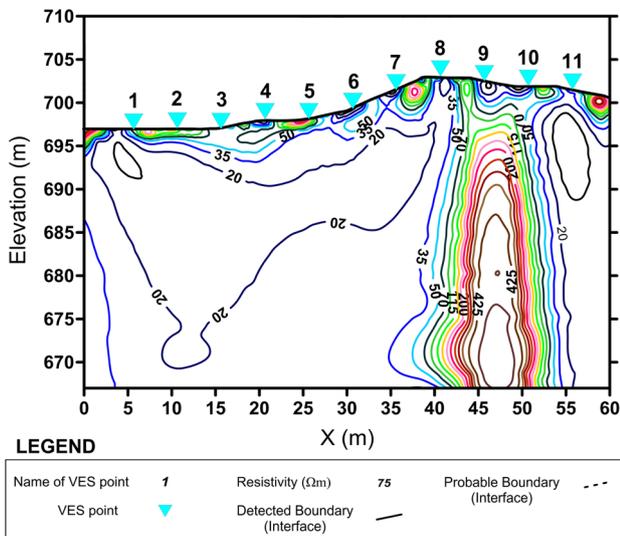


Fig. 10 The 1.5D and 2D resistivity models over a bituminous body near Gilan-gharb city, Iran

Groundwater investigation at Khomein

This investigation was a routine VES implementation for groundwater assessment in Khugan village, Khomein city, Markazi Province, Iran. It is a logarithmically spaced VES dataset. Maximum current electrode spacing and the maximum potential electrode spacing were 90 m and 5 m, respectively. The results of 1D and 2D inversion algorithms are provided in Fig. 12. Interestingly, the main features are relatively similar in terms of shape. However, the thickness and resistivity parameters are different in 1.5D and 2D models. Although the 2D model provided a better image of the subsurface with higher accuracy, the result of 1.5D model is also reliable. From the viewpoint of survey cost and also the required accuracy of the subsurface model, implementing

the VES technique to provide a 1.5D model is preferred in these situations.

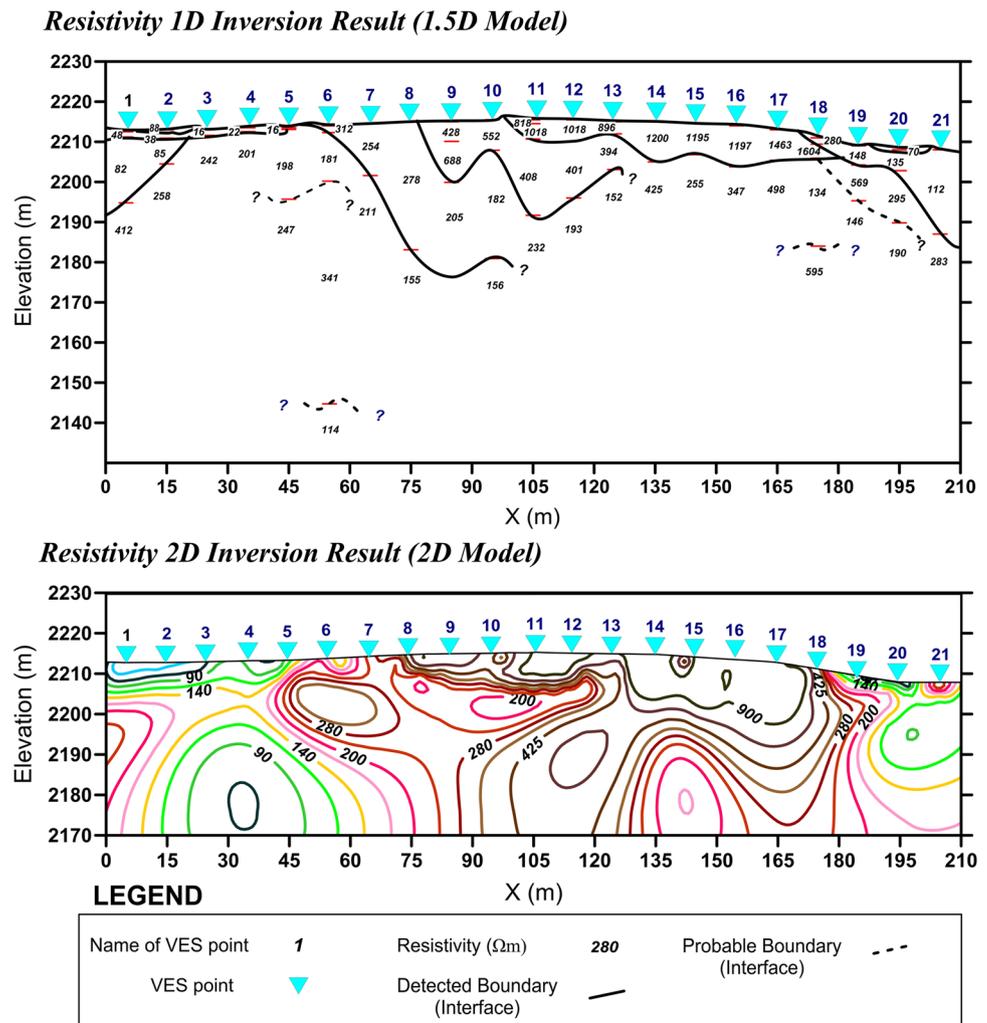
Discussion

Previous studies about 2D/3D effects on VES data

As the 1D resistivity technique does not account for lateral changes in subsurface resistivity, one-dimensional inverse modeling can lead to errors in the interpretation of resistivity and thickness of layers (Loke 2019). Furthermore, high noise levels can have dramatic effects on the interpretation of VES data. Başokur (1999) discussed the sources of noise and error in the framework of VES data while trying to introduce an automated 1D interpretation method. As he discussed, “random noises” can be defined as fluctuations in resistivity readings that can be highlighted by repeated measurements (providing a measure of data instability). However, “systematic noises” are related to some faulty actions and/or problems related to data measurement system, field crew mistakes, etc., that usually affect a series of measurements rather than a few readings. Başokur (1999) has represented a synthetic model with a small 2D body located on the surface of a horizontally layered model. For each VES station, different phenomena occurred according to the location of VES station (with respect to 2D body) and the existing resistivity contrast (positive or negative). Some of the effects discussed in our research could be found in Başokur (1999), i.e., the appearance of “outlier” or “V-shape fluctuation” in the apparent resistivity curve whenever the VES station is far from a surface resistivity inhomogeneity. If there are no measures of data stability, these mentioned fluctuations could also be misleadingly taken as “random noise” (Başokur 1999) even if the true reason is the effect of 2D/3D irregularities (Bentley and Gharibi 2004) and vice versa. Here, we have expressed the reason for resistivity variations based on the sensitivity concept that can improve our understanding in these situations to highlight whether the data is affected by noise or by a 2D/3D violating feature. This seems easier when there are several VES stations in a profile as you can plot the 1.5D model to correct the possible inaccuracies and mistakes in your data analysis. The same concept is believed by some other researchers who have proposed to invert one-dimensional VES data using a series of VES stations concurrently in a joint inversion technique, i.e., 1.5D inversion method (Gyulai and Ormos 1999).

Beard and Morgan (1991) have tested the use of 1D inversion of VES data over some 2D structures for Schlumberger and Wenner arrays. They have created some 2D models and calculated their equivalent 1.5D model to see the possible inaccuracies and distortions (some of his models are pretty similar to the ones represented in this paper). Interestingly,

Fig. 11 The 1.5D and 2D resistivity models over a skarn copper deposit near Malayer city, Iran



the same negative effects and distortions were observed in the 1.5D models. To overcome these distortions, Beard and Morgan (1991) introduced a new methodology for VES data analysis in which 1D inverted model parameters are interpolated to create a “cross section of resistivity”. Note that this “cross section” is different from the well-known “pseudo-section plot” because the former uses the inverted resistivity values while the latter gets apparent resistivity values for the interpolation. Beard and Morgan (1991) concluded that contours based on inverted layers (i.e., the introduced cross section) are advantageous with respect to 1.5D models since the interfaces are more accurately determined and a good resistivity cutoff would be helpful to predict the geometry of the anomalous target. On the other hand, the negative effects, distortions, artifacts, etc., discussed here in this paper (also observed in Beard and Morgan (1991)) would directly affect their introduced methodology of VES data interpretation.

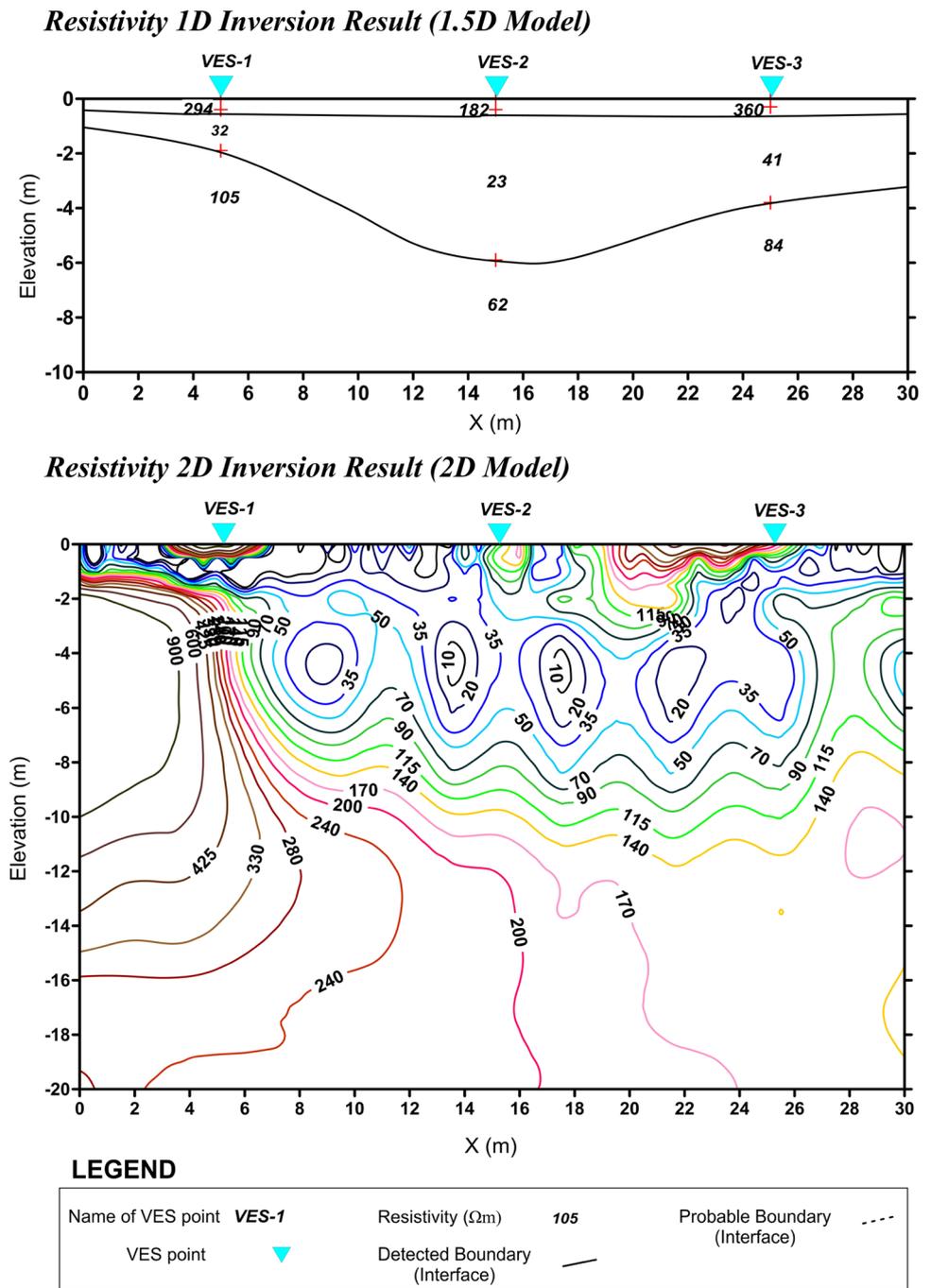
Geological settings with 1D violating features

The assessment of synthetic models showed that 1D assumption violations can have noticeable effects on the accuracy and precision of 1D inverted models resulted from VES data analysis. This paper has taken a more practical overview of the problem by discriminating between logarithmically and linearly spaced VES datasets and by considering a limited number of datapoints rather than the perfect datasets consist of tens of measurements for each VES station. We would like to focus that these simple considerations have greatly improved our results since the real case studies can prove the effects observed in the analysis of synthetic models.

Here, some examples of the common 1D violating features in the geological environments are listed:

1. Vertical/sub-vertical movement of geological units: it can happen due to faulting, subsidence, sliding and so on.

Fig. 12 The 1.5D and 2D resistivity models in a groundwater assessment survey at Khugan village, Khomein city, Iran



2. The crushed zone of faults: it usually happens for the major faults in the region. The crushed zone is usually characterized by a decrease in resistivity (this case is a vertical/sub-vertical resistivity interface in the subsurface).
3. Sedimentary interlayers: it will not be problematic if the surveying location is far from the edge of the interlayers.
4. Karstic features: they can create undulating, vertical, sub-vertical and even horizontal resistivity interfaces since karstic features can be in any shape.
5. Magmatic dikes and sills: this can be a difficult problem in some areas especially when there is no information about the subsurface.
6. Undulating bedrock/interfaces with extreme variations over short distances.

7. Salt domes.

The mentioned geological features exist within the environments that most of the hydrogeophysical VES surveys are implemented. As a result, it is good to remind the geophysicists that these situations are not uncommon at all. So, by gathering the available information about the study area, we can understand the different 1D violating features that are expected to have in the subsurface, and, to have an eye for the possible negative effects on the VES measurements. It is important to note that the inverted models (1D, 1.5D and 2D models) can sometimes help us to identify the undesirable features pretty easily. Consequently, it is worthy to access the models more precisely, searching for the possible negative effects or modeling artifacts before proposing the final interpretations.

Conclusions

This research is performed to examine the possible effects of 1D assumption violations for linearly spaced and logarithmically spaced VES measurements (1D electrical survey). To do the assessment more effectively, some synthetic resistivity models are provided and processed by 1D and 2D inversion algorithms. Furthermore, several real field case studies are also examined and some of them presented in the paper to support the modeling results.

After synthetic modeling, several distortions and inversion artifacts are characterized. It is observed that linearly spaced VES measurements are not really suitable for 1D inversion in the case of 1D assumption violations and logarithmically spaced VES can better handle these problematic features. For semi-infinite horizontal layers and also small surface resistivity inhomogeneities, logarithmically spaced VES datasets mostly provide acceptable 1.5D model while linearly spaced VES datasets suffer from remarkable artifacts. In the case of vertical structures, both linearly spaced and logarithmically spaced VES techniques suffer from remarkable artifacts. If these kinds of violating features exist, artifacts in the form of “extra layer” may appear in those VES stations that are adjacent to the undesirable vertical structure. However, for the VES stations located on the vertical structure, no extra layer appears in the 1D inversion result. Some discussions about the 2D inversion of VES datasets are also provided. As it was not the aim of this paper to assess the accuracy and precession of 2D inversed models of VES datasets, it is not discussed in detail. In the end, it must be emphasized that 1D violating features are not improbable in many geological situations, so they must be considered in mind when processing and interpreting the geophysical models.

References

- Asfahani J (2016) Hydraulic parameters estimation by using an approach based on vertical electrical soundings (VES) in the semi-arid Khanasser valley region, Syria. *J Afr Earth Sci* 117:196–206. <https://doi.org/10.1016/j.jafrearsci.2016.01.018>
- Barker RD (1989) Depth of investigation of collinear symmetrical four-electrode arrays. *Geophysics* 54(8):1031–1037
- Başokur AT (1999) Automated 1D interpretation of resistivity soundings by simultaneous use of the direct and iterative methods. *Geophys Prospect* 47:149–177. <https://doi.org/10.1046/j.1365-2478.1999.00123.x>
- Beard LP, Morgan FD (1991) Assessment of 2-D resistivity structures using 1-D inversion. *Geophysics* 56(6):874–883. <https://doi.org/10.1190/1.1443106>
- Bentley LR, Gharibi M (2004) Two- and three-dimensional electrical resistivity imaging at a heterogeneous remediation site. *Geophysics* 69(3):674–680. <https://doi.org/10.1190/1.1759453>
- Chambers JE, Ogilvy RD, Kuras O, Cripps JC, Meldrum PI (2001) 3D electrical imaging of known targets at a controlled environmental test site. *Environ Geol* 41:690–704. <https://doi.org/10.1007/s00254-001-0452-4>
- Dahlin T, Bernstone C, Loke MH (2002) A 3-D resistivity investigation of a contaminated site at Lernacken, Sweden. *Geophysics* 67(6):1692–1700. <https://doi.org/10.1190/1.1527070>
- Danielsen BE, Dahlin T (2009) Comparison of geoelectrical imaging and tunnel documentation at the Hallandsås Tunnel, Sweden. *Eng Geol* 107:118–129. <https://doi.org/10.1016/j.enggeo.2009.05.005>
- Greenhalgh S, Wiese T, Marescot L (2010) Comparison of DC sensitivity patterns for anisotropic and isotropic media. *J Appl Geophys* 70:103–112. <https://doi.org/10.1016/j.jappgeo.2009.10.003>
- Gupta PK, Nirwas S, Gaur VK (1997) Straightforward inversion of vertical electrical sounding data. *Geophysics* 62(3):775–785. <https://doi.org/10.1190/1.1444187>
- Gyulai Á, Ormos T (1999) A new procedure for the interpretation of VES data: 1.5-D simultaneous inversion method. *J Appl Geophys* 41:1–17
- Gyulai A, Ormos T, Dobroka M (2010) A quick 2-D geoelectric inversion method using series expansion. *J Appl Geophys*. <https://doi.org/10.1016/j.jappgeo.2010.09.006>
- Hauck C, Kneisel C (2008) *Applied geophysics in periglacial environments*. Cambridge University Press, Cambridge
- Hodlur GK, Dhakate R, Andrade R (2006) Correlation of vertical electrical sounding and borehole-log data for delineation of saltwater and freshwater aquifers. *Geophysics* 71(1):G11–G20. <https://doi.org/10.1190/1.2169847>
- Kumar D, Ahmed S, Krishnamurthy NS, Dewandel B (2007) Reducing ambiguities in vertical electrical sounding interpretations: a geostatistical application. *J Appl Geophys* 62:16–32. <https://doi.org/10.1016/j.jappgeo.2006.07.001>
- Loke MH Rapid 2D resistivity forward modelling using the finite-difference and finite-element methods (RES2DMODE Manual). Geotomo Software, Malaysia (www.geoelectrical.com)
- Loke MH (2019) Tutorial: 2-D and 3-D Electrical Imaging Surveys. Geotomo Software, Malaysia (<https://www.geotomosoft.com/coursenotes.zip>)
- Loke MH, Chambers JE, Rucker DF, Kuras O, Wilkinson PB (2013) Recent developments in the direct-current geoelectrical imaging method. *J Appl Geophys* 95:135–156. <https://doi.org/10.1016/j.jappgeo.2013.02.017>
- Mashhadi SR, Ramazi H (2018) The application of resistivity and induced polarization methods in identification of skarn alteration

- haloes: a case study in Qale-alimoradkhan area. *J Environ Eng Geophys* 23(3):363–368. <https://doi.org/10.2113/jeeq23.3.363>
- Mashhadi SR, Mostafaei K, Ramazi H (2017) Improving bitumen detection in resistivity surveys by using induced polarization data. *Explor Geophys*. <https://doi.org/10.1071/EG17032>
- Mota R, Santos FAM, Mateus A, Marques FO, Gonçalves MA, Figueiras J, Amaral H (2004) Granite fracturing and incipient pollution beneath a recent landfill facility as detected by geoelectrical surveys. *J Appl Geophys* 57:11–22. <https://doi.org/10.1016/j.jappgeo.2004.08.007>
- Pedromo S, Ainchil JE, Kruse E (2014) Hydraulic parameters estimation from well logging resistivity and geoelectrical measurements. *J Appl Geophys* 105:50–58. <https://doi.org/10.1016/j.jappgeo.2014.02.020>
- Porsani JL, Filho WM, Elis VR, Shimeles F, Dourado JC, Moura HP (2004) The use of GPR and VES in delineating a contamination plume in a landfill site: a case study in SE Brazil. *J Appl Geophys* 55:199–209. <https://doi.org/10.1016/j.jappgeo.2003.11.001>
- Reynolds JM (2011) An introduction to applied and environmental geophysics, 2nd edn. Wiley, London
- Rucker DF, Glaser DR (2015) Standard, random, and optimum array conversions from two-pole resistance data. *J Environ Eng Geophys* 20:207–217
- Sajinkumar KS, Castedo R, Sundarajan P, Rani VR (2015) Study of a partially failed landslide and delineation of piping phenomena by vertical electrical sounding (VES) in the Wayanad Plateau, Kerala, India. *Nat Hazards* 75:755–778. <https://doi.org/10.1007/s11069-014-1342-x>
- Sattar GS, Keramat M, Shahid S (2014) Deciphering transmissivity and hydraulic conductivity of the aquifer by vertical electrical sounding (VES) experiments in Northwest Bangladesh. *Appl Water Sci*. <https://doi.org/10.1007/s13201-014-0203-9>
- Singh UK, Tiwari RK, Singh SB (2005) One-dimensional inversion of geo-electrical resistivity sounding data using artificial neural networks—a case study. *Comput Geosci* 31:99–108. <https://doi.org/10.1016/j.cageo.2004.09.014>
- Song SH, Lee JY, Park N (2007) Use of vertical electrical soundings to delineate seawater intrusion in a coastal area of Byunsan, Korea. *Environ Geol* 52:1207–1219. <https://doi.org/10.1007/s00254-006-0559-8>
- Sundararajan N, Sankaran S, Al-Hosni TK (2012) Vertical electrical sounding (VES) and multi-electrode resistivity in environmental impact assessment studies over some selected lakes: a case study. *Environ Earth Sci* 65:881–895. <https://doi.org/10.1007/s12665-011-1132-7>
- Tizro AT, Voudouris K, Basami Y (2012) Estimation of porosity and specific yield by application of geoelectrical method—a case study in western Iran. *J Hydrol* 454–455:160–172. <https://doi.org/10.1016/j.jhydrol.2012.06.009>
- Topolewska S, Stêpień M, Kowalczyk S (2016) Mapping of the north-eastern part of Kozłowicka buried valley based on geoelectrical data. *Stud Quat* 33:91–101. <https://doi.org/10.1515/squa-2016-0009>
- Veeraiah B, Babu GA (2014) Deep resistivity sounding (DRS) technique for mapping of sub-trappean sediments—a case study from central India. *J Appl Geophys* 105:112–119. <https://doi.org/10.1016/j.jappgeo.2014.03.012>
- Zohdy AAR (1989) A new method for the automatic interpretation of Schlumberger and Wenner sounding curves. *Geophysics* 54(2):245–253. <https://doi.org/10.1190/1.1442648>



Novel wide-angle AVO attributes using rational function

Yaru Xue^{1,2} · Jialin Xiang^{1,2} · Xin Xu^{1,2}

Received: 3 July 2019 / Accepted: 28 November 2019 / Published online: 9 December 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

Conventional AVO inversion employs Zoeppritz equations and various approximations to them to obtain the reflection coefficients of plane-waves, which are confined to a certain (small) angle range (mostly below 40°). However, near the critical angles (wide-angle), reflections at the post-critical angles provide much more potential for velocity and density inversion because of the large amplitudes and phases-shifted waveforms, while the Zoeppritz equations are not applicable anymore. Hence, there is a strong demand for the research into wide-angle AVO. With reflection coefficients at wide-angle corresponding to the features of rational function, we try to approximate the seismic data with vector fitting which is used to obtain the rational zero-pole and residual properties of wide-angle AVO. We apply this technique to classify AVO type and recognize the lithology. Our experiment shows that extending our research into wide-angle AVO is very promising in gathering richer data for a more accurate seismic analysis.

Keywords Seismic exploration · Lithological distinction · Wide-angle AVO · Rational function fitting

Introduction

In seismic exploration, AVO (Amplitude Variation with Offset) is a technology which studies the lithology and detects hydrocarbon using amplitude information. The AVO inversion is based on the AVO characteristics to estimate the elastic parameters of rock and deduce the lithology of medium according to seismic data. Before inversion, the critical and post-critical reflections are traditionally muted, because NMO correction will lead to the stretching of remote offset and to avoid the complexity of interferences of reflected and head waves (Krail and Brysk 1983; Winterstein 1985). However, with further research, it is found that the anomalous amplitude and phase anomalies near the critical angles

are favorable for inversion. AVO studies have shown that large-offset information is needed to extract density information. (Debski and Tarantola 1995; Downton and Ursenbach, 2006).

However, in AVO analysis, the traditional method to approximate the AVO response is by using linear formulas including the triangular function of reflection angle (Bortfeld 1961; Aki and Richards 1980; Shuey 1985). But it is only applicable at the small incidence angles, because there is a strong impedance difference near the critical angles, the Zoeppritz approximation is not applicable (O'Brien 1963; Macdonald et al. 1987). Therefore, how to characterize wide-angle AVO is an important and meaningful issue.

The introduction of long recording cables and new acquisition methods, such as submarine nodes, makes it possible to record large-offset reflections. The critical angle is easy to reach in structures with large velocity ratios, such as salts, volcanic rocks and carbonates. Therefore, the seismic reflections at the angle above the critical angle are becoming more and more common (Zhu and Mcmechan 2015).

In general, when seismic exploration is carried out, the actual acquired seismic data are excited by the point source, which produces spherical wave (Červený and Hron 1961; Haase 2004; Ayzenberg et al. 2009; Ursenbach et al. 1949). And so far, a lot of studies have been devoted to the reflection of spherical waves on a planar interface in two-layer

✉ Yaru Xue
xueyaru@cup.edu.cn

✉ Jialin Xiang
xiangjialin12@163.com

✉ Xin Xu
XuXinIsJiangYeYe@outlook.com

¹ College of Information Science and Engineering, China University of Petroleum (Beijing), Beijing 102249, China

² State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, China

homogeneous half-space. We can find that the spherical-wave reflection coefficients are related to rational function.

Studies have found that when the incident angle is larger than the critical angle, the offset will be larger and the signal-to-noise ratio of the seismic data will be higher. In the case of ignoring the loss of seismic wave energy, the energy of the wide-angle seismic reflection wave is much greater than that of the non-wide-angle reflection wave, which is a very favorable condition for the processing of seismic data. Development of wide-angle seismic exploration is very necessary.

In this paper, we use rational function to fit seismic data and characterize AVO attributes by zero-pole and residual properties. The main body of this paper is divided into two parts., fitting and applying. In the fitting part, we use rational function to fit seismic coefficient curves whose form will be elaborated in the theory, and to obtain zero-pole and residual properties and discuss the feasibility of these attributes; in the applying part, we apply this method to classify AVO, recognize hydrocarbon signatures, and compare with the traditional polynomial fitting technique. This new method is used to describe the characteristics of seismic waves more accurately, and to explore more useful information from seismic record.

Theory

AVO describes the amplitudes variation with offset, while PVO means the phases variation with offset; these two attributes can be formulated as an complex function with variable θ (the incident angle):

$$H(\theta) = |H(\theta)|e^{i\phi(\theta)} \tag{1}$$

where $|H(\theta)|$ represents AVO and $\phi(\theta)$ represents PVO; this formula is very similar to the frequency response of a circuit system. Formula (1) can be expanded as summation of rational function:

$$H(s) = \sum_{k=1}^n \frac{C_k}{s - A_k} + D \tag{2}$$

where $s = j2\pi r$, A_k denotes the pole, C_k denotes the corresponding residual for each pole, and r is defined as offset, D is the delay factor that controls the amount of delay to fit the data.

By summarizing the previous research, the spherical-wave reflection coefficient can be expressed as follows for a homogeneous half-space medium model (Li et al. 2016):

$$R_{pp}^{sph} = \frac{\int_1^0 B(x)J_0(\omega r\sqrt{1-x^2}/v_1)e^{i\alpha x(h+z)/v_1}dx + i\int_0^\infty B(x)J_0(\omega r\sqrt{1+x^2}/v_1)e^{-\alpha x(h+z)/v_1}dx}{\int_1^0 J_0(\omega r\sqrt{1-x^2}/v_1)e^{i\alpha x(h+z)/v_1}dx + i\int_0^\infty J_0(\omega r\sqrt{1+x^2}/v_1)e^{-\alpha x(h+z)/v_1}dx} \tag{3}$$

where B is the plane PP-wave reflection:

$$B(x) = \frac{\rho_2 v_2 x - \rho_1 v_1 \sqrt{1 - v_2^2/v_1^2(1 - x^2)}}{\rho_2 v_2 x + \rho_1 v_1 \sqrt{1 - v_2^2/v_1^2(1 - x^2)}} \tag{4}$$

v_1, v_2, ρ_1, ρ_2 denotes the P-wave velocity and density of the upper and lower medium, respectively, J_0 is the zero-order Bessel function, r is the source–receiver offset ($r = (h + z) * \tan\theta$), h and z are the vertical distance from the source and the receiver to the interface, respectively, i is the imaginary unit and $x = \cos\theta$, ω is the angular frequency.

The formula (3) shows the spherical-wave reflection coefficient follows the rational function form, so we propose to use rational function (2) to fit seismic reflection coefficients and use the least-square method to solve the problem; finally, the zero-pole and residual attributes are achieved.

In the formula (2), the unknowns A_k appear in the denominator, it is a nonlinear problem, so the Vector Fitting algorithm (Gustavsen and Semlyen 1999) is used to solve the problem, the nonlinear problem can become a linear problem, and then the least-square method is used to solve it.

Specify a set of initial poles A_k in (2), then multiply $H(s)$ with a formula $\sigma(s)$, where $\sigma(s)$ is a rational approximation to be determined; this gives the augmented problem:

$$\begin{pmatrix} \sigma(s)H(s) \\ \sigma(s) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^n \frac{C_k}{s - A_k} + D \\ \sum_{k=1}^n \frac{\tilde{C}_k}{s - A_k} + 1 \end{pmatrix} \tag{5}$$

Note that in (5) the rational approximation for $\sigma(s)$ and the approximation for $\sigma(s)H(s)$ have the same poles.

Multiplying the second row in (5) with $H(s)$ draw the following relation:

$$\sum_{k=1}^n \frac{C_k}{s - \tilde{A}_k} + D = \left[\sum_{k=1}^n \frac{\tilde{C}_k}{s - \tilde{A}_k} + 1 \right] H(s) \tag{6}$$

Or

$$(\sigma H)_{fit}(s) = \sigma_{fit}(s)H(s) \tag{7}$$

Equation (6) is linear in its unknowns C_k, \tilde{C}_k . So writing for several incident angle points, the overdetermined linear problem is given:

$$A\tilde{x} = b \tag{8}$$

where the unknowns are in the solution vector \tilde{x} . Equation (8) is solved as a least-square problem as follows: for a given incident angle θ_k we can get s_k and:

$$A_k \tilde{x} = b_k \tag{9}$$

where

$$A_k = \left[\frac{1}{s_k - \tilde{A}_1} \cdots \frac{1}{s_k - \tilde{A}_N} \quad 1 \quad \frac{-H(s_k)}{s_k - \tilde{A}_1} \cdots \frac{-H(s_k)}{s_k - \tilde{A}_N} \right] \tag{10}$$

$$\tilde{x} = [c_1 \cdots c_N \quad D \tilde{c}_1 \cdots \tilde{c}_N] \tag{11}$$

$$b_k = H(s_k) \tag{12}$$

Thus, a rational function approximation for $H(s)$ is easy to get from (6) now.

If each sum of partial fractions of (6) is written as a fraction, it becomes obvious that it can be expressed as the form of zeros and poles:

$$(\sigma H)_{\text{fit}}(s) = h \frac{\prod_{k=1}^{n+1} (s - z_k)}{\prod_{k=1}^n (s - \tilde{A}_k)}, \sigma_{\text{fit}}(s) = \frac{\prod_{k=1}^n (s - \tilde{z}_k)}{\prod_{k=1}^n (s - \tilde{A}_k)} \tag{13}$$

where $z_k, (k = 1, 2, 3 \dots n)$ are the zeros of $(\sigma H)_{\text{fit}}(s), \tilde{A}_k$ are the poles of $\sigma_{\text{fit}}(s)$ and $(\sigma H)_{\text{fit}}(s), \tilde{z}_k$ are the zeros of $\sigma_{\text{fit}}(s)$. Finally we can get:

$$H(s) = \frac{(\sigma H)_{\text{fit}}(s)}{\sigma_{\text{fit}}(s)} = h \frac{\prod_{k=1}^{n+1} (s - z_k)}{\prod_{k=1}^n (s - \tilde{z}_k)} \tag{14}$$

Equation (14) shows the poles of $H(s)$ become equal to the zeros of $\sigma_{\text{fit}}(s)$. Thus, by calculating the zeros of $\sigma_{\text{fit}}(s)$ we can get a good set of poles for fitting the original function $H(s)$. And the residuals for $H(s)$ can be directly calculated from Eq. (14).

Vector Fitting is equally well suited for fitting vectors as it is for scalars. By replacing the scalar by a vector, it will result in all elements of the fitted vector sharing the same poles.

Results

Fitting

In this example, the spherical-wave reflection coefficient model of Castagna et al. (1998) and Li et al. (2016) is used, shown in Table 1, where v_1, v_2, ρ_1, ρ_2 as mentioned earlier.

Table 1 Parameters of AVO models (two-layer models)

Model	$v_1/(m s^{-1})$	$\rho_1/(g cm^{-3})$	$v_2/(m s^{-1})$	$\rho_2/(g cm^{-3})$
A (avo1 gas/sand)	3093	2.40	4050	2.21
B (avo1 wet/sand)	3093	2.40	4114	2.32
C (avo2 gas/sand)	2642	2.29	2781	2.08
D (avo2 wet/sand)	2642	2.29	3048	2.23
E	2000	2.40	2933	2.20

Firstly, the rational fitting property is demonstrated by fitting the AVO coefficient curves in frequency domain. Because the spherical-wave reflection coefficient is frequency dependent, so in the frequency domain, the rational function fitting is done to the original wide-angle seismic data corresponding to each frequency.

The initial poles are predetermined; first, the optimum fitting conditions, including poles and order, of each frequency are determined by observing the relative error tolerance in the iteration. When the relative error tolerance is less than -40 dB, the iterations are stopped and optimal fitting amplitude and phase are achieved. The point source has a Ricker wavelet with dominant frequency of 10 Hz, and hence we choose the curve of 10 Hz to analysis.

The frequency-domain AVO coefficient curve's amplitudes and phases variation with angles of model A–D are shown in Fig. 1 [each sub-figure (a) and (c)], and the fitting results are shown in the same figure. It is clear that the complex AVO curves are almost fitted. The poles distribution and the relative residuals are plotted in each sub-figure (b) and (d). From the results, we can see that the smaller the critical angle, the more serious the oscillation after it, and the more poles are needed to fit. Also, the curves with the larger critical angle can be fitted with less poles, so the critical angle has an effect on the order of rational fitting.

Although much more poles are needed to fit the complex AVO, the first few residuals are much more than the others, so we guess the first few residuals are the main attribute of the complex AVO. In Fig. 2 Model E (fit result), at least 12 pole and residual pairs are needed to fit the curve with the tolerance less than -40 dB. Now the first 6 large poles and residuals are used to reconstruct the curve, and results are shown in Fig. 2 Model E (reconstruct result); it can be seen the curve reconstructed by using these 6 poles and residuals are basically close to the original curve, and only a few oscillations after the critical angle are not fitted. So these larger poles and residuals do contain most information of the

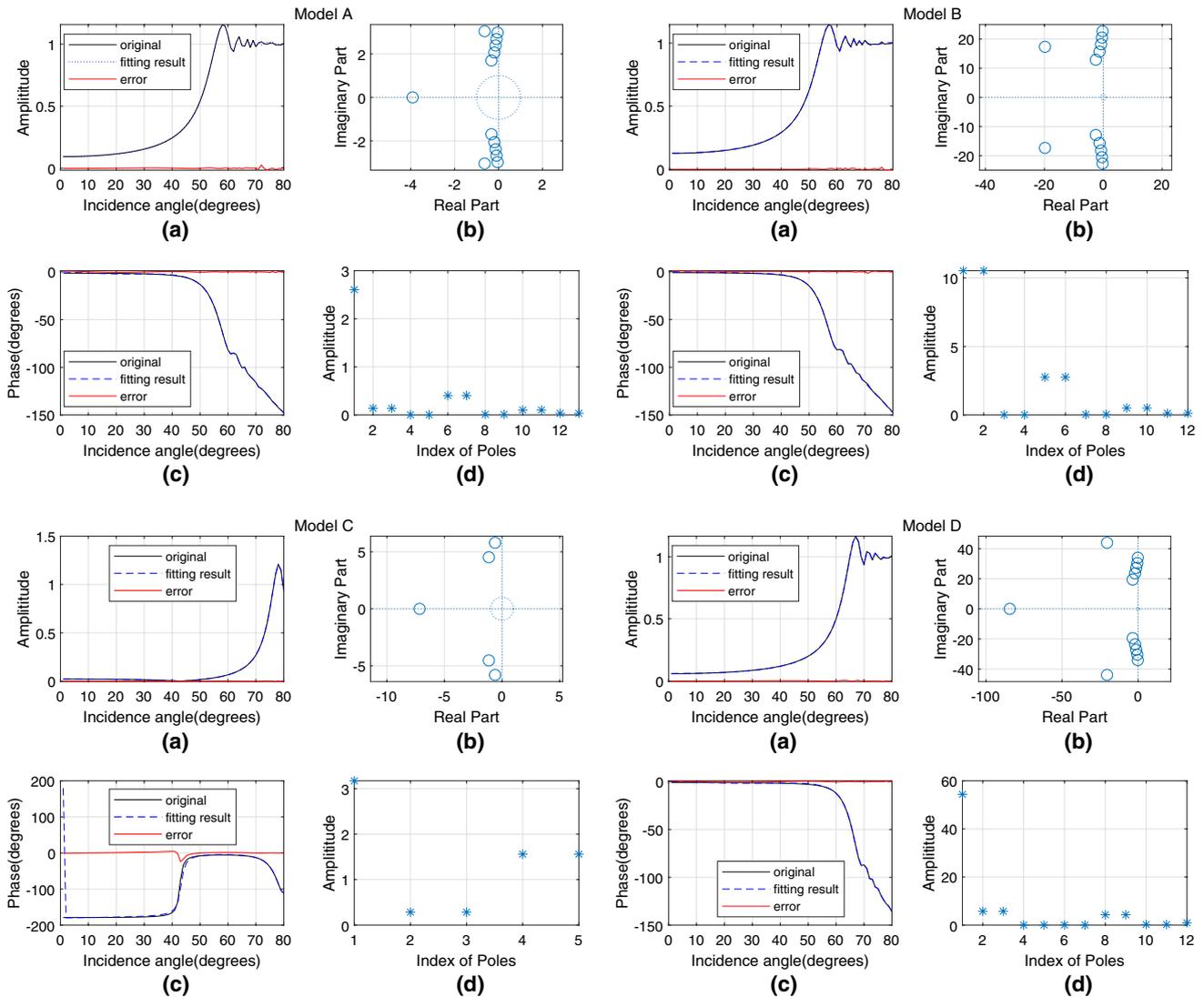


Fig. 1 One of the original AVO coefficient curves (solid line) and the rational function fitting result curves (dotted line) in frequency domain of models A–D: (a) amplitudes. (c) Phases; the AVO attributes extracted from fitting curves (b) poles distribution. (d) Amplitudes of residuals

curve, and we can use these poles and residuals as attributes to represent wide-angle AVO.

In order to observe the sensitivity between residual and the intercept-gradient value of polynomial fitting technique (Shuey 1985) to the change of velocity or density, we analyze the property variation with velocity and density shown in Figs. 3 and 4 of model E.

Figure 3 tells us that when one of two layers of medium models velocity changes, the amplitudes of the residuals, the intercept or gradient values and velocities are both linear correlation, but the variations of the amplitudes of

the residuals caused by the velocities are significantly bigger than the variations of intercept or gradient values; the residual attributes are more sensitive to velocity.

And as shown in Fig. 4, density has the same conclusion with velocity. Therefore, the rational function fitting method we put forward is reasonable and has advantages; it can characterize wide-angle AVO properties, provide more sensitive attributes for wide-angle AVO than intercept-gradient properties. Model A and model B are tested, and they got the same results.

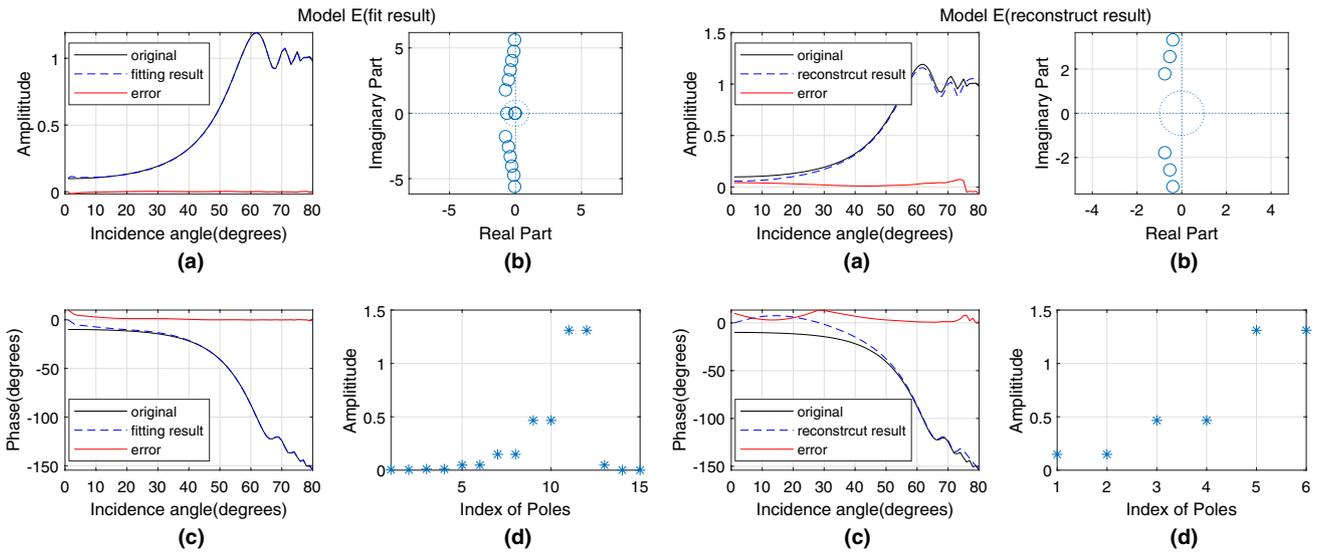
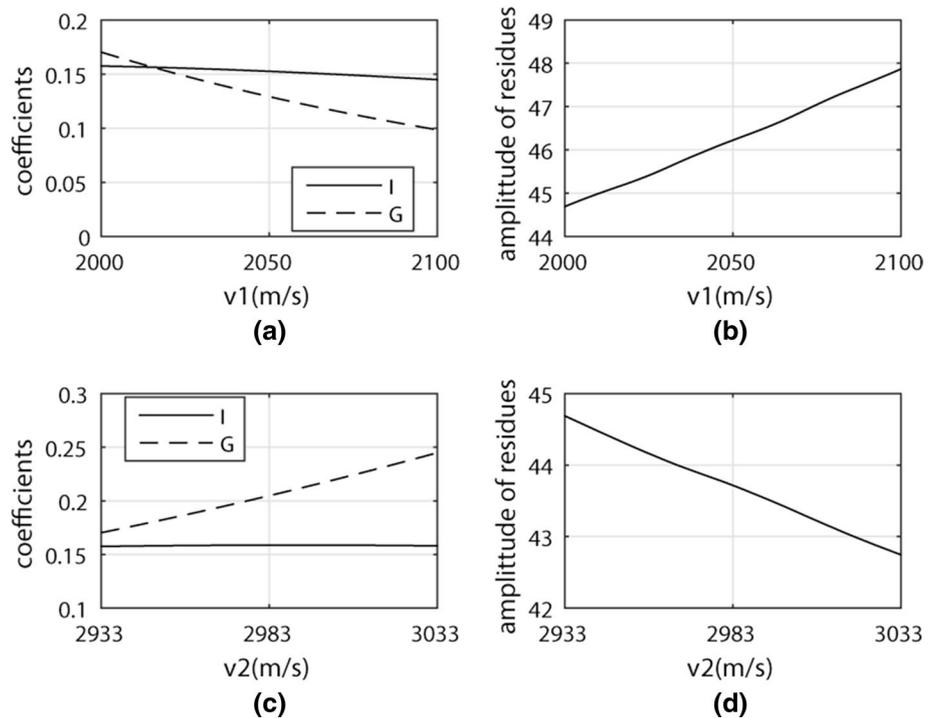


Fig. 2 One of the original AVO coefficient curves (solid line) and the fit/reconstruct result curves (dotted line) in frequency domain of model E: **a** Amplitudes. **c** Phases. **b** Poles distribution. **d** Amplitudes of residuals

Fig. 3 Model E: **a** intercept-gradient variation with v_1 . **b** Amplitudes of residuals variation with v_1 . **c** Intercept-gradient variation with v_2 . **d** Amplitudes of residuals variation with v_2



Applying

Based on previous research results, in this section, we use this method to classify AVO and recognize hydrocarbon signatures. For the different properties of models A–D, we

make cross-correlation analysis of them. For vertical comparison to classify AVO, we compared models A and C, they all belong to gas/sand model, but model A represents AVO class 1, while model C represents class 2; models B and D, and they all belong to wet/sand model, but model

Fig. 4 Model E: **a** Intercept-gradient variation with ρ_1 . **b** Amplitudes of residuals variation with ρ_1 . **c** Intercept-gradient variation with ρ_2 . **d** Amplitudes of residuals variation with ρ_2

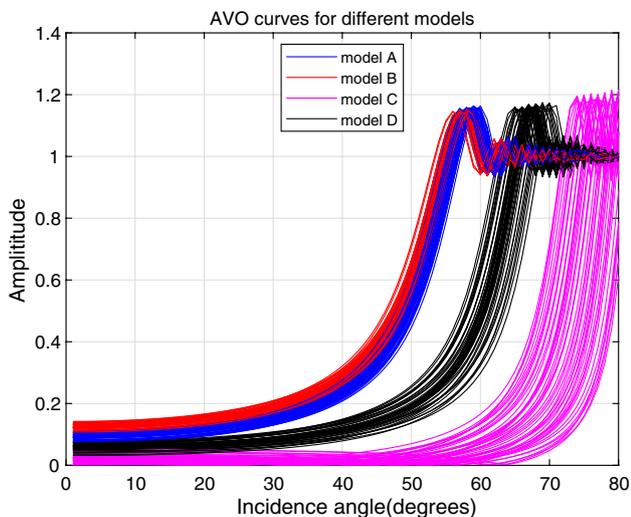
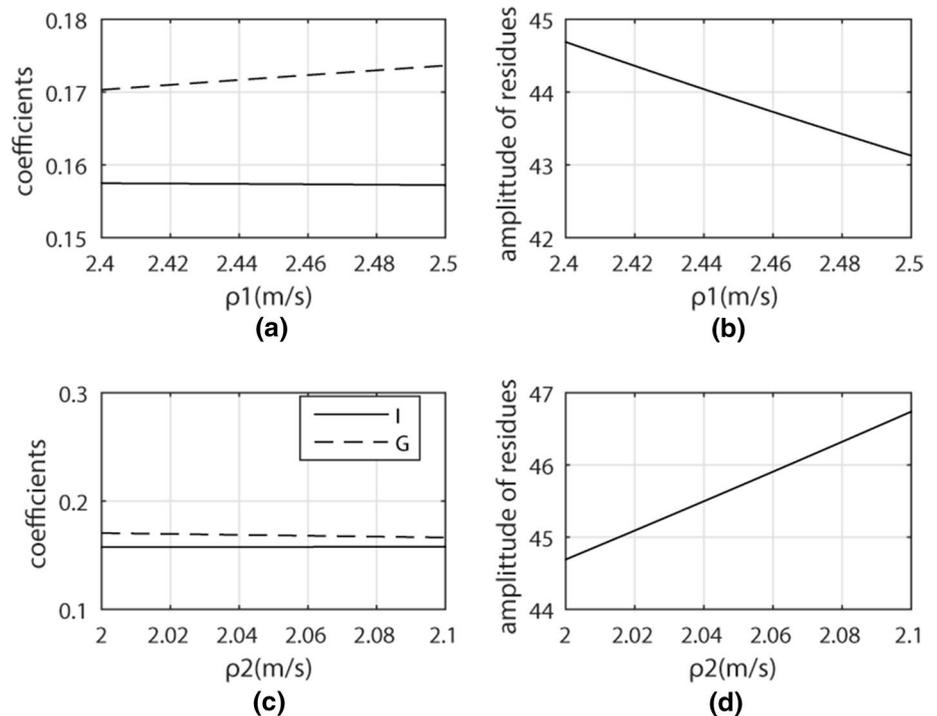


Fig. 5 AVO curves' amplitudes for different models

B represents AVO class 1, while model D represents class 2. For horizontal comparison to recognize hydrocarbon signatures, we compared models A and B, and they all belong to AVO class 1, but model A represents gas/sand model, while model B represents wet/sand; models C and D all belong to AVO class 2, but model C represents gas/sand model, while model D represents wet/sand.

First of all, we get a set of curves of models A–D, and these curves are obtained by adding random perturbations to geological parameters. The amplitudes of these curves are as Fig. 5 shows.

Then as in the preceding fitting process, the rational function zero-poles and residuals are obtained. At the same time, for comparison, we also make polynomial fitting of the data and get the intercept-gradient parameters for small angles lowered 40° .

As mentioned earlier, the first residual often contains more information about the curves, so we compare the intercept-gradient attribute with the first residual under the same scale. The contrast result of model A and C are shown in Fig. 6. From these experiments, Fig. 6a shows the intercept-gradient property and Fig. 6b shows the residuals distribution. In each picture, the circle represents the model A, and the cross denotes the model C.

It can be seen from the diagram that, under the same scale, the two AVO classes expressed by the intercept-gradient attribute are almost overlapped and can't be distinguished, while the two AVO classes expressed by the residual attributes can be distinguished clearly. And the results of models B and D are shown in Fig. 7, and we can get the same results as Fig. 6 shows.

In addition to classify AVO, our method can also recognize hydrocarbon signatures and classify reflector responses. Figure 8 shows the contrast results of models A and B, and

Fig. 6 Comparison of AVO attributes from polynomial and rational function fitting method (model A: circle, model C: cross). **a** Intercept-gradient cross-plot (polynomial fitting). **b** Largest residuals distribution (rational function fitting)

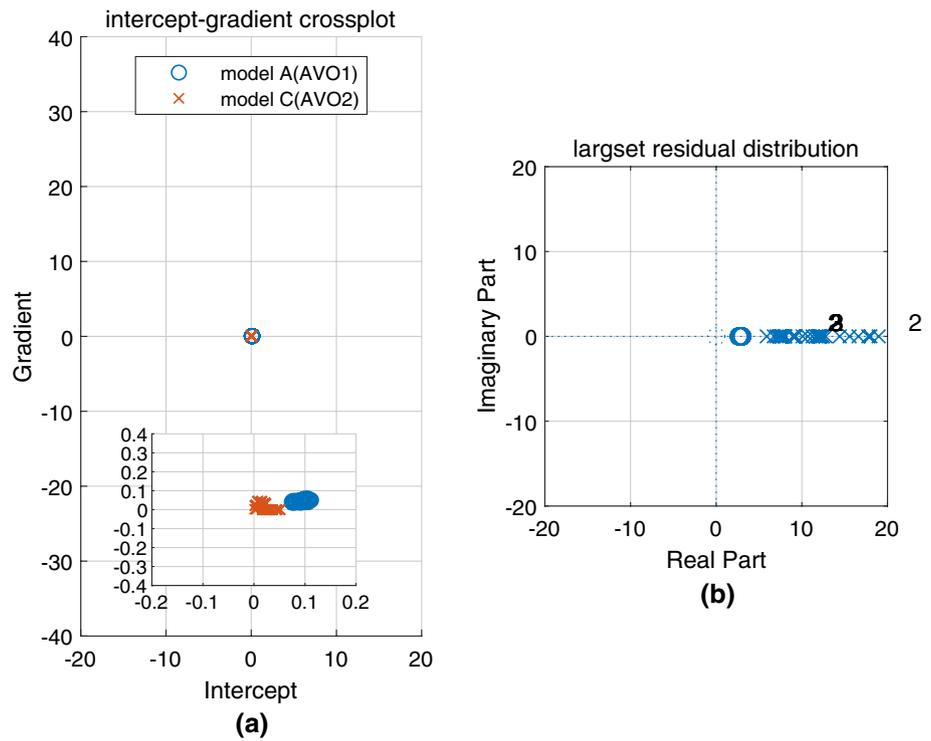
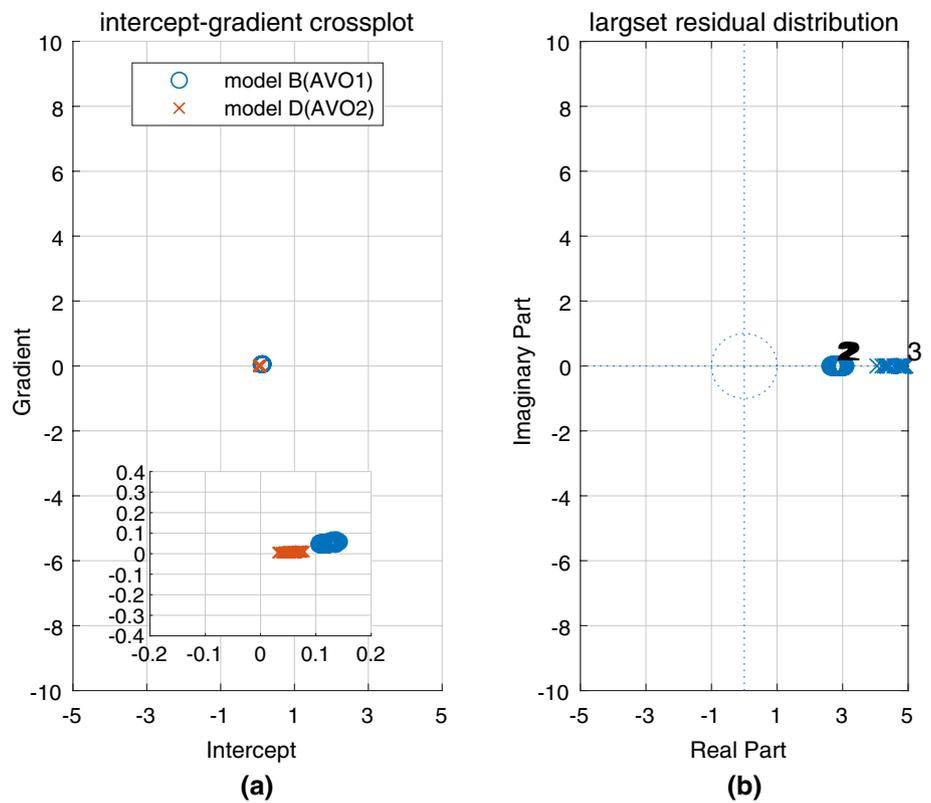


Fig. 7 Comparison of AVO attributes from polynomial and rational function fitting method (model B: circle, model D: cross). **a** Intercept-gradient cross-plot (polynomial fitting). **b** Largest residuals distribution (rational function fitting)



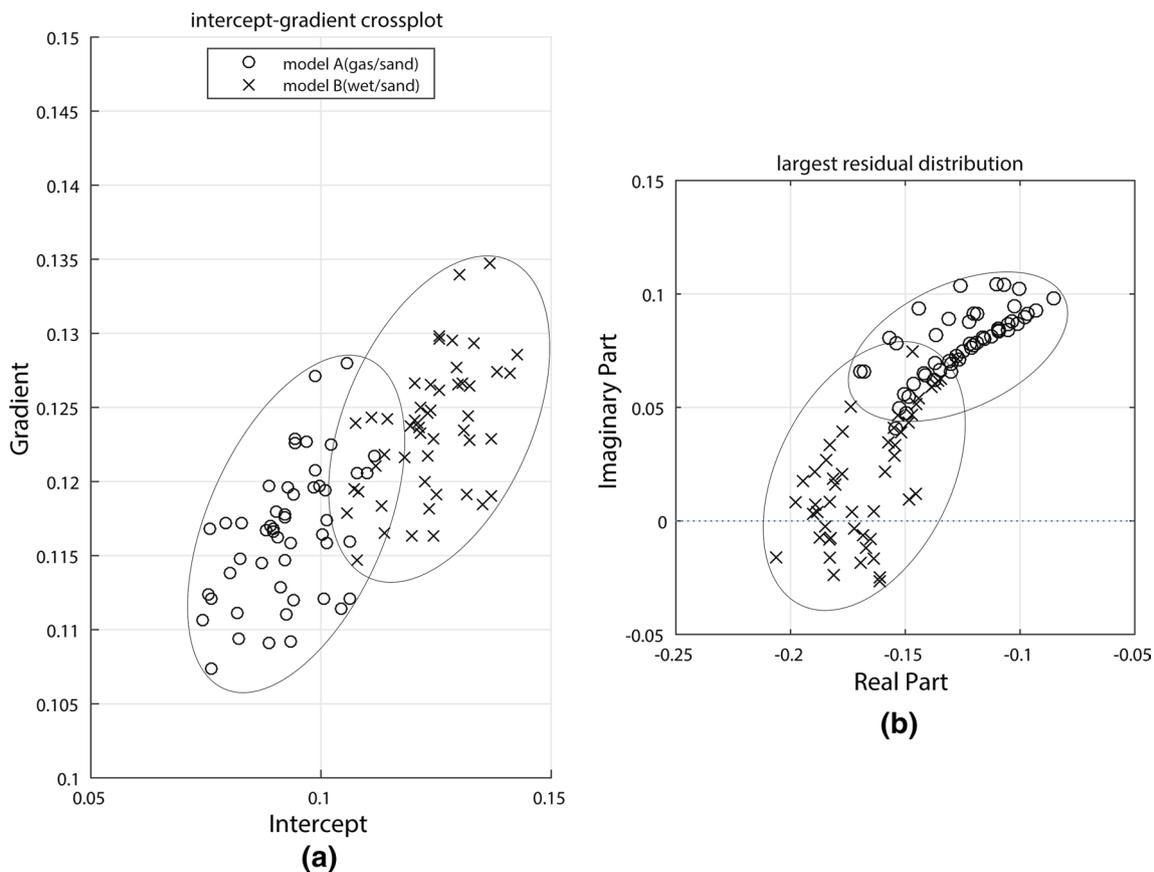


Fig. 8 Comparison of AVO attributes from polynomial and rational function fitting method (model A: circle, model B: cross). **a** Intercept-gradient cross-plot (polynomial fitting). **b** Largest residuals (rational function fitting)

Fig. 9 shows the contrast results of models C and D. And by calculating the center distance between the AVO models, the intercept-gradient distance between the models A and B is about 0.0483 and between models C and D is about 0.0557; however, according to the zero-pole and residual characteristics of rational function fitting, the distance between the largest residuals of models A and B is about $0.1006 > 0.0483$, between models C and D is about $36.0907 > 0.0557$. Therefore, the distance calculated by our method is larger than the original polynomial fitting method, and the use of rational zero-pole and residual properties can classify reflector responses more accurately.

Conclusion

The purpose of this work is clear: to extract more useful information out of seismic data by extending the validity range of AVO analyses to larger angles. In order to achieve this goal, we have applied a new method to represent

wide-angle AVO, that is rational function fitting, based on Vector Fitting algorithm.

By doing rational function fitting to AVO curves, we obtain better approximations of wide-angle AVO and get sparse zero-pole and residual properties and thereby can improve the classification of AVO and reflector responses.

By using this method, the research range can be extended to wide-angle, and the fitting effect is ideal in large offsets. The residual properties are more sensitive to medium velocity or density than intercept-gradient properties, which can be more accurately characterized AVO attributes. The limitation of polynomial fitting technique only in the small incidence angles can be improved. And under the same scale, the intercept-gradient attributes of AVO models are not easy to classify, while the zero-pole and residual attributes can be easily classified. So, the AVO or reflector responses can be classified more accurately by using rational function fitting than polynomial fitting. In this paper, our research is limited in the two-layer medium, and the medium analysis with more layers is our future work.

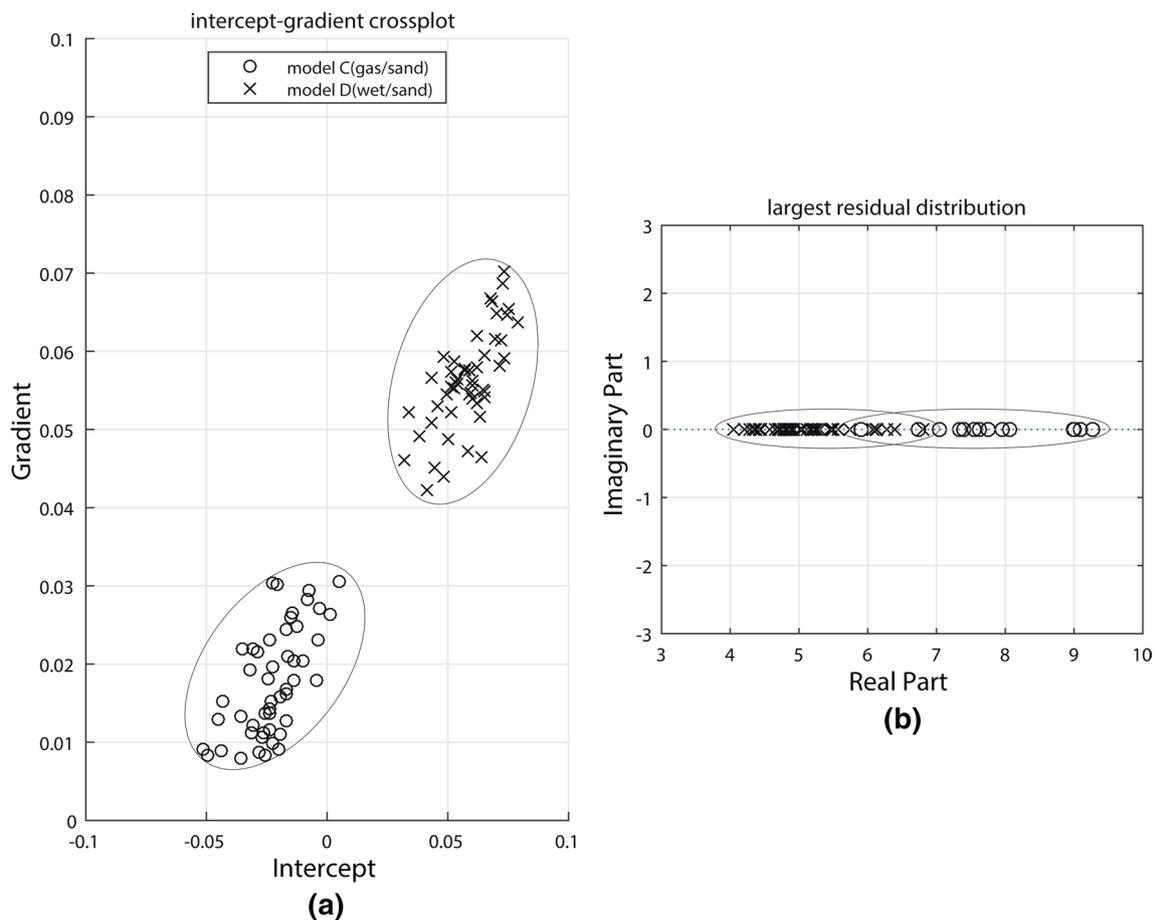


Fig. 9 Comparison of AVO attributes from polynomial and rational function fitting method (model C: circle, model D: cross). **a** Intercept-gradient cross-plot (polynomial fitting). **b** Largest residuals (rational function fitting)

Acknowledgements The research leading to this paper is supported by the China University of Petroleum (Beijing). We would like to thank a lot of people who have been involved in the AVO/AVA theory research and Vector Fitting algorithm, and especially want to thank Jingnan Li for spherical-wave reflection coefficient codes and Guangtan Huang for his advice and guidance.

References

- Aki K, Richards PG (1980) Quantitative seismology, theory and methods. I. W. H. Freeman, New York
- Ayzenberg M, Tsvankin I, Aizenberg A, Ursin B (2009) Effective reflection coefficients for curved interfaces in transversely isotropic media. *Geophysics* 74(5):WB33–WB53
- Bortfeld R (1961) Approximations to the reflection and transmission coefficients of plane longitudinal and transverse waves. *Geophys Prospect* 9(4):485–502
- Castagna JP, Foster DJ, Swan HW (1998) Framework for avo gradient and intercept interpretation. *Geophysics* 63(3):948–956
- Červený V, Hron F (1961) Reflection coefficients for spherical waves. *Studia Geophysica Et Geodaetica* 5(2):122–132
- Debski W, Tarantola A (1995) Information on elastic parameters obtained from the amplitudes of reflected waves. *Geophysics* 60(5):1426–1436
- Downton JE, Ursenbach C (2006) Linearized amplitude variation with offset (avo) inversion with supercritical angles. *Geophysics* 71(5):E49–E55
- Gustavsen B, Semlyen A (1999) Rational approximation of frequency domain responses by vector fitting. *IEEE Trans Power Deliv* 14(3):1052–1061
- Haase AB (2004) Spherical wave avo modeling of converted waves in isotropic media. *Seg Tech Progr Expand Abstr* 23(1):2586
- Krail PM, Brysk H (1983) Reflection of spherical seismic waves in elastic layered media. *Geophysics* 48(6):655–664
- Li J, Xu WS, Hui DC, Yi YS, Bo WJ (2016) Study on frequency-dependent characteristics of spherical-wave PP reflection coefficient. *Chin J Geophys* 59:3810–3819
- Macdonald C, Davis PM, Jackson DD (1987) Inversion of reflection traveltimes and amplitudes. *Geophysics* 52(5):606–617
- O'Brien P (1963) A note on the reflection of seismic pulses with application to second event refraction shooting. *Geophys Prospect* 11(1):59–72
- Shuey RT (1985) A simplification of the Zoeppritz equations. *Geophysics* 50(5):609–614

- Ursenbach C, Haase AB, Downton JE (1949) An efficient method for avo modeling of reflected spherical waves. *Seg Tech Progr Expand Abstr* 24(1):202–205
- Winterstein DF (1985) Supercritical reflections observed in p- and s-wave data. *Geophysics* 50(2):185
- Zhu X, Mcmechan GA (2015) Elastic inversion of near- and post-critical reflections using phase variation with angle. *Geophysics* 77(4):149



The added value of spatially distributed meteorological data for simulating hydrological processes in a small Mediterranean catchment

Ahlem Gara^{1,2} · Khoulood Gader² · Slaheddine Khelifi¹ · Marnik Vanclooster³ · Donia Jendoubi⁴ · Christophe Bouvier⁵

Received: 14 August 2018 / Accepted: 17 October 2019 / Published online: 25 October 2019

© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

The purpose of this paper was to demonstrate the added value of the spatial distribution of rainfall and potential evapotranspiration (PE) in the prediction of the discharge for a small Mediterranean catchment located in the Medjerda basin in Tunisia, i.e. the Raghay. We compare therefore the performance of a conceptual hydrological model available in the ATHYS platform, using global and spatial distributed input data. The model was implemented in two different ways. The first implementation was in a spatially distributed mode, and the second one was in a non-distributed lumped mode by using spatially averaged data weighed with a Thiessen-interpolated factor. The performance of the model was analysed for the distributed mode and for the lumped mode with a cross-validation test and through several modelling evaluation criteria. Simultaneously, the impact of the spatial distribution of meteorological data was assessed for the two cases when estimating the model parameters, the flow and water amounts, and the flow duration curves. The cross-validation of the split-sample test shows a preference for the spatially distributed model based on accuracy criteria and graphical comparison. The distributed mode required, however, more simulation time. Finally, the results reported for the Raghay indicated that the added value of the spatial distribution of rainfall and PE is not constant for the whole series of data, depending on the spatial and temporal variability of climate data over the catchment that should be assessed prior to the modelling implementations.

Keywords Medjerda · ATHYS · Hydrological modelling · Distributed PE and rainfall · Accuracy criteria

Introduction

The assessment of the water resources as support for water management can be based either on monitoring or hydrological modelling programs (Ibrahim et al. 2015). In Africa, however, the hydrological monitoring programs are often deficient. Hydrological monitoring gauges are often limited to a few observed stations, mainly focused on the main watersheds within a region. Small watersheds of less than one thousand square kilometres are generally not well monitored (Schuol et al. 2008; Ibrahim et al. 2015). For small water catchments with limited data, hydrological modelling remains the second option to assess the water resources and different hydrological functions (Oudin et al. 2008; Coustau et al. 2012; Tegegne et al. 2017). Rainfall–runoff (RR) models are therefore essential tools that support decision-making and water management (Andréassian et al. 2004).

Distributed hydrological RR models are now very popular because they can describe rainfall–runoff response

✉ Ahlem Gara
ahlem_gara@yahoo.fr

¹ U-R:Gestion Durable des Ressources en Eau et en Sol (GDRES), High School of Engineering of Medjez el Bab (ESIM), University of Jendouba, Jendouba, Tunisia

² National Agronomic Institute of Tunisia (INAT), University of Carthage, Tunis, Tunisia

³ Earth and Life Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

⁴ Centre for Development and Environment (CDE), University of Bern, Bern, Switzerland

⁵ HydroSciences Montpellier, Institut de Recherche pour le Développement (IRD), Montpellier, France

with spatially distributed input, which can be managed with modern GIS and remote sensing technology (Bao et al. 2017). A more detailed representation of the spatial variability in RR modelling is definitely sounder from a conceptual basis. Distributed RR modelling is, however, subjected to a lot of uncertainties. The uncertain representation of the hydrological modelling input is an example. Uncertainties related to the spatial variability of climate data have been demonstrated to affect hydrological modelling performance (Sangati and Borga 2009). Distributed RR modelling may also become constraining from an operational perspective. When implementing an RR model, an appropriate selection has to be made considering the trade-offs between model efficiency, validity and robustness (Duan et al. 2003). Model identifiability, i.e. the capacity to implement and parameterize the model in an operational context, can be considered as a part of model efficiency (Gábor et al. 2017). Model identifiability will be high when the model structure is simple; it will be low when the model structure is complex. Hence, spatially distributed hydrological models may suffer from identifiability issues.

In order to reduce the complexity and hence increase the identifiability of distributed RR models, it will be interesting to reduce the number of parameters. A straightforward strategy consists in running the spatially distributed model with partially lumping a part of the spatially distributed parameters. Although it is assumed that such lumping will generally reduce model performance in terms of goodness of fit, the associated results may remain acceptable in terms of identifiability. It is therefore important to analyse carefully the trade-off between goodness of fit and overall modelling efficiency. This problem has been largely studied, and many viable results have been found when focusing on comparing between lumped and distributed modes' performance in order to pick out the best-selected model based on its performance and the running duration. Brulebois et al. (2018) confirmed that the results given by two models, a semi-distributed physically based model SWAT and a global conceptual model GR4j, are rather close when tested over contrasted climate periods, with slightly higher robustness for the SWAT model. In the same context, Coustau et al. (2012) brought to the light the important benefits for the use of distributed hydrological models that take into account the radar data without increasing the number of parameters or the complexity of the model. However, distributed models do not always give better results than the lumped models. Vansteenkiste et al. (2014) found out that lumped models actually gave similar results as the distributed models based on accuracy criteria for different spatial resolutions. Both types of models produced well the seasonality with a preference to lumped models, especially for overall water balance terms and subsurface flow.

Similarly, Khakbaz et al. (2012) certified that lumped implementations were able to sufficiently endorse the effects of spatial variability in precipitation on stream-flow prediction. These results were consistent with those found in many studies comparing lumped and distributed hydrological models, suggesting that the two types of model can give similar results in terms of accuracy and that there is no superior model if several measures of model performance are considered (Ajami et al. 2004; Breuer et al. 2009; Koren et al. 2012; Apip et al. 2012; Lobligeois et al. 2014). According to Lobligeois et al. (2014), the results are largely determined by the catchment complexity and the existing spatial and temporal variability of the hydro-climatic data. Yet, the way how catchment complexity and spatial and temporal variability of hydrologic attributes determines the ultimate performance of distributed versus lumped models needs further investigation.

This paper aims to contribute to this challenge by addressing the following question: How can spatially distributed climate data improve hydrological model prediction in poorly gauged catchments? We select to address this issue of the Raghay catchment which is a small sub-catchment of the Medjerda basin in Tunisia. The Medjerda is suffering from hydrometric data availability in quantity and in quality, which jeopardizes the management of droughts and floods that typically characterize the Mediterranean regions (Ludwig et al. 2011). While analysing this aspect, we also try to answer the following scientific question: Is the selected hydrological conceptual model suitable for reproducing the hydrological response in the Raghay catchment?

In an attempt to respond to this question, the paper is organized as follows: In the first part, a concise description of the runoff and the routing models that are used is given. In addition, the adopted validation method is summarized. The second section describes the study site, followed by a presentation of the available data used to run the model. The third part of this paper illustrates and discusses the use of reanalysis hydro-climatic data in addition to the observed ones. This is followed by an outlining of the main results for the two calibration phases by comparing the goodness of fit, scatter plots and flow duration curves when implementing a conceptual runoff hydrological model in a fully spatial distributed mode (case 1) and in a lumped mode (case 2). The final section summarizes the main conclusions of the paper.

Rainfall–runoff modelling implementation

We implemented RR models for the catchment using the hydrological modelling platform ATHYS (Atelier Hydrologique Spatialisé, Bouvier et al. 2013). This open and free-use software platform (www.athys-soft.org) is continuously updated to allow the improvement of results and

to minimize the simulation time. It presents the possibility to combine many production functions and transfer functions via a consistent and easy-to-use environment, including processing of hydro-meteorological and geographical data. For our case study, we decided to use a modified distributed version of the runoff model of the Soil Conservation Service (SCS) with the integration of potential evaporation (PE) combined with a lag-and-route (LR) routing model. The SCS model was widely used to predict runoff, especially for poorly gauged catchments (Mishra et al. 2003). The LR routing model was also tested in several catchments in southern France (Bentura and Michel 1997; Tramblay et al. 2011).

The model was implemented in two different ways: once in a fully spatial distributed mode (case 1), by using distributed rainfall and PE input data, and once in a non-distributed lumped mode (case 2), by using averaged rainfall and PE input weighted by Thiessen coefficients (Table 1).

This distributed model operates over a grid mesh of regular cells using as input the digital elevation model and permitting the generation of its derived maps such as slopes, flow directions and sub-basin layers (Gara et al. 2015). The time step used to run the model is a daily time step for continuous hydro-climate data to permit taking into account the dry and the wet phases of the hydrological cycle and based on the available data series for a long enough period of observations. Rainfall and PE were interpolated using Thiessen interpolation method and were integrated into the runoff model in order to calculate flows for each grid cell. The LR model processed the elementary hydrograph for each cell and routed this hydrograph to the discharge point of the catchment to provide a complete simulated runoff (Bouvier et al. 2008).

The modified runoff model with PE integration

The Soil Conservation Service (SCS) model, developed by US Soil Conservation Service, has been largely used for

estimating runoff from medium- and small-sized watersheds (Hawkins 1993; Lewis et al. 2000; Liu and Li 2008; Tramblay et al. 2011). The model was also applied to model runoff in Tunisian catchments (Sellami et al. 2013; Dakhlaoui et al. 2017) based on the fact that this model requires only commonly available terrain data. In this paper, we associated the Soil Moisture Accounting (SMA) procedure to the SCS method in the ATHYS platform.

We used here the SCS–SMA formulated by Michel et al. (2005), relating rainfall and runoff based on the relationship between:

$$\begin{aligned} R(t) &= p(t) \times \left(\frac{V(t)-Sa}{S} \right) \times \left(2 - \frac{V(t)-Sa}{S} \right) & \text{if } V(t) > Sa \\ R(t) &= 0 & \text{otherwise} \end{aligned} \quad (1)$$

where $R(t)$ [$L T^{-1}$] denotes the runoff rate, $p(t)$ [$L T^{-1}$] the rainfall intensity at the time t , S [L] the maximal capacity of the soil reservoir, $V(t)$ the level in the soil reservoir at the time t [L] and Sa [L] the initial losses.

The advantage of this model is that the runoff is directly related to the level in the soil reservoir, which eases the use of discharge of this soil reservoir and the formulation of a delayed runoff as a part of this discharge (Fig. 1).

The runoff rate, $R(t)$, can be simplified as the product of the runoff coefficient $C(t)$ [–] and rainfall intensity at time t , $p(t)$, where

$$C(t) = \left(\frac{V(t) - Sa}{S} \right) \times \left(2 - \frac{V(t) - Sa}{S} \right) \quad (2)$$

The soil reservoir is then forced by the intensity of infiltration at time t , $f(t)$ [$L T^{-1}$], which is expressed as follows:

$$f(t) = (1 - C(t)) \times p(t) \quad (3)$$

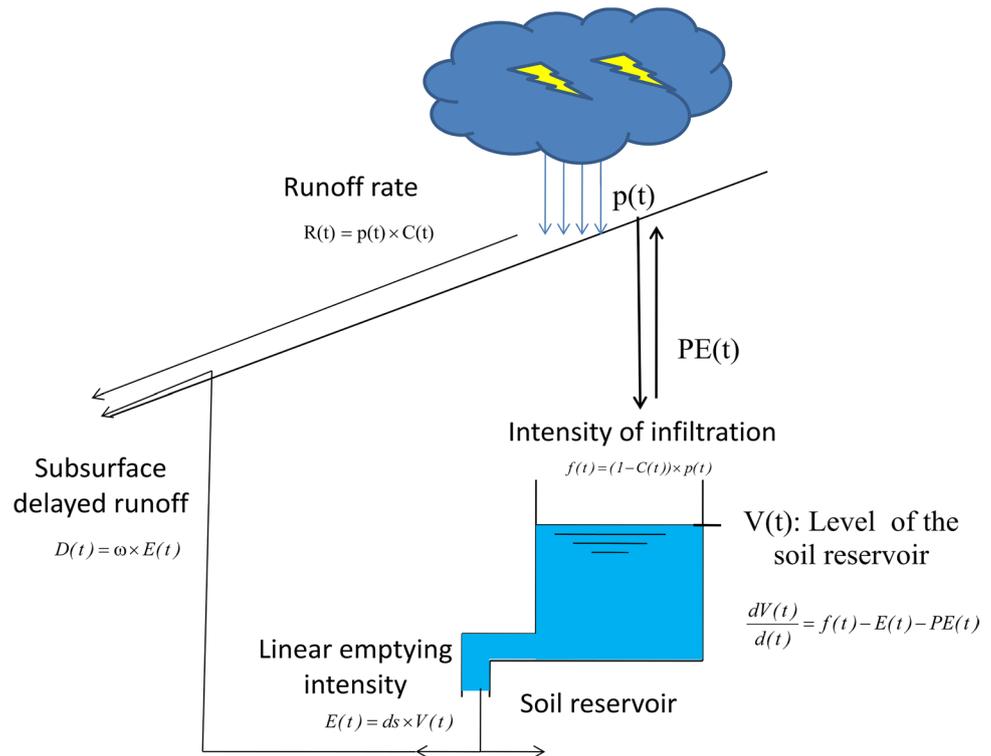
To consider the reduction in the runoff coefficient for non-rainy periods, we used the linear emptying intensity coefficient of the soil reservoir at time t , $E(t)$ [$L T^{-1}$], to

Table 1 Summary description of the methodology used for the two cases of modelling in the Raghay catchment

	Case 1	Case 2
Model description	Distributed SCS–SMA–ETP–simple lag-and-route model	Lumped SCS–SMA–ETP–simple lag-and-route model
Model parameters	$S, \omega, Sa/S, ds, V_0, K_0$	$S, \omega, ds, Sa/s, V_0, K_1$
Input description	*Spatially distributed observed rainfalls (data from six rainfall stations, 18-year duration) *Spatially distributed PE (data from five climatic CFSR data + 1 climatic observed station, 18-year duration) *DEM ^a (30 m × 30 m)	*Averaged observed rainfalls (Thiessen interpolated from six rainfall stations, 18-year duration) *Averaged PE (Thiessen interpolated from five climatic CFSR data + 1 climatic observed station, 18-year duration) *The watershed is considered as one cell
Output description	Simulated runoff as the sum of all the EH produced for each cell	Simulated runoff in only one EH
Cross-validation accuracy criteria	NS, RMSE, RSR, PBIAS, EFF	

^aDigital Elevation Model obtained from the mission of 19 June 2014: <https://gdex.cr.usgs.gov/gdex/>

Fig. 1 Simplified presentation of the runoff model



be dependent on the level $V(t)$ in the soil reservoir and the discharge coefficient ds [T^{-1}]:

$$E(t) = ds \times V(t) \quad (4)$$

assuming to simplify the discharge coefficient due to losses by percolation to the deep aquifer and lateral flow, ds , into a constant parameter over the catchment permitting the decrease in the runoff coefficient during non-rainy periods. Theoretically, we can subdivide the parameter ds into two parts: ds_1 taking account of the losses by delayed flow caused by the lateral flow and ds_2 which is considering the part of percolation to the deep aquifer.

A part ω of this discharge flows back to the outlet of the catchment, as subsurface-delayed runoff (Coustau et al. 2012).

This additional runoff $D(t)$ [$L T^{-1}$]:

$$D(t) = \omega \times E(t) \quad (5)$$

must be added to the surface runoff. In the model, a simple presentation of the hydrological process led to neglecting the part of the delayed runoff caused by the percolation to the deep aquifer. Then, ds_2 is set into 0 and ds_1 , which is directly related to the subsurface-delayed runoff, which can be replaced by ω in the calibrated parameters.

The total calculated runoff $R_t(t)$ [$L T^{-1}$] given by the SCS–SMA procedure is then expressed as follows:

$$R_t(t) = p(t) - f(t) + D(t) \quad (6)$$

The reservoir level $V(t)$ is derived from the balance between the water input (infiltration) and the water output (PE, reservoir discharge due to deep percolation and subsurface flow):

$$\frac{dV(t)}{d(t)} = f(t) - E(t) - PE(t) \quad (7)$$

where the initial level of the soil reservoir is V_0 [L].

Definitely, the SCS–SMA model in this simplified version comprises five parameters to be calibrated: the maximal capacity of soil reservoir S [$L T^{-1}$], the coefficient of the linear emptying of the soil reservoir only by percolation to the deep aquifer ds_2 [T^{-1}], the part of the soil discharge to simulate the delayed runoff ω (dimensionless), the fraction limiting the lateral emptying of the reservoirs Sa/S (dimensionless), and the initial level in the soil reservoir V_0 [L].

The used formula to calculate the PE is developed by Oudin et al. (2005).

$$PE = \frac{Re}{\lambda} \times \frac{Ta+5}{100} \text{ if } Ta + 5 > 0$$

$$PE = 0 \text{ otherwise,} \quad (8)$$

where PE (mm/day), Re is extraterrestrial radiation (MJ/m^2 day), λ is the latent heat flux (MJ/kg) and Ta is the mean daily air temperature ($^{\circ}C$), derived from the long-term average.

The used formula to calculate the PE relies on the results obtained by Andréassian et al. (2004) and Oudin et al. (2005)

when comparing results for the 27 PE-tested formulations. These results proved that a very simple version method which only requires extraterrestrial radiation and mean daily temperature is a sufficient and robust method compared to more complex methods.

Routing model

The SCS–SMA model was connected to the LR routing model. This LR model allows the runoff to be routed from every grid cell to the discharge point of the catchment (Coustau et al. 2012) (Fig. 2). The calculation of the time of travel, T_m , is given by:

$$T_m = \sum \frac{l_k}{V_k} \tag{9}$$

where l_k and V_k are, respectively, the lengths and the velocity over the k -cells of the trajectory between the cell m and the outlet. Here, V_k will be assumed to be uniform and constant at the catchment scale, $V_k = V_o$, for the sake of simplicity.

For each grid cell, the time of propagation (T_m) was forecasted by the routing model and a diffuse time (K_m) is computed as follows:

$$K_m = K_o \times T_m \tag{10}$$

where K_o (dimensionless) is the diffusion coefficient (lag), which has been preset to 0.7.

The elementary discharge $q(t)$ due to the runoff $R(t_0)$ of cell m at time t_0 (Tramblay et al. 2011) is calculated as follows:

$$q_m(t) = \begin{cases} 0 & \text{when } t < t_0 + T_m \\ \frac{R(t_0)}{K_m} \exp\left(-\frac{t-(t_0+T_m)}{K_m}\right) \cdot A & \text{otherwise} \end{cases} \tag{11}$$

Then, the final obtained runoff for the catchment consists of the sum of all the elementary hydrographs produced at the outlet of each grid cell.

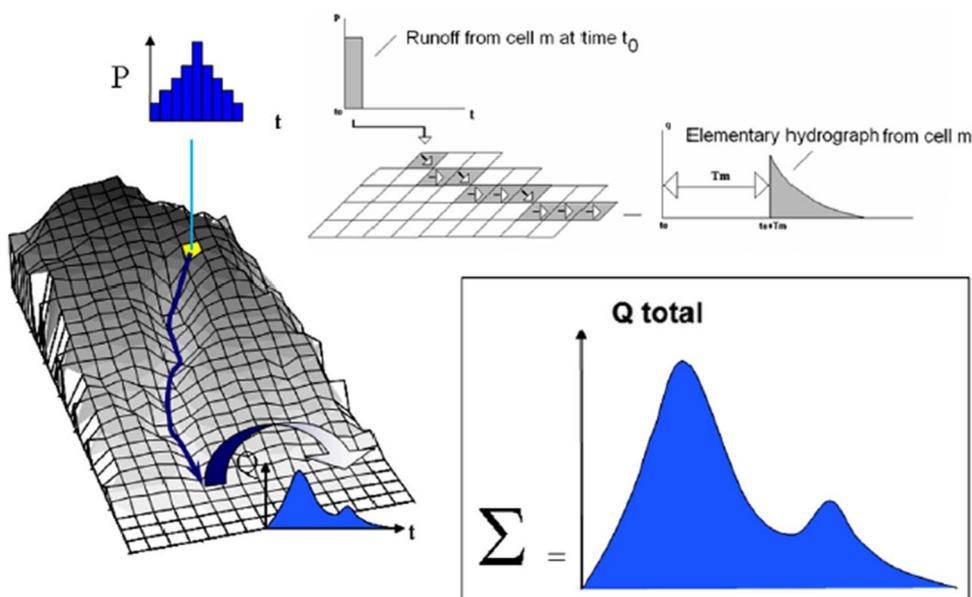
When the LR model is used in the lumped case, only one cell is considered, of which the area equals the total area of the catchment. Therefore, there is only one elementary hydrograph produced at the outlet. It yields the simulated runoff, routed to the outlet by means of a simple linear reservoir, with one parameter K_1 (mn), to be calibrated. The V_o parameter is no more useful and should be set to a high value ($V_o = 1000$ m/s).

The parameters of the LR model to be calibrated are thus either V_o in distributed mode (K_o is empirically set to 0.7) or K_1 in lumped mode. The production parameters are the same than those for the distributed model and are considered to be constant over the Raghay catchment; then, the distribution only concerned the rainfall and PE interpolated for each cell.

Cross-validation procedure and performance evaluation

The model performance was assessed for the two cases: (1) distributed mode and (2) lumped mode, using a split-sample cross-validation test (Klemeš 1986). Since we have access to a 18 years' time series of hydro-climatic data, we subdivided this series of data into two equal parts. The first part (P1) presents data going from September 1990 to August 1999, and the second part (P2) includes data ranging from September 1999 until August 2008. Afterwards, we went through a calibration for P1 and validation for P2. This operation is called the first step of the split-sample test procedure. Subsequently, we switched the data by using P2 for the calibration and P1 for the validation for the second step of the cross-validation. In order

Fig. 2 Diagram of the lag-and-route routing model (Tramblay et al. 2011)



to obtain the parameters for calibration for the two steps of the cross-validation, we used the BLUE (best linear unbiased estimator) algorithm.

The BLUE algorithm, also referred to the error covariance matrix analysis method, is a method of optimization allowing to minimize a cost function (evaluation criteria) based on both the differences between values simulated by the model and observed values (volumes, flow rates, soil moisture content, velocities, etc.), and the differences between initial and optimal values of the model parameters (Henderson 1975). This method is derived from the formulation of the Sherman–Morrison–Woodbury algorithm (Sherman and Morrison 1950). The minimization algorithm of the cost function is based on a gradient method allowing to minimize the running time of the model and then to converge rapidly. The adopted version for the ATHYS platform introduces constraints on the variations of the parameters to be optimized. These constraints are directly linked to the confidence of their initial estimate. This method allows optimizing all the parameters to be calibrated in the RR model. However, the user must specify confidence levels on parameter estimates and observations. The user must also specify the increment (disruption) steps for calculating the derivative of the incremental cost function. Therefore, the number of iterations in the optimization procedure corresponds to an effective realization of an equivalent number of iterations. Then, this optimization is essentially based on the knowledge of the user of the hydrological behaviour for the studied catchment in addition to the status and parameter estimation of the model.

We assessed the runoff prediction accuracy of the model by calculating several performance indicators for hydrological modelling such as the Nash–Sutcliffe efficiency (NS) (Nash and Sutcliffe 1970), the root-mean-square error (RMSE), the ratio of RMSE to the standard deviation of the observations (RSR) (Moriassi et al. 2007) and the Percent BIAS (PBIAS) (Singh et al. 2005). These performance indicators are, respectively, expressed as follows:

$$NS = 1 - \frac{\sum_{i=1}^n (S_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (12)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (S_i - O_i)^2}{N}} \quad (13)$$

$$RSR = \frac{\left(\sqrt{\sum_{i=1}^n (O_i - S_i)^2} \right)}{\left(\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \right)} \quad (14)$$

$$PBIAS = \frac{\sum_{i=1}^n (O_i - S_i)}{\sum_{i=1}^n (O_i)} \times 100 \quad (15)$$

where O_i is the observed runoff; S_i : the simulated runoff; \bar{O} : the mean observed runoff; n : the number of pairs of the measured and simulated variables.

The NS is widely used for efficiency determination of hydrologic models. It is considerably acceptable when it is greater than 0.5 (Wöhling et al. 2013).

The RMSE is widely used as a goodness-of-fit indicator that describes the difference between the observed and predicted values in the same units. Smaller RMSE values describe a better model for runoff prediction. The RSR can also provide additional information and can be applied to a variety of different constituents. The RSR is considered satisfactory when it is below 0.7.

Similarly, the PBIAS quantifies a model's tendency to underestimate or overestimate values, where a value of zero (optimum) shows a perfect fit. Positive (negative) bias results indicate model underestimation (overestimation). PBIAS is considered satisfactory when it is below 0.25. A performance rating based on PBIAS was used by Moriassi et al. (2007).

To further compare between the simulation of the two model implementations at the same time, we calculated the EFFiciency index (EFF) (Brocca et al. 2010), which is expressed as follows:

$$EFF = 1 - \frac{\sum (Q_{\text{sim_distributed}}(t) - Q_{\text{Obs}}(t))^2}{\sum (Q_{\text{sim_global}}(t) - Q_{\text{Obs}}(t))^2} \quad (16)$$

where $Q_{\text{sim_distributed}}$ is the simulated discharge with spatial rainfall and PE.

$Q_{\text{sim_global}}$ is the simulated discharge using mean Thiessen-interpolated rainfall and PE in the lumped model, and Q_{Obs} is the observed discharge.

The EFF has been used to evaluate the efficiency of spatially distributed meteorological data compared to uniform meteorological data for RR modelling. If EFF is greater than 0, then the use of spatially distributed rainfall produces an improvement in the runoff simulation by the model (Tramblay et al. 2011).

These quantitative evaluations were reinforced by visually evaluating the model performance in comparison with the combined observed and estimated runoff values through scatter plots.

To additionally enhance the emphasis of the spatial variability for the observed rainfalls in the Raghay catchment, we introduced a climatic index calculated for the two parts of the data, P1 and P2. The Standardized Precipitation Index (SPI), developed by McKee et al. (1993), is mainly used in order to outline and to quantify the shortage or the

abundance of the precipitation in regional or local scale for a specified period. This index remains an effective and flexible tool to highlight the severity of the drought (Zkhiri et al. 2019).

This index is expressed as follows:

$$\text{SPI}_{12} = \frac{P_i - P_m}{\sigma} \quad (17)$$

where SPI₁₂ is the annual calculated SPI (without dimension), P_i is the annual precipitation for a given year (mm/year), P_m is the mean annual precipitation (mm/year) and σ is the temporal standard deviation calculated for each rain gauge during P1 and P2 for annual rainfalls.

Because of its capability to classify the severity of the drought or the humidity, we applied this index in an attempt to assess the spatial and temporal variability over the catchment calculated for the six observed rain gauges during P1 and P2.

Study area and data assessment

Study area

The Raghay catchment is a Mediterranean catchment located in the rural area in the north-western of Tunisia (Fig. 3). This catchment is one of the main tributaries of the high valley of the Medjerda river, which is the only permanent waterway in the country. The length of the main Raghay tributary is about 35 km, which stretches from the confluence with the mainstream network of Medjerda watershed up to the Algerian border and draining an area of 362 km² among which almost 40 km² being in Algeria (Gara et al. 2015). The catchment has a contrasted topography: having an altitude range varying between 138 and 1183 m, characterized by a high slope in the upstream mountainous part and a weak one in the downstream part, ranging then from 2 to 4%. Dominant land-use types are mostly forest with a typical Mediterranean vegetation cover. The soils are relatively thin, especially in the mountainous part, from 10 to 15 cm at the top of the hill slopes to 120 cm near to the river bed. Soil types are mainly silty loam in the upstream parts and sandy loam in the downstream parts and near the river bed.

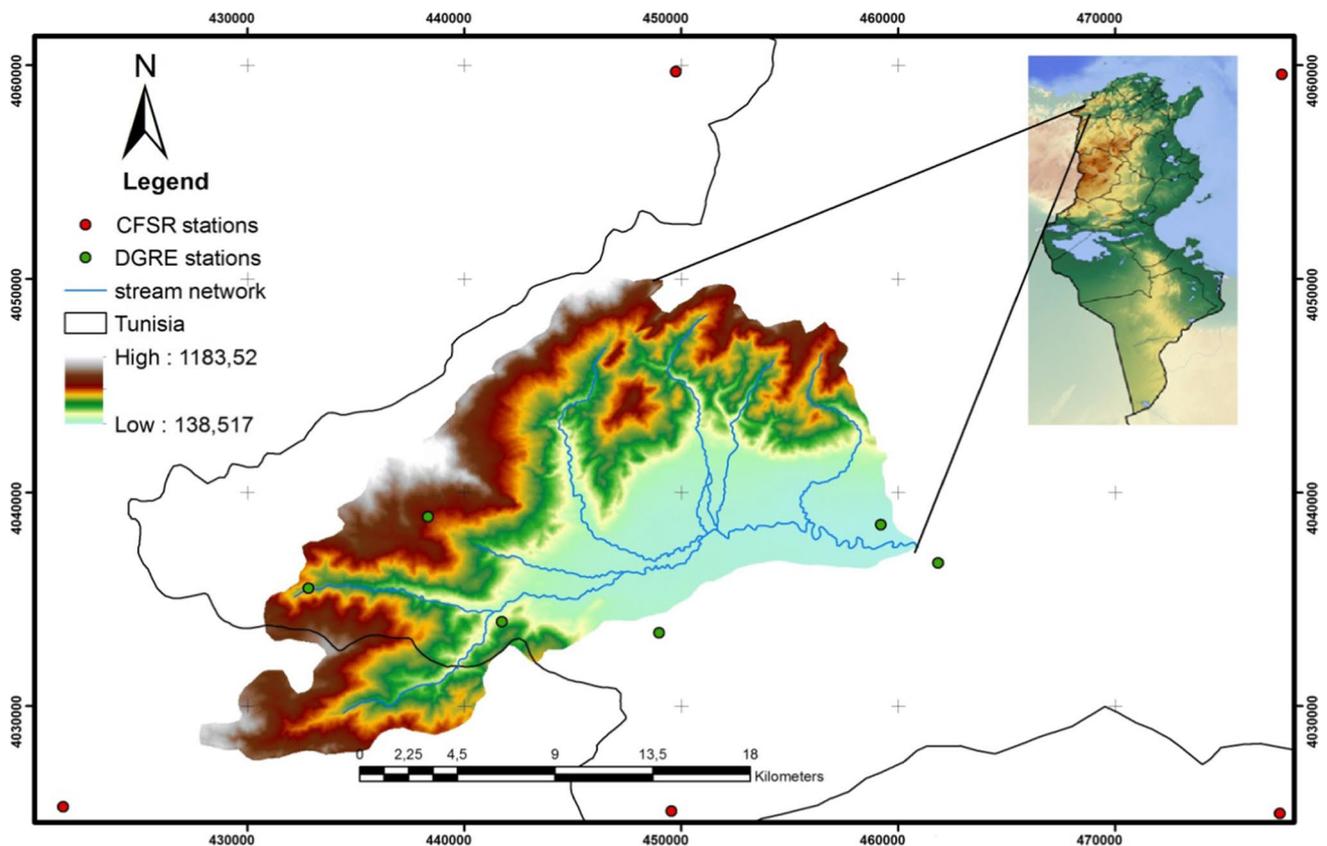


Fig. 3 Location of the Raghay catchment study site

From a climatic point of view, the zone of study presents a certain heterogeneity as it belongs to two different bioclimatic stages: a humid Mediterranean bioclimatic stage with moderate winter and a rainfall from 800 and 1200 mm/year, and a sub-humid bioclimatic stage which receives an annual rainfall varying between 450 and 700 mm/year (Ghorbel 1976). Rainfall is decreasing when going from the Western to the Eastern part in the catchment. The runoff generation process in the western part is expected to be different from the eastern part.

The Raghay catchment has a typical Mediterranean climate, with intense rainfall in the fall and winter seasons. Most of the rainfalls are recorded in winter (40 to 50% of the annual rainfall), while only 3% of the annual rainfall occurs in summer. The wet season starts from October to May, and the dry season starts from June to September. The floods mainly occur during very intense rainy events that may reach more than one hundred mm less than 24 h. In February 2012 for instance, a maximum flow of 1490 m³/s was recorded at the Ghardimaou hydrometric station, controlling a surface area of 1490 km² of the Medjerda basin. Simultaneously, 647 m³/s of the flows was measured in the Raghay discharge station, causing then catastrophic damages affecting downstream cities. The Raghay catchment is considered as a small sub-catchment in the Medjerda basin which has a huge contribution during floods. The great severity of the floods in this catchment is explained by the short rising times of the flow (less than 1 day), related directly to the concentration time (*t_c*) of the catchment and to the initial soil moisture.

The Raghay catchment is a representative study site on which the modelling procedure will be carried on.

Input data

For the model implementation, several inputs are necessary for combining hydro-climatic time-series data and cartographic data.

Cartographic input data

The topography of the study area is illustrated by the digital elevation model (DEM) provided by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). This DEM presents a resolution of 30 m for the mission of the 19 June 2014 with the Universal Transverse Mercator (UTM) coordinate system projection. This model with a fine grid mesh resolution allows outlining the elevation of any point at the Raghay catchment. This DEM was used to define the Raghay catchment and to identify the drainage and the slopes of the studied catchment.

Conventional time-series data

The daily observed rainfall and runoff data of the Raghay catchment were provided by the Tunisian General Directorate of Water Resources (DGRE). The rainfall was measured with six rain gauges (Table 2). These rain gauges have a weak percentage of gaps varying between a minimum gap ratio of 2% at the Raghay Supérieur rain gauge and a maximum gap ratio of almost 11% for the recorded period 1979–2008 at the Feija SM station. The rain gauge stations exhibit a large altimetry variation. The Chemtou Ferme station, for instance, situates at the minimum altitude of 172 m, while the Feija SM station situates at the maximum altitude of 730 m. In order to highlight the spatial variability for the Raghay catchment, we calculated several statistical parameters for the same durations, and then, we compared between the spatially distant rain gauges' values. The large spatial variability is confirmed with descriptive statistical parameters calculated separately for each station, such as the mean annual rainfalls, the standard deviation (σ) and the coefficient of variation (CV), permitting to highlight the importance of using several rain gauges in the Raghay (Table 3). The σ measures the temporal dispersion of the data around the annual mean rainfalls for each rain gauge. For the observed stations, σ varies between 124 mm (Chemtou Ferme station) and 265 mm (Feija SM) proportional to the mean annual rainfall values (417 mm and 967 mm for Chemtou Ferme station and Feija SM, respectively). The CV, which is the percentage of the ratio between σ and the mean annual rainfalls, is greater than 20% affirming the large spatial variability in the Raghay catchment for the conventional rainfall data. This result is consistent with studies emphasizing the large spatiotemporal variability of rainfall in Medjerda catchment (Gader et al. 2015). Based on the results reported for the rainfall observed data, it can be confirmed that the Raghay catchment presents a large spatial and temporal rainfall annual variability.

Runoff data were collected at the hydrometric discharge station 'Raghay Plaine.' The water levels were registered via a mechanical OTT 20™ stream gauge, between 1969 and 2008.

The minimum and maximum temperatures were measured at Jendouba meteorological station and controlled by the National Institute of Meteorology.

The common period of observation for the climatic and hydrometric data for the Raghay catchment is from 1990 to 2008, and it will be adopted for model implementation.

CFSR data

Data from a single observed climatic station for assessing PE is considered insufficient compared to the available spatially distributed observed rainfall data. To increase

Table 2 Detailed description of different hydro-climatic input data used in the SCS-SMA LR model implementation of the Raghay catchment

	Data type	Source	Coordinates UTM ^a (m)			Starting year	Period of observation	Gaps ^b (%)
			X	Y	Z			
Raghay Plaine	Daily runoff	DGRE ^c	460,817	4,037,710	145	1969	1990–2008	0
Chemtou Raouedet SM	Daily rainfall	DGRE	459,198	4,038,783	179	1966	1967–2012	2
Chemtou Ferme	Daily rainfall	DGRE	461,828	4,036,984	172	1969	1969–2012	4
Feija SM	Daily rainfall	DGRE	438,327	4,039,153	730	1889	1980–2010	11
Oued Mliz	Daily rainfall	DGRE	448,968	4,033,722	200	1973	1974–2012	5
Raghay Supérieur	Daily rainfall	DGRE	441,727	4,034,231	310	1977	1978–2012	2
Sraya Ecole	Daily rainfall	DGRE	432,827	4,035,805	600	1969	1975–2008	6
Jendouba climate station	Daily weather data ^d	INM ^e	449,166	4,033,752	143	1949	1985–2015	1
36481	Daily weather data	CFSR ^f	449,542	4,025,366	1041	1979	1979–2014	0
36484	Daily weather data	CFSR	477,574	4,025,248	860	1979	1979–2014	0
36488	Daily weather data	CFSR	477,664	4,059,878	229	1979	1979–2014	0
36784	Daily weather data	CFSR	421,824	4,060,206	374	1979	1979–2014	0
36788	Daily weather data	CFSR	449,744	4,059,996	306	1979	1979–2014	0

^aUTM: Universal Transverse Mercator coordinate system, Zone 32 Carthage North

^bThe percentage of the gaps is calculated for the common period of available climatic data: 29 years starting from 1979 to 2008

^cIn situ data obtained for free from the General Directorate of Water Resources, Ministry of Agriculture and Water Resources

^dWeather described data are: rainfall, maximum temperature, minimum temperature, wind speed, relative humidity, solar radiation

^eIn situ data obtained from National Institute of Meteorology in Tunisia

^fReanalysis data obtained from: <https://globalweather.tamu.edu/>

the spatial resolution of PE input data, global reanalysis data are useful data to represent the spatial distribution patterns of observed weather data (Zhang et al. 2013; Dile and Srinivasan 2014). We used the National Centers for Environmental Prediction's Climate Forecast System Reanalysis (CFSR) climate data to increase the spatial resolution of PE.

The CFSR is a worldwide, high resolution, coupled atmosphere–ocean–land surface–sea ice data system created in order to estimate weather data. The data are provided by the National Centers for Environmental Prediction (NCEP), which produces daily data for precipitation, maximum and minimum temperature, wind speed, relative humidity and solar radiation for the period of 1979–2014 with a resolution of 38 km (Tolera et al. 2018; Saha et al. 2010). All available conventional and satellite observations are included in the CFSR. Satellite observation data are integrated for radiance and are bias-corrected with a full-resolution 'spin-up' runs, taking into account green gas emissions. The CFSR data are completed by a spectral model including the parameterization of all major physical processes (Roth and Lemann 2016). After the calculation of the PE with the method described in "The modified runoff model with PE integration" section, we assessed the same description and statistical parameters as for the rainfall observed data. This allowed validating the use of these reanalysis data and evaluating the distributed or lumped implementation of the RR model (Tables 2, 3). We reveal a large spatial variability

of rainfall reanalysis data confirmed by high σ and CV for the mean annual reanalysis rainfall data.

For the PE data, the σ and CV are considered weak since spatial variability for temperature, which is the major driver for PE, is an order of magnitude smaller than for rainfall. For PE, CV is calculated as the percentage of the ratio between σ and the mean annual PE. When having a CV greater than 1%, PE calculated using CFSR data is variable compared to the mean values, validating the use of additional PE data derived from the CFSR data. We further compared between the available observed rainfalls recorded from six rain gauge stations and the rainfalls provided by the gridded CFSR reanalysis data, using ranked box plots and exceedance probability plots. We assessed the comparison in the distribution of the two types of data based on several evaluation criteria for each rank. The rainfall reanalysis data were only used in this paper for comparison with observed rainfalls in order to validate the use of this source of data. As observed rainfall data are spatially distributed, the model was implemented using only these data. In order to obtain spatially distributed PE, the CFSR data were applied for the Raghay catchment, which makes it possible to fill in the gaps in the PE to be integrated into the modelling implementations.

Table 3 Spatial variability description of observed rainfall and mixed PE (gridded data and observed data) of the Raghay catchment

	Rainfall (1979–2008)			Potential evapotranspiration (1990–2008) ^a		
	Mean annual rainfall ^b (mm)	SD ^c (σ)	Coefficient of variation ^d	Mean annual PE ^e (mm)	SD ^f (σ)	Coefficient of variation ^g
Chemtou Raouedet SM	446	124.96	27.99	–	–	–
Chemtou Ferme	417	133.55	31.95	–	–	–
Feija SM	967	265.35	27.42	–	–	–
Oued Mliz	497	142.46	28.63	–	–	–
Raghay Supérieur	508	163.57	32.14	–	–	–
Sraya Ecole	635	188.28	29.63	–	–	–
Jendouba climate station	381	97.81	25.67	1134	25.99	2.29
36481	857	193.99	22.62	1064	26.75	2.51
36484	723	166.95	23.09	1079	23.59	2.18
36488	636	154.97	24.35	1094	23.05	2.1
36784	826	177.26	21.45	1077	20.69	1.91
36788	746	164.86	22.09	1090	20.85	1.91

^aPE is calculated only for the common period of observation between rainfalls and runoffs

^bMean annual rainfalls (mm): the mean annual rainfalls calculated for the observed and CFSR rain gauges during the period of observation 1979–2008

^cSD (σ) is the temporal standard deviation for annual rainfalls calculated for the observed and CFSR rain gauges during the period of observation 1979–2008

^dCoefficient of variation (CV) is the temporal coefficient of variation for annual rainfalls calculated for the observed and CFSR rain gauges during the period of observation 1979–2008

^eMean annual PE (mm) is the mean annual PE calculated for the climatic stations, during 1990–2008

^fSD (σ) is the temporal standard deviation for annual PE calculated for the climatic stations, during 1990–2008

^gCoefficient of variation (CV) is the temporal coefficient of variation for annual PE calculated for the climatic stations, during 1990–2008

Results and discussion

Comparison between observed data (DGRE) and reanalysis gridded data (CFSR)

DGRE and CFSR rainfall for the Raghay catchment were compared on a statistical and graphical basis. Based on ranked box-plot (BP) representation of rainfall for the two types of data (Fig. 4), extreme high rainfall data having an appearance probability (AP) of 5% are similar, with a value of approximately 25 mm/day. However, there is a high difference for mean rainfall values between CFSR data and DGRE data, where CFSR data vary between 1 to 9 mm/day and DGRE data vary between 1 to 4 mm/day, with an AP of 50%. The presence of dry days in the Raghay catchment is validated in both cases of data based on weak and extremely weak AP in the BP having then the highest probability to occur in both cases (respectively, 70 and 95%).

The ranked exceedance probability (EP) was also assessed to give a detailed comparison of the two types of rainfall data (Fig. 5). For the high rainfall values, DGRE data and CFSR have almost the same AP (for EP with 5 and

10% of AP). The difference is clear for rainfall values with a medium AP (50%) in which DGRE data are lower than CFSR data. For dry days, when no rainfalls are recorded, both CFSR and DGRE data had a similar high and extreme high AP (respectively EP for 70% and 95% of AP). Thus, these results are consistent with the results obtained for the BP.

A large similarity was revealed between the extreme values of CFSR and DGRE data, with, respectively, a Nash–Sutcliffe (NS) higher than 85% and a coefficient of determination (R^2) higher than 90%, affirming that the two types of data have similar AP. These results demonstrate the appropriateness of the reanalysis data for flooding studies in Mediterranean poorly gauged catchments.

For EP with AP smaller than 10%, the two types of data are also similar, mainly for data higher than 10 mm/day, confirmed by an NS > 82% and R^2 > 90%. However, the difference is considerable for values smaller than 1 mm/day, confirmed by the AP of 50%. This questions the use of reanalysis data for drought studies in this region.

The estimated differences between the two rainfall types are probably due to the fact that the DGRE rainfall stations

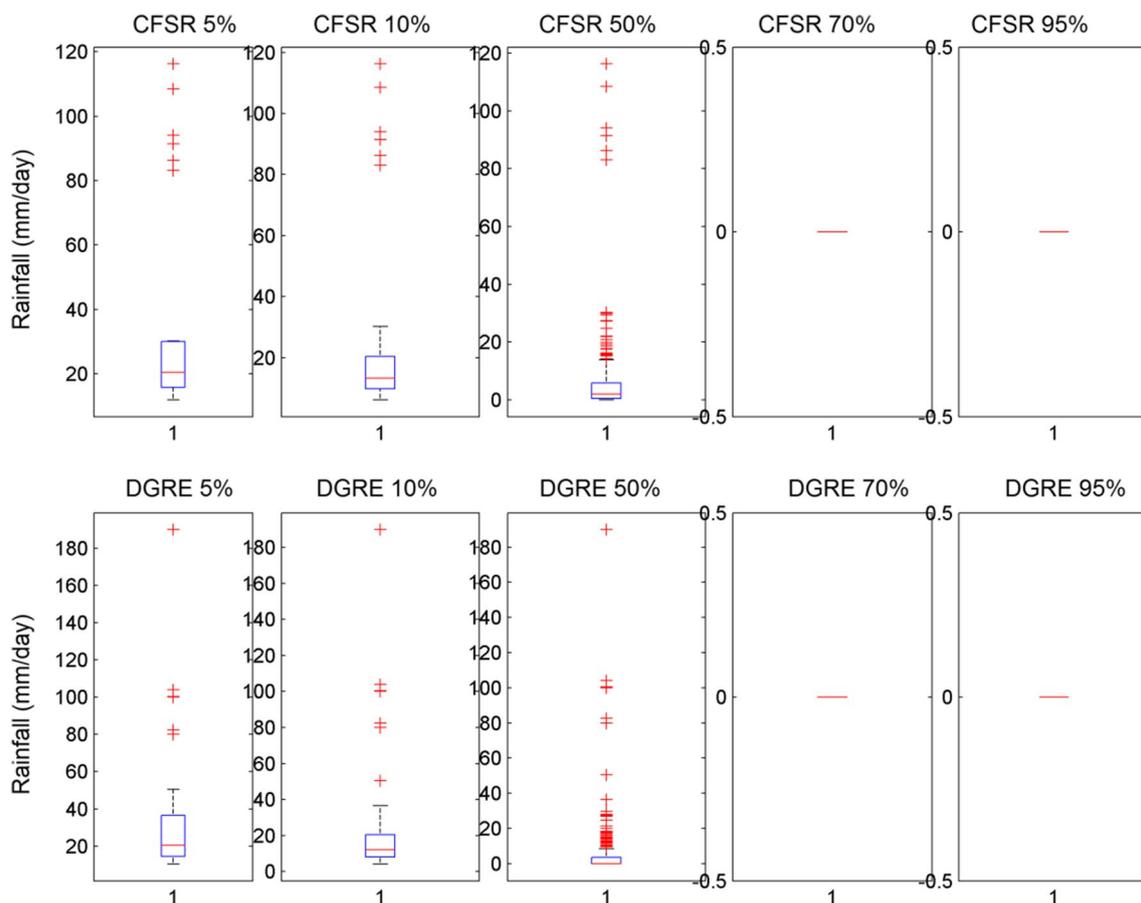


Fig. 4 Ranked box-plot comparison between DGRE data and CFSR data. Five percentage of the recorded data: extremely high recorded rainfall values with a lower probability of appearance (extremely rainy days); 10% of the data: high or important recorded rainfall val-

ues (with a probability of appearance 10%); 50% of the data: rainfall with a medium probability of appearance; 70% and 95%: weak and extremely weak rainfall values: they represent dry days with no recorded rainfall having the highest probability of appearance)

are located quite far away from the CFSR grid points (22 to 55 km), which can seriously affect CFSR efficiency in a mountainous area. Generally, reanalysis data overestimate rainfall data with about 20% compared to the observed ones in the study catchment. This difference is essentially caused by weak rainfall having values lesser than 1 mm/day. These results are coherent with the results obtained in other studies. According to Dile and Srinivasan (2014), there is no significant difference between the water balance simulation using conventional weather data and CFSR data, but the average annual rainfall from CFSR weather was higher than the annual rainfall from the conventional weather. Similarly, Fuka et al. (2014) affirmed that adding CFSR data to the suite of watershed modelling tools provides new opportunities for meeting the challenges of modelling in a watershed with scarce climate data.

To run the model, we used the PE derived from the six climatic stations available in the study site with PE calculated from climate data provided by the climatic station of Jendouba in addition to five CFSR climatic reanalysis stations.

For the rainfalls, we used the observed data provided from six rain gauges available for the Raghay catchment.

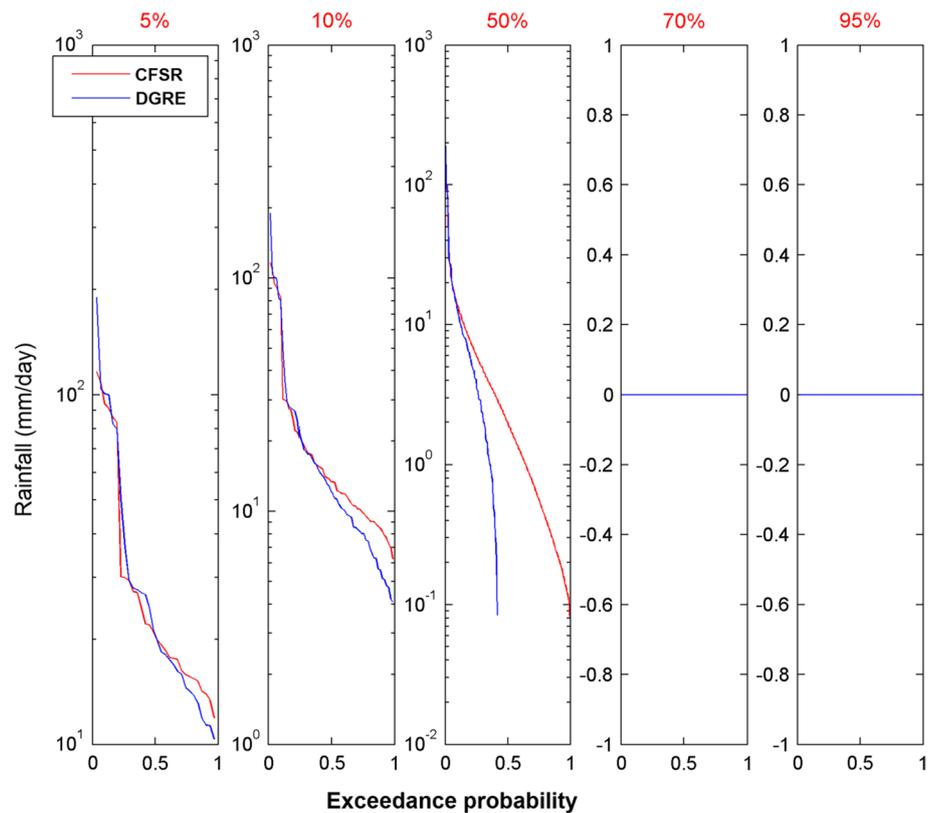
Rainfall–runoff modelling results comparison

The previously described model was implemented in the study area with two types of data schemes: firstly we run the RR model using distributed PE and rainfall, and secondly, using mean areal values of rainfall and PE, interpolated with the Thiessen interpolation method. We evaluated the two types of modelling implementations with a cross-validation split-sample test.

Impact of spatial distribution of PE and rainfall on model parameterization

For implementing the model, the parameters ds_2 ($ds_2 = 0$) and K_0 ($K_0 = 0.7$) were fixed. This last value was empirically deduced from the calibration of more than thirty catchments located in the south of France and having climate similarity

Fig. 5 Ranked exceedance probability comparison between DGRE data and CFSR data. Five percentage of the recorded data: extremely high recorded rainfall values with a lower probability of appearance (extremely rainy days); 10% of the data: high or important recorded rainfall values (with a probability of appearance 10%); 50% of the data: rainfall with a medium probability of appearance; 70% and 95%: weak and extremely weak rainfall values: they represent dry days with no recorded rainfall having the highest probability of appearance)



to the case study. The setting up of these two parameters is intended to make the calibration of the other parameters, such as S , ω , Sa/S and V_0 , more robust and in order to minimize the run-time for the RR model.

When we calibrated the RR model for the two phases of the split-sample test, we obtained completely different values for all the parameters for the distributed and lumped modes.

As for the first step of the cross-validation procedure, the value of S for the distributed mode is 122.4 mm. For the lumped case, the S value is around 87.5 mm. The difference is also noticeable for the Sa/S , where we obtain 0.34 and 0.56, respectively, for distributed and lumped modes. The ω parameter reaches 0.82 for the distributed mode and 0.53 for the lumped one.

The value of the S parameter is directly related to the capacity of storage for the soil reservoir which depends on soil characteristics (depth, heterogeneity, porosity, hydraulic conductivity, subsurface dip, etc.). A physical interpretation would be considered to highlight S as the product of the soil depth by the average porosity on a vertical profile (in the case of dominant processes of the contributing area type). This parameter partially explains the runoff rate for each mesh when running the model in the distributed mode. However, for the lumped mode, S was calculated for a unique single mesh. This can explain the differences in the parameter's values between the two types of model implementations.

The Sa/S emphasizes the proportion of initial losses in relation to the maximum capacity of the soil reservoir. Generally, the used value of this parameter is fixed to 0.3 (Michel et al. 2005). Nevertheless, we optimized this parameter in order to further highlight the differences in the parameterization for the two model implementations. This optimization featured that the distributed mode yields a parameter value that is approximately similar to the presumed value. It allows concluding that the soil reservoir for the lumped mode lost most of its capacity at the beginning of the calibrated period P1.

The parameter ω reflects the delayed flow that results from the emptying of the upper soil profiles, and it must be calibrated on observed floods. This parameter can be linked to the base flow obtained after the end of the event. According to the values calibrated for this parameter for the two model implementations, it can be presumed that the distributed mode has the biggest fraction of the soil reservoir emptying to contribute to runoff. This leads to a predicted base flow that is higher as compared to the predictions with the lumped mode.

Similar results were obtained for the calibration of the second step of the split-sample test. Yet, the difference is more important, especially for the S parameter varying between 213 mm for the distributed mode and 71 mm for the lumped one. Nevertheless, the difference in the LR model for the two types of model implementations is impossible

to quantify due to the difference in the used parameters. Regarding the lumped mode, the calibrated parameter is K_1 (mn), while for the distributed case it is V_0 (m/s). The differences may be partially related to the model identifiability (Gábor et al. 2017). The BLUE objective function for model optimization relies on the estimated initial parameters before running the model, which may cause errors.

For the distributed mode, there is no significant change in the calibrated V_0 for the two phases of cross-validation with a value of 1.12 m/s for the first step and 0.95 m/s for the second step. Still, the difference is important when calibrating the parameter K_1 for the lumped mode varying between 378 min for the first step of the cross-validation and 1642 min for the second step. This difference highlights variation in the lag, illustrating that the needed time for the produced runoff to reach the outlet for P2 is higher than for P1.

Impact of spatial distribution of PE and rainfall on flow and water amount simulations

The first calibration during the cross-validation process shows that the two model implementations (spatially distributed and lumped) tend to underestimate the peak flow for some flash floods. This underestimation is about 30% compared to the observed peak flow for the two cases and about 19% as a difference in water amount for the distributed mode, while for the lumped one, the difference between predicted and observed water amounts is about 24%. Based on the accuracy parameters comparison, we can highlight a considerable difference between the two model implementations (cases 1 and 2).

For the first case, with a spatially distributed input data, the model gives a NS of 0.61 and a PBIAS of 19.3, which is considered as satisfactory. For the second case, with lumped input data, the modelling is considered unsatisfactory with an NS about 0.57 and a PBIAS of 28.1. The simulated and observed discharges are presented in Fig. 6. The summary of calibrated parameters and accuracy criteria for the two modelling schemes of the first step of the cross-validation test is summarized in Table 4.

For case 1, when using distributed climate input data, the NS is considered acceptable, but the RSR is smaller than 0.7 and the PBIAS is almost unacceptable. Yet, all the accuracy parameters are unacceptable for case 2 when using lumped climatic data. This shows that the goodness of fit for the distributed model is better than for the lumped one. This interpretation is reinforced by the direct comparison between the differences in peak flow and in water amount. This is much higher for the lumped implementation, reaching 45% as for the difference in water amount and 34% for the difference in peak flow. However, for the spatially distributed mode, the difference in water amount and in peak flow is, respectively, 36 and 26%. This observation can be interpreted in terms of

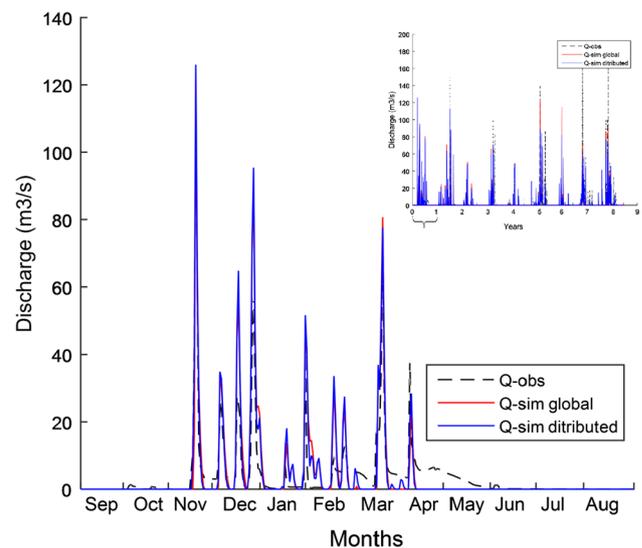


Fig. 6 Comparison between the simulation of the distributed model (case 1) and the global model (case 2) with the observed runoff for the calibration of P1 (first stage of cross-validation calibration): zoom in the first year (September 1990 to August 1991)

the difference between calibrated parameters, especially for the fraction causing the delayed runoff. This difference suggests that the obtained parameters for the distributed mode are more realistic and represent better the hydrological process in the Raghay catchment as compared to the lumped mode.

Another comparison is possible when using a graphical comparison matrix, the scattered plot matrix, comparing at the same time the observed and the simulated runoffs for the two cases. This comparison confirms that the distributed mode significantly yields better results, and it is considered as more efficient to predict runoff at the Raghay catchment for the validation of the first step of cross-validation procedure (Fig. 7).

For the second phase of the split-sample test, we obtained similar results as for the first phase referring to the evaluation criteria (Fig. 8). The distributed mode leads to lesser results than the one found for the calibration of the first step, even if the improvement is much smaller. We found different parameters for the calibration with lumped and distributed data using BLUE optimization. The slightly superior NS results obtained through the distributed mode are associated with a slight improvement of the other accuracy criteria for the calibration procedure as compared with the lumped mode implementation. These results can rely directly on the estimation of initial parameters for calibration, which can significantly improve the modelling optimization for the used objective function.

Our results show therefore that the distributed mode improves the model response for the Raghay catchment,

Table 4 Summary presentation of the calibrated parameters and accuracy criteria for the first step of the split-sample test of the two modelling implementations

	Calibration on P1		Validation on P2	
	Distributed mode (case 1)	Global mode (case 2)	Distributed mode (case 1)	Global mode (case 2)
<i>Calibrated parameters</i>				
S (mm)	122.4	87.5	–	–
ω (dimensionless)	0.82	0.53	–	–
ds_2 (1/day)	0	0	–	–
Sa/S (dimensionless)	0.34	0.56	–	–
V_0 (m/s)	1.12	1000	–	–
K_0 (dimensionless)	0.7	0	–	–
K_1 (mm)	0	378.8	–	–
<i>Evaluation criteria</i>				
NS	0.61	0.57	0.55	0.47
RMSE	7.36	7.62	10.43	11.45
RSR	0.48	0.59	0.61	0.74
PBIAS	19.27	28.1	26.7	46.8
EFF	0.15		0.38	

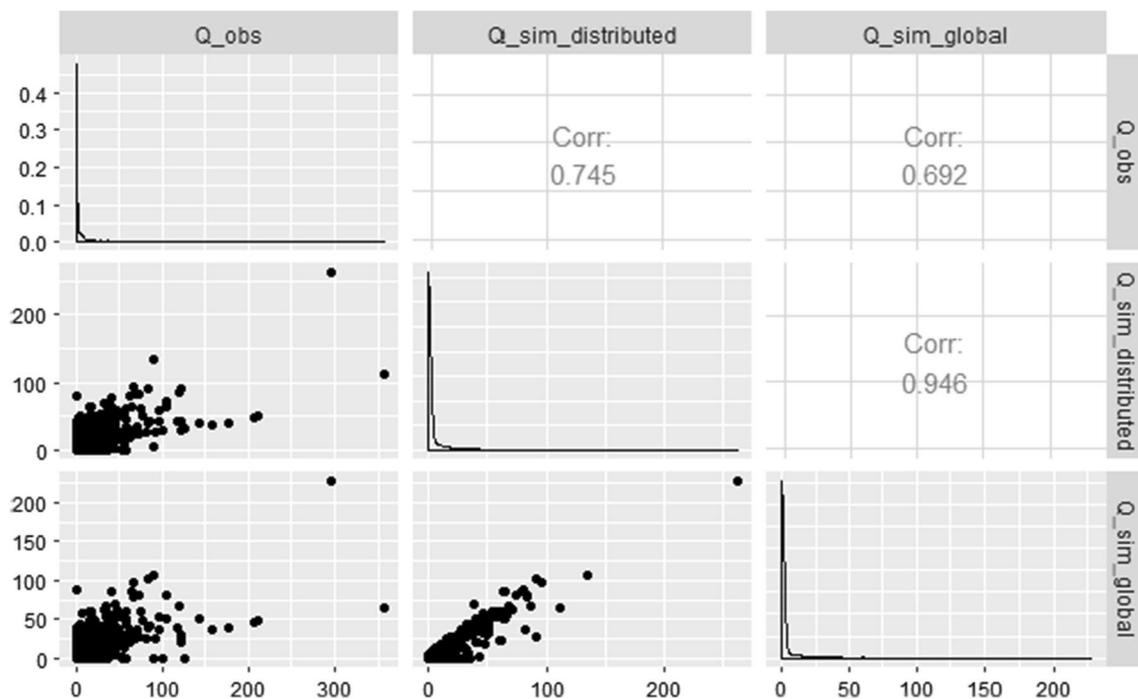


Fig. 7 Scattered plot matrix of the observed versus simulated discharge with global and distributed models in Raghay catchment for the validation of the first part of split-sample test

considering NS, where we obtained superior results for 7% leading to a good NS. Nevertheless, the NS obtained for the lumped mode is only considered as acceptable according to Moriasi et al. (2007). This difference in model responses can be linked to the fact that the model was calibrated and validated for the Raghay catchment for

periods P1 and P2 having a completely different hydro-climatic process.

For P1, going from September 1990 to August 1999, the High Valley of Medjerda basin went through a sequence of intense dry periods that marked the drought phenomena in the region (Bargaoui et al. 2008), with only nine important

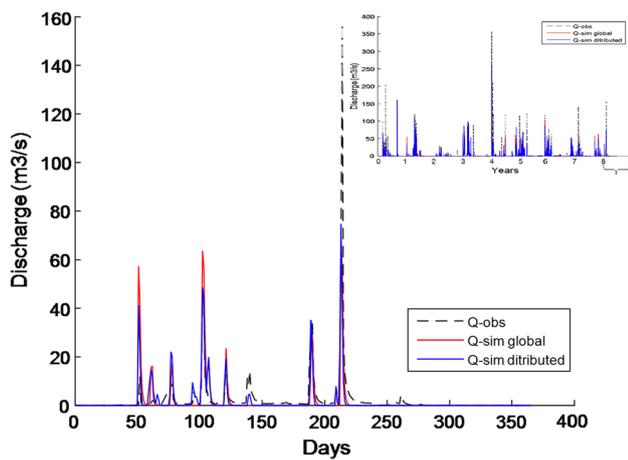


Fig. 8 Comparison between the simulation of the distributed model (case 1) and the global model (case 2) with the observed runoff for the calibration of P2 (second stage of cross-validation calibration): zoom in the last year (September 2007 to August 2008)

events mainly occurred in the autumn season. These events were recorded for the rain gauges located in the mountains and leading to important flash floods in the downstream parts, where no rainfall data were recorded. However, the data for P2 were pronounced with reciprocation of exceptional floods intermitted with intense dry periods. This period was marked by high spatial and temporal variability deduced from the hydro-climatic data, which yielded to note the impact of climate change in the southern Mediterranean region.

A more in-depth assessment is featured when examining the accuracy criteria for the two cases of model implementation (Table 5). For the spatially distributed model, all the accuracy parameters are considered good, with an NS equal to 0.61, an RSR below 0.6 and a PBIAS less than ± 0.15 . These evaluation criteria are only considered satisfactory ($NS < 0.6$, $PBIAS > \pm 0.15$, $0.6 < RSR < 0.7$) for the lumped model. However, when comparing the differences in peak flow and water amount, the difference is trivial (less than 5%). This result proves that the improvement of the model simulation given by the spatially distributed input is less significant in this phase. This result is more detailed based on the scattered plot matrix, proving that the regression between the two types of simulation is considered as non-significant for this part of the cross-validation procedure (Fig. 9).

By comparing the results obtained for the two phases of split-sample test for the two modelling schemes, we determined the important added value for the distributed model compared with the global one. We demonstrated that using distributed data produces better results for the Raghay catchment, on the basis of several statistical accuracy criteria and graphical comparison. This is consistent with the current state of the art in hydrological modelling. This result ties

Table 5 Summary presentation of the calibrated parameters and accuracy criteria for the second step of the split-sample test of the two modelling implementations

	Calibration on P2		Validation on P1	
	Distributed mode (case 1)	Global mode (case 2)	Distributed mode (case 1)	Global mode (case 2)
<i>Parameters</i>				
S (mm)	213.6	71.7	–	–
ω (dimensionless)	1.31	1.86	–	–
ds2 (1/day)	0	0	–	–
Sa/S (dimensionless)	0.18	0.28	–	–
V_0 (m/s)	0.95	1000	–	–
K_0 (dimensionless)	0.7	0	–	–
K_1 (mn)	0	1642.4	–	–
<i>Evaluation criteria</i>				
NS	0.58	0.55	0.61	0.54
RMSE	9.72	10.19	7.2	7.8
RSR	0.60	0.62	0.59	0.67
PBIAS	24.2	30.4	14.1	19.1
EFF	0.32		0.12	

well but is less pronounced with previous hydrological studies within the Mediterranean catchments using the same platform for hydrological modelling (Tramblay et al. 2011; Coustau et al. 2012). The improvement in our study suggests that ATHYS generally performs better for distributed models than lumped ones. One of the possible explanations for the difference between our study and previous studies with ATHYS in Mediterranean catchments is mainly because the models implemented for these previous cases were event-based with a finer time step and then the daily time step used in this research study. This kind of model implementations directly depends on the initial soil moisture for each event, an issue that does not appear in continuous RR models. Nevertheless, because of the lack of a finer observed data for the studied catchment, we decided to not investigate an infra-daily continuous modelling response.

Another possible explanation is that the hydrological process in Raghay catchment presents a significant fluctuation between dry and wet spells. This fluctuation is clearly noticeable for the part P2 of the data where the effect of climate change was highly prevalent.

As for the time step selected for this case of study, we opted to use the daily hydro-climatic data. To confirm this choice, we calculated the t_c through several formulas. We obtained a large difference between the obtained results, between 9 and 22 h, according to the used formula and to

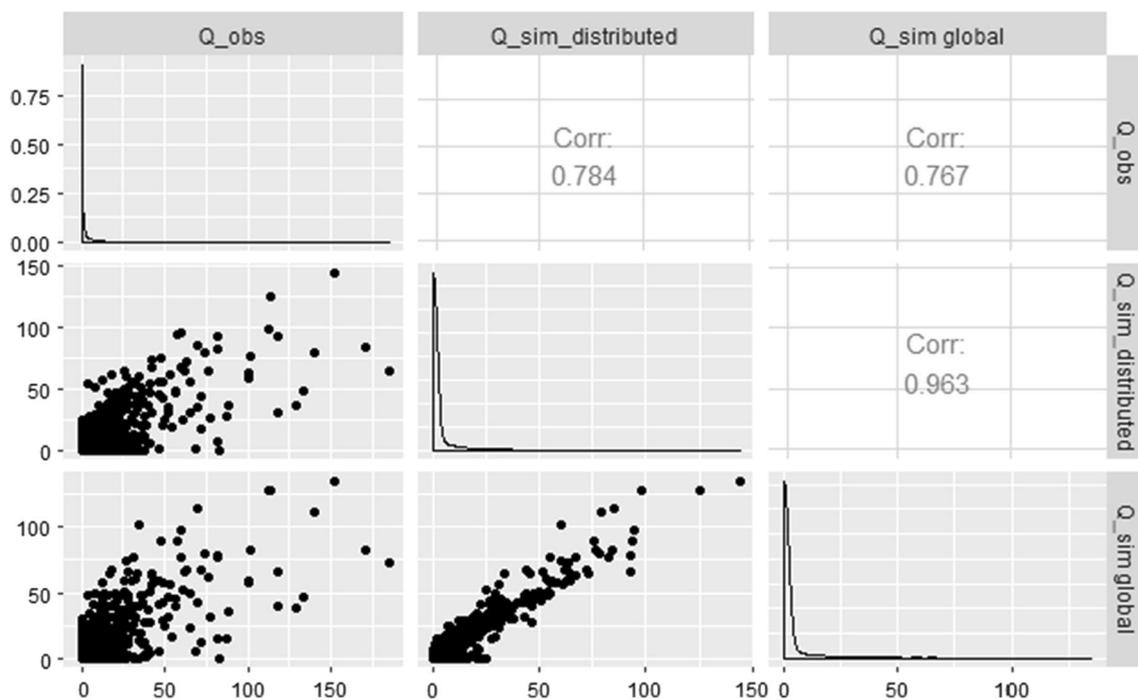


Fig. 9 Scattered plot matrix of the observed versus simulated discharge with global and distributed models in Raghay catchment for the validation of the second part of split-sample test

the estimated curve number. The obtained results are in concordance with the ones obtained by Fang et al. (2008), who used five different formulas to calculate the t_c . The obtained values of t_c are varying between 13 and 20 h for catchments with similar areas as for our case of study. The time of concentration of the catchment is thus less than 1 day. However, the purpose of this paper was to demonstrate the added value of the spatial distribution of rainfall and potential evapotranspiration (PE) in the prediction of the discharge for a small Mediterranean catchment. The fact that the time of concentration is less than 1 day when using daily rainfall data is not a problem, provided the models when compared with the same time step of the input rainfalls: spatially distributed or averaged rainfalls. Furthermore, the fact that the time of concentration is less than the time step of the input data makes that the model cannot retrieve accurate estimation of the flood dynamics. So it cannot be applied for hydrological applications such as flood forecasting, because the daily discharge could be much less than the instantaneous peak flow. But the model could be convenient for assessing water resources or hydrological budgets, which the daily discharges are convenient for.

We remark that for the two cases of modelling implementations, the RR model underestimates several important floods occurring in the autumn season. For some cases, the lumped mode overestimates the base flow, especially in the validation of the two contrasted phases. However, the

running time required by the lumped mode is much less than the time needed for the spatially distributed mode.

When comparing this research study with previous studies using the same time step, many studies claimed that the two model implementations were rather close. Brulebois et al. (2018) pointed out the same conclusion by obtaining rather close results when tested over contrasted climate periods, with minor higher stability for the semi-distributed model. The partial improvement of model efficiency criteria was also assessed by Lobligeois et al. (2014), where they found for a large sample of tested floods and catchment that results were associated with the catchment characteristics. As discussed, the Raghay catchment presents some improvements in terms of accuracy when using spatially distributed rainfall and PE data.

Impact of spatial distribution of PE and rainfall on the flow duration curve (FDC) at the Raghay catchment

Among the various indicators of runoff variability in watersheds, FDC of daily flow was assessed in this part to give a further comparison between distributed and lumped modes. The FDC represents the relationship between the magnitude and frequency of stream flows at the catchment discharge point when an adequate number of stream-flow observations are available (Botter et al. 2008). These observations may have different time step of entries;

however, mean daily stream-flow values are generally used (Ganora et al. 2009). Due to their ability to give a simple and comprehensive graphical view of the overall historical variability of stream flows over the catchment, from floods to low flows, empirical FDCs are widely used to represent the runoff regime in several water-related studies (Pugliese et al. 2014; Castellarin et al. 2004). We can virtually arrange the flow percentiles into three different ranks with different intervals of FDC: high flow for the segment between 0 and 10% of time flow equal or exceeded, median flows (segment between 10 and 50%) and low flows (segment between 50 and 100%).

The 9-year FDC obtained for the validation of the first step of the split-sample test highlights a large difference between the two modelling cases (Fig. 10). For the segment with values below 10%, there is a large resemblance for the two cases compared with the observed FDC, with a slight preference to the distributed mode. However, the difference is quite considerable for the two segments of median flows and low flows, respectively (10–50%) and (50–100%). Overall, the distributed mode gives better predicted FDC than the one predicted for the lumped mode for the first step of cross-validation.

Concerning the 9-year FDC obtained for the validation of the second step of the split-sample test, there is no significant difference between the two modelling cases having almost the same slope (Fig. 11). This is elucidated in “Impact of spatial distribution of PE and rainfall on flow and water amount simulations” section, proving that for the second validation, the difference between the two cases is not significant. Compared to the observed FDC, the difference between the two cases is based on medium and low flows, while the probability of high flows remains the same as for the first validation.

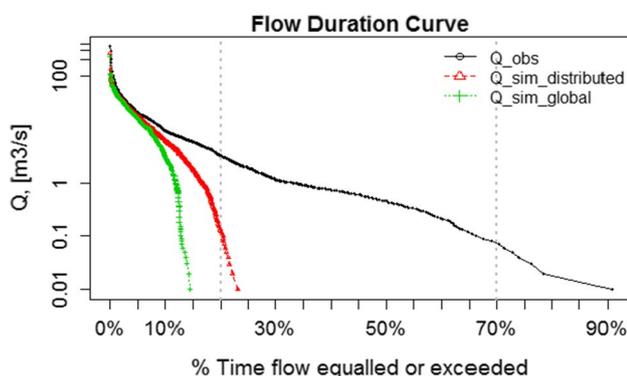


Fig. 10 Nine-year FDC simulation comparison between lumped and distributed modes at the Raghay for the validation of the first step of cross-validation test (validation on P2)

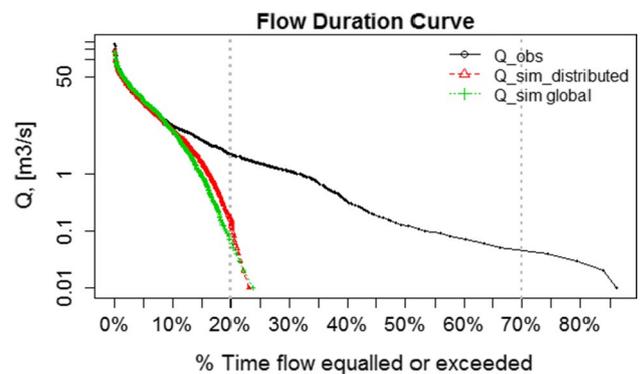


Fig. 11 Nine-year FDC simulation comparison between lumped and distributed modes at the Raghay for the validation of the second step of cross-validation test (validation on P1)

PE and rainfall spatial distribution global impact on flood simulation

In this section, the global impact is assessed based on the EFF, detailed in “Cross-validation procedure and performance evaluation” section. This criterion compares the simulated runoffs of distributed and lumped modes at the same time with the observed ones.

For the first validation, the EFF, about 0.38 which is much greater than 0, proves that the model simulation shows progress using spatially distributed input. However, this difference is less significant for the validation of the second phase of cross-validation, where the EFF is less than the one found in the validation of the first stage (0.12), yet, still higher than 0 confirming then the positive consequence for the use of distributed input data on model efficiency at the Raghay catchment. In order to additionally investigate the origins of the differences between the values of the EFF, we calculated the mean annual values, σ and CV, for annual input values of rainfalls and PE which are introduced to run the model, separately for the two periods of the data, P1 and P2. These mean annual temporal descriptive statistics were calculated for the contrasted durations of 9 years each and for all the spatially distant stations demonstrate a large spatial and temporal variability mainly noticed through the CV. This coefficient presents a noticeable difference during P2 for the rainfalls of the six rain gauges, varying between 17.6 and 34.21%, where during P1, it fluctuates between 23 and 34.7% (Table 6). For the PE, CV is around 2% for the two parts P1 and P2 and for all the climatic stations since the spatial and temporal variability of the mean annual temperature is not really consistent in the studied catchment. However, using spatially distributed data has slightly improved the modelling results. The SPI has been also used as an index to assess the climate variability for the different rain gauges in the Raghay over the contrasted

Table 6 Observed rainfall and mixed PE (gridded data and observed data) used to run the model in the Raghay catchment for the cross-validation periods P1 (1990–1999) and P2 (1999–2008)

	Mean annual (mm) ^c		SD (σ) ^d		Coefficient of variation ^e	
	P1 ^a	P2 ^b	P1	P2	P1	P2
<i>Rainfall</i>						
Chemtou Raouedet SM	472.99	453.27	145.82	148.55	30.83	32.77
Chemtou Ferme	466.69	452.35	162.09	136.31	34.73	30.13
Feija SM	930.92	844.37	221.74	148.58	23.82	17.60
Oued Mliz	505.04	525.69	160.04	170.94	31.69	32.52
Raghay Supérieur	522.22	557.89	174.92	190.86	33.50	34.21
Sraya Ecole	654.22	540.61	160.25	181.54	24.49	33.58
<i>Potential evapotranspiration</i>						
Jendouba climate station	1135.18	1133.00	30.1	22.95	2.65	2.03
36481	1052.10	1077.78	25.25	22.64	2.40	2.10
36484	1070.01	1088.14	24.76	19.62	2.31	1.80
36488	1085.93	1104.02	24.31	18.84	2.24	1.71
36784	1069.93	1085.15	21.17	18.21	1.98	1.68
36788	1083.23	1097.92	21.63	18.31	2.00	1.67

^aP1: the part of the data with a duration of 9 years, going from September 1990 to August 1999

^bP2: the part of the data with a duration of 9 years, going from September 1999 until August 2008

^cMean annual (mm): the mean annual rainfalls and PE calculated, respectively, for the observed rain gauges and the climatic stations, during P1 and P2 separately

^dSD (σ) is the temporal standard deviation for annual rainfalls and PE calculated, respectively, for the observed rain gauges and for the climatic stations, during P1 and P2 separately

^eCoefficient of variation (CV) is the temporal coefficient of variation for annual rainfalls and PE calculated, respectively, for the observed rain gauges and for the climatic stations, during P1 and P2 separately

periods. During P1, the SPI shows very similar variations for all the rain gauges with only small differences in terms of drought amplitude, varying between severe humidity and severe drought. Nevertheless, during P2, the variability in the SPI calculated between the different rain gauges is really spectacular fluctuating between extreme drought and extreme humidity (Fig. 12). The variability in the SPI during P2 reflects the large variability in the rainfalls. This variability can be explained by different possible origins mainly through the marked climate change impacts in the two last decades where its effects were strengthened by the important difference in altitudes and slopes over the Raghay.

Conclusions

In this paper, we assessed the performance of the SCS–SMA LR model with the PE integration, from ATHYS platform. We implemented the conceptual daily continuous model in the Raghay catchment with two different types of implementations: once in a fully spatially distributed mode (case 1) by using spatially distributed rainfall and PE, and once in a lumped mode using Thiessen-interpolated PE and rainfall (case 2). The used model requires a few numbers of

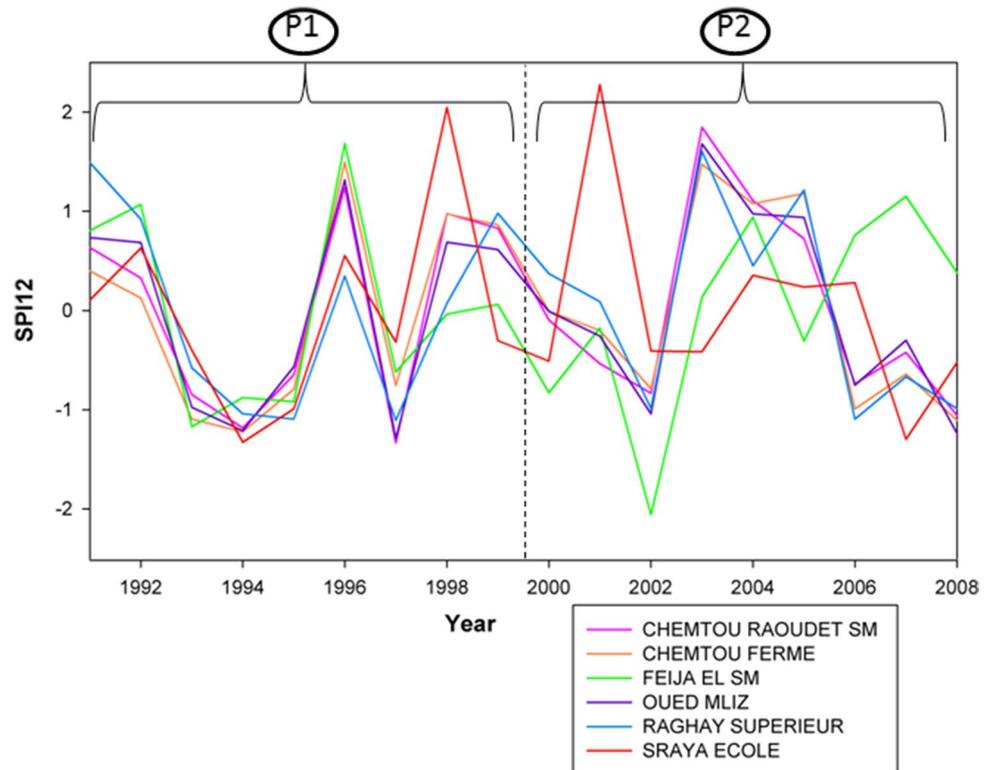
parameters to be calibrated letting for simple and efficient modelling phases.

In order to reinforce and to fill in the gaps for the PE data required to run the model in the spatially distributed mode for the studied catchment, we integrated CFSR data in addition to the observed PE provided from one observed climatic station, and the observed rainfalls recorded on the level of six rain gauges.

A cross-validation procedure was implemented on the two equal parts of the hydro-climatic data (P1 and P2) for the two modelling cases. The main conclusions that can be drawn for the validation of the first and second phases of the split-sample test is that the RR model presents a slight preference to the spatially distributed implementations, on the basis of different model accuracy criteria (NS, RSR, RMSE, PBIAS and EFF) which are performing less good in the lumped mode. However, this improvement is not important, especially for the NS accuracy criterion, which is presenting a difference that does not exceed 7%, for the calibration and for the validation of the part P2 of the data.

To further investigate the two modelling cases response, we integrated graphical and statistical comparisons, such as the scattered plot matrix showing the difference between observed and simulated runoffs for the two cases, and the 9-year duration FDCs for the two cases of model

Fig. 12 Comparison between the SPI 12 index calculated for the six rain gauges during P1 and P2 in the Raghay catchment



implementations, which were leading to prove the same results found on the basis of the goodness-of-fit criteria.

These findings confirm that according to the validation of the second step of the cross-validation procedure, the improvement of the model simulation given by the spatially distributed input is less significant in this phase, with an $EFF = 0.12$ as compared to $EFF = 0.38$ for the validation of the first stage. Nevertheless, these criteria cast the need to use of spatially distributed data because it is higher than 0. Statistical descriptive indexes were calculated for the input data over the contrasted periods, P1 and P2, in an attempt to explain by climate reasoning the obtained difference in the EFF through the cross-validation procedure. These statistical coefficients reinforced the large spatial and temporal variability in the rainfalls for the second part of the data, P2. The SPI also confirmed this assumption when giving a large variability in its values for the different rain gauges as for the same period, P2.

Based on the analyses and the results reported in this study, it can be concluded that spatially distributed rainfall and PE increase the efficiency of the model in the Raghay catchment. However, this improvement is less important for the second half part of the data, P2, used in this comparison study. The interest of the spatial distribution of rainfalls and PE depends on the spatial and temporal variability of climate data over the catchment. This spatial and temporal variability is mainly controlled by the large variability in the altitudes and the slopes over

the catchment coupled with the present impact of climate change assessed in recent decades.

Future research should consider the potential effects of other types of models presented in the ATHYS platform more carefully. For example, the main issue when trying to compare accuracy between models having a different number of input parameters (i.e. conceptual based versus physically based model implementations in the ATHYS platform), we have to refer to other accuracy criteria linking between the number of model parameters and the goodness of fit, in order to evaluate the model efficiency. The goodness-of-fit criteria give only a partial understanding of model performance that should be completed with an additional parsimony analysis. This is an issue for future research to explore.

It could be also with great importance to apply the same methodology for other catchments situated in the Medjerda basin in Tunisia. However, the availability of distributed climate data needed to calculate PE remains the main challenge for Tunisian catchments. Further simplifications may be possible by using a unique PE for each catchment, based on the assumption that the NS rarely exceeds 1% when using spatially distributed PE.

We believed that apart from looking for model responses in the actual conditions, it would be an objective for future studies to investigate how different these two models will produce the projected hydrological process for the studied catchment under different climate change scenarios' projections. This will aim to replicate the differences in the results

of the two model schemes regardless of their complexity. Future investigations are necessary to summarize this kind of comparisons.

Acknowledgements This research was realized at the Sustainable Management of Water and Soil Resources (UR17AGR03). The authors received funding from the project ‘Adaptation of water resources management in the Medjerda watershed to the challenges of climate change’ through the project 11—axis 2 of the Wallonia Brussels International Grant No. (Aso/CFo/Tunisie/15.1218/cf) and Tunisia Joint Commission 2016–2018. Moreover, part of this research study was accomplished in the Laboratory of HydroSciences in Montpellier with the funding of mobility program provided by the Ministry of Higher Education and Scientific Research in Tunisia. Authors also would like to thank the Tunisian General Directorate of Water Resources (DGRE) for providing conventional data for the study area. Finally, the authors would like to express their gratitude for the anonymous reviewers of this paper.

Compliance with ethical standards

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

References

- Ajami NK, Gupta H, Wagener T, Sorooshian S (2004) Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *J Hydrol* 298:112–135. <https://doi.org/10.1016/J.JHYDROL.2004.03.033>
- Andréassian V, Oddos A, Michel C, Anctil F, Perrin C, Loumagne C (2004) Impact of spatial aggregation of inputs and parameters on the efficiency of rainfall–runoff models: a theoretical study using chimera watersheds. *Water Resour Res*. <https://doi.org/10.1029/2003WR002854>
- Apip Sayama T, Tachikawa Y, Takara K (2012) Spatial lumping of a distributed rainfall–sediment–runoff model and its effective lumping scale. *Hydrol Process* 26:855–871. <https://doi.org/10.1002/hyp.8300>
- Bao H, Wang L, Zhang K, Li Z, Wang L (2017) Application of a developed distributed hydrological model based on the mixed runoff generation model and 2D kinematic wave flow routing model for better flood forecasting. *Atmos Sci Lett* 18:284–293. <https://doi.org/10.1002/asl.754>
- Bargaoui Z, Hamouda D, Houcine A (2008) Modélisation pluie-débit et classification hydroclimatique. *Rev des Sci l’eau* 21:233–245. <https://doi.org/10.7202/018468ar>
- Bentura PLF, Michel C (1997) Flood routing in a wide channel with a quadratic lag-and-route method. *Hydrol Sci J* 42(2):169–189
- Botter G, Zanardo S, Porporato A, Rodriguez-Iturbe I, Rinaldo A (2008) Ecohydrological model of flow duration curves and annual minima. *Water Resour Res*. <https://doi.org/10.1029/2008WR006814>
- Bouvier C, Marchandise A, Brunet P, Crespy A (2008) Un modèle pluie-débit distribué événementiel parcimonieux pour la prédétermination et la prévision des crues éclair en zone méditerranéenne. Application au bassin du Gardon d’Anduze. IWRA congress 2008, p 10. https://www.iwra.org/congress/2008/resource/authors/abs811_article.pdf. Accessed 23 June 2018
- Bouvier C, Crespy A, L’Aour-Dufour A, Crès FN, Desclaux F, Marchandise A (2013) Distributed hydrological modeling—the ATHYS platform. Modeling software. Wiley, Hoboken, pp 83–100
- Breuer L, Huisman JA, Willems P, Bormann H, Bronstert A, Croke B, Frede H, Gräff T, Hubrechts L, Jakeman A, Kite G, Lanini J, Leavesley G, Lettenmaier D, Lindström G, Seibert J, Sivapalan M, Viney N (2009) Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: model intercomparison with current land use. *Adv Water Resour* 32:129–146. <https://doi.org/10.1016/j.advwatres.2008.10.003>
- Brocca L, Melone F, Moramarco T, Wagner W, Naeimi V, Bartalis Z, Hasenauer S (2010) Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Earth Syst Sci* 145194:1881–1893. <https://doi.org/10.5194/hess-14-1881-2010>
- Brulebois E, Ubertosi M, Castel T, Richard Y, Sauvage S, Sanchez Perez JM, Le Moine N, Amiotte-Suchet P (2018) Robustness and performance of semi-distributed (SWAT) and global (GR4J) hydrological models throughout an observed climatic shift over contrasted French watersheds. *Open Water J* 5(1):41–56. <https://scholarsarchive.byu.edu/openwater/vol5/iss1/4>
- Castellari A, Galeati G, Brandimarte L, Montanari A, Brath A (2004) Regional flow-duration curves: reliability for ungauged basins. *Adv Water Resour*. <https://doi.org/10.1016/j.advwatres.2004.08.005>
- Coustaou M, Bouvier C, Borrell-Estupina V, Jourde H (2012) Flood modelling with a distributed event-based parsimonious rainfall–runoff model: case of the karstic Lez river catchment. *Nat Hazards Earth Syst Sci* 12:1119–1133. <https://doi.org/10.5194/nhess-12-1119-2012>
- Dakhlaoui H, Ruelland D, Tramblay Y, Bargaoui Z (2017) Evaluating the robustness of conceptual rainfall–runoff models under climate variability in northern Tunisia. *J Hydrol* 550:201–217. <https://doi.org/10.1016/J.JHYDROL.2017.04.032>
- Dile YT, Srinivasan R (2014) Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: an application in the blue Nile river basin. *J Am Water Resour Assoc* 50:1226–1241. <https://doi.org/10.1111/jawr.12182>
- Duan Q, Gupta HV, Sorooshian S, Rousseau A, Turcotte R (2003) Calibration of watershed models. American Geophysical Union, Washington
- Fang X, Thompson DB, Cleveland TG et al (2008) Time of concentration estimated using watershed parameters determined by automated and manual methods. *J Irrig Drain Eng* 134:202–211. [https://doi.org/10.1061/\(ASCE\)0733-9437\(2008\)134:2\(202\)](https://doi.org/10.1061/(ASCE)0733-9437(2008)134:2(202))
- Fuka DR, Walter MT, Macalister C, Degaetano A, Steenhuis T, Easton Z (2014) Using the climate forecast system reanalysis as weather input data for watershed models. *Hydrol Process*. <https://doi.org/10.1002/hyp.10073>
- Gábor A, Villaverde AF, Banga JR (2017) Parameter identifiability analysis and visualization in large-scale kinetic models of bio-systems. *BMC Syst Biol* 11:54. <https://doi.org/10.1186/s12918-017-0428-y>
- Gader K, Gara A, Slimani M, Mahjoub MR (2015) Study of spatio-temporal variability of the maximum daily rainfall. In: 2015 6th International conference on modeling, simulation, and applied optimization (ICMSAO). IEEE, pp 1–6
- Ganora D, Claps P, Laio F, Viglione A (2009) An approach to estimate nonparametric flow duration curves in ungauged basins. *Water Resour Res*. <https://doi.org/10.1029/2008WR007472>
- Gara A, Gader K, Bargaoui M, Mahjoub MR (2015) Assessment of the hydrological response of the watershed through a distributed physically-based modeling for extreme events: application in the Raghay catchment (Medjerda) (Northern Tunisia). In: 6th international conference on modeling, simulation, and applied optimization, ICMSAO 2015. IEEE, pp 1–6
- Ghorbel A (1976) Etude hydrologique de l’oued Rarai. 26
- Hawkins RH (1993) Asymptotic determination of runoff curve numbers from data. *J Irrig Drain Eng* 119:334–345. [https://doi.org/10.1061/\(ASCE\)0733-9437\(1993\)119:2\(334\)](https://doi.org/10.1061/(ASCE)0733-9437(1993)119:2(334))

- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423. <https://doi.org/10.2307/2529430>
- Ibrahim B, Wisser D, Barry B, Fowe T, Aduna A (2015) Hydrological predictions for small ungauged watersheds in the Sudanian zone of the Volta basin in West Africa. *J Hydrol Reg Stud* 4:386–397. <https://doi.org/10.1016/J.EJRH.2015.07.007>
- Khakbaz B, Imam B, Hsu K, Sorooshian S (2012) From lumped to distributed via semi-distributed: calibration strategies for semi-distributed hydrologic models. *J Hydrol* 418–419:61–77. <https://doi.org/10.1016/J.JHYDROL.2009.02.021>
- Klemeš V (1986) Operational testing of hydrological simulation models. *Hydrol Sci J*. <https://doi.org/10.1080/02626668609491024>
- Koren V, Zhang Z, Zhang Y, Reed SM, Cui Z, Morea F, Cosgrove BA, Mizukami N, Anderson EA (2012) Results of the DMIP 2 Oklahoma experiments. *J Hydrol* 418–419:17–48. <https://doi.org/10.1016/J.JHYDROL.2011.08.056>
- Lewis D, Singer MJ, Tate KW (2000) Applicability of SCS curve number method for a California oak woodlands watershed. *J Soil Water Conserv* 55:226–230
- Liu X, Li J (2008) Application of SCS model in estimation of runoff from small watershed in Loess Plateau of China. *Chin Geogr Sci* 18:235. <https://doi.org/10.1007/s11769-008-0235-x>
- Lobligeois F, Andréassian V, Perrin C, Tabary P, Loumagne C (2014) When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrol Earth Syst Sci* 18:575–594. <https://doi.org/10.5194/hess-18-575-2014>
- Ludwig R, Roson R, Zografos C, Kallis G (2011) Towards an interdisciplinary research agenda on climate change, water and security in Southern Europe and neighboring countries. *Environ Sci Policy*. <https://doi.org/10.1016/j.envsci.2011.04.003>
- Mckee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. In: Eighth conference on applied climatology, pp 179–184
- Michel C, Andréassian V, Perrin C (2005) Soil conservation service curve number method: How to mend a wrong soil moisture accounting procedure? *Water Resour Res*. <https://doi.org/10.1029/2004wr003191>
- Mishra SK, Singh VP, Sansalone JJ, Aravamathan V (2003) A Modified SCS-CN Method: characterization and Testing. *Water Resour Manag* 17:37–68. <https://doi.org/10.1023/A:1023099005944>
- Moriasi DN, Arnold JG, Van Liew MW, Bingner R, Harmel R, Veith T (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans ASABE* 50:885–900
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—a discussion of principles. *J Hydrol* 10:282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Oudin L, Hervieu F, Michel C, Perrin C, Andréassian V, Anctil F, Loumagne C (2005) Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2—towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *J Hydrol* 303:290–306. <https://doi.org/10.1016/J.JHYDROL.2004.08.026>
- Oudin L, Andréassian V, Perrin C, Michel C, Le Moine N (2008) Spatial proximity, physical similarity, regression and ungauged catchments: a comparison of regionalization approaches based on 913 French catchments. *Water Resour Res*. <https://doi.org/10.1029/2007WR006240>
- Pugliese A, Castellarin A, Brath A (2014) Geostatistical prediction of flow-duration curves in an index-flow framework. *Hydrol Earth Syst Sci* 18:3801–3816. <https://doi.org/10.5194/hess-18-3801-2014>
- Roth V, Lemann T (2016) Comparing CFSR and conventional weather data for discharge and soil loss modelling with SWAT in small catchments in the Ethiopian Highlands. *Hydrol Earth Syst Sci* 20:921–934. <https://doi.org/10.5194/hess-20-921-2016>
- Saha S, Moorthi S, Pan HL, Wu X, Wang J, Nadiga S, Tripp P, Kistler R, Woollen J, Behringer D, Liu H, Stokes D, Grumbine R, Gayno G, Wang J, Hou Y, Chuang H, Juang H, Sela J, Iredell M, Treadon R, Kleist D, Van Delst P, Keyser D, Derber J, Ek M, Meng J, Wei H, Yang R, Lord S, van den Dool H, Kumar A, Wang W, Long C, Chelliah M, Xue Y, Huang B, Schemm J, Ebisuzaki W, Lin R, Xie P, Chen M, Zhou S, Higgins W, Zou C, Liu Q, Chen Y, Han Y, Cucurull L, Reynolds R, Rutledge G, Goldberg M (2010) The NCEP climate forecast system reanalysis. *Bull Am Meteorol Soc* 91:1015–1058. <https://doi.org/10.1175/2010BAMS3001.1>
- Sangati M, Borga M (2009) Influence of rainfall spatial resolution on flash flood modelling. *Nat Hazards Earth Syst Sci* 9:575–584
- Schul J, Abbaspour KC, Yang H, Srinivasan R, Zehnder A (2008) Modeling blue and green water availability in Africa. *Water Resour Res*. <https://doi.org/10.1029/2007WR006609>
- Sellami H, La Jeunesse I, Benabdallah S, Vanclooster M (2013) Parameter and rating curve uncertainty propagation analysis of the SWAT model for two small Mediterranean catchments. *Hydrol Sci J*. <https://doi.org/10.1080/02626667.2013.837222>
- Sherman J, Morrison WJ (1950) Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Ann Math Stat* 21:124–127
- Singh J, Knapp HV, Demissie M (2005) Hydrologic modeling of the Iroquois river watershed using HSPF and SWAT. *J Am Water Resour Assoc* 41(2):343–360. <https://doi.org/10.1111/j.1752-1688.2005.tb03740.x>
- Tegegne G, Park DK, Kim Y-O (2017) Comparison of hydrological models for the assessment of water resources in a data-scarce region, the Upper Blue Nile River Basin. *J Hydrol Reg Stud* 14:49–66. <https://doi.org/10.1016/J.EJRH.2017.10.002>
- Tolera M, Chung I-M, Chang S (2018) Evaluation of the climate forecast system reanalysis weather data for watershed modeling in Upper Awash Basin, Ethiopia. *Water* 10:725. <https://doi.org/10.3390/w10060725>
- Tramblay Y, Bouvier C, Ayrat P-A, Marchandise A (2011) Impact of rainfall spatial distribution on rainfall–runoff modelling efficiency and initial soil moisture conditions estimation. *Nat Hazards Earth Syst Sci* 11:157–170. <https://doi.org/10.5194/nhess-11-157-2011>
- Vansteenkiste T, Tavakoli M, Van Steenberghe N, De Smedt F, Bate-laan O, Pereira F, Willems P (2014) Intercomparison of five lumped and distributed models for catchment runoff and extreme flow simulation. *J Hydrol* 511:335–349. <https://doi.org/10.1016/J.JHYDROL.2014.01.050>
- Wöhling T, Samaniego L, Kumar R (2013) Evaluating multiple performance criteria to calibrate the distributed hydrological model of the upper Neckar catchment. *Environ Earth Sci* 69:453–468. <https://doi.org/10.1007/s12665-013-2306-2>
- Zhang Q, Heiner K, Holmgren K (2013) How well do reanalyses represent the southern African precipitation? *Clim Dyn*. <https://doi.org/10.1007/s00382-012-1423-z>
- Zkhiri W, Tramblay Y, Hanich L et al (2019) Spatiotemporal characterization of current and future droughts in the High Atlas basins (Morocco). *Theor Appl Climatol* 135:593–605. <https://doi.org/10.1007/s00704-018-2388-6>



Estimation of groundwater potential using GIS modeling in Kohistan region Jamshoro district, Southern Indus basin, Sindh, Pakistan (a case study)

Aneela Memon^{1,2} · Kamran Ansari¹ · Abdul Ghani Soomro^{1,2} · Muhammad Afzal Jamali³ · Babar Naeem¹ · Arshad Ashraf²

Received: 24 May 2019 / Accepted: 25 October 2019 / Published online: 8 November 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

The subsurface groundwater potential was calculated through the Geographic Information System (GIS) tool in district Jamshoro, Sindh, Pakistan; research was carried out in the vulnerable and arid climatic regions of the study area, which experienced unpredictable droughts and extreme floods. The primary recharging source of the aquifer is rainfall only in monsoon season from July to September. The current situation of water resources is critical in the region; in the present research work, two principal groundwater aquifers were encountered present in lithology of Laki formation mostly in sand and gravel lithology, which are shallow and deep aquifers. The GIS has been employed as a preservative tool to estimate groundwater potential. The estimated groundwater potential was 29,330.22 million cubic meter (MCM) under the monitoring area of 3207.25 square kilometers (km²); thus, the volume of groundwater was 9.14 MCM/km² in the selected cluster zone of the study area. It is realized that using a GIS tool combats the susceptible situation and shows authenticated results of groundwater potential.

Keywords Groundwater potential · Lithology · Stratigraphy · GIS · Water resources

Introduction

Water is one of the most vital natural sources for survival on the earth and a precious commodity for human development. Fresh water in abundant volume is readily available in the subsurface of the earth. A gradual increase in the population leads to more water demand for agriculture and economic activities in the last three decades (Murasingh and Jha 2013). However, groundwater is perceived as one of the most significant and reliable fresh sources of water globally

(Todd and Mays 2004), which contributes about 0.61% in the world (Fetter 1980).

Pakistan faces water scarcity problem in the last decades due to increased water demand. The condition of the country becomes worst by the influence of many factors such as the growing trend of population, climate change, unauthorized, and unjustified use of water resources (Sadaf et al. 2019). In the present scenario, the water availability is less than 1000 m³/capita/year, which is projected to decrease up to 837 m³/capita/year by December 2025 (Gardner-Outlaw and Engelman 2019). The increasing population rate, boosting urbanization, and industrializations would be more challenging, which lead to the land degradation, demand of more water resources, and threaten food security. The current scenario of the country is more problematic and complicated. Thus, over-exploiting groundwater leads to negative impacts on water resources. The need for water resources management for sustainable development is the most important, especially in the arid areas of Pakistan, which depends on precipitation.

Pakistan depends on the groundwater as a secondary source of irrigation; 90% of total cereal are produced under

✉ Aneela Memon
aneelahameem@gmail.com

¹ United States-Pakistan Center for Advanced Studies in Water (USPCAS-W), Mehran University of Engineering and Technology (MUET), Jamshoro, Pakistan

² Pakistan Agricultural Research Council (PARC)/NARC, Islamabad, Pakistan

³ Centre for Pure and Applied Geology, University of Sindh, Jamshoro, Pakistan

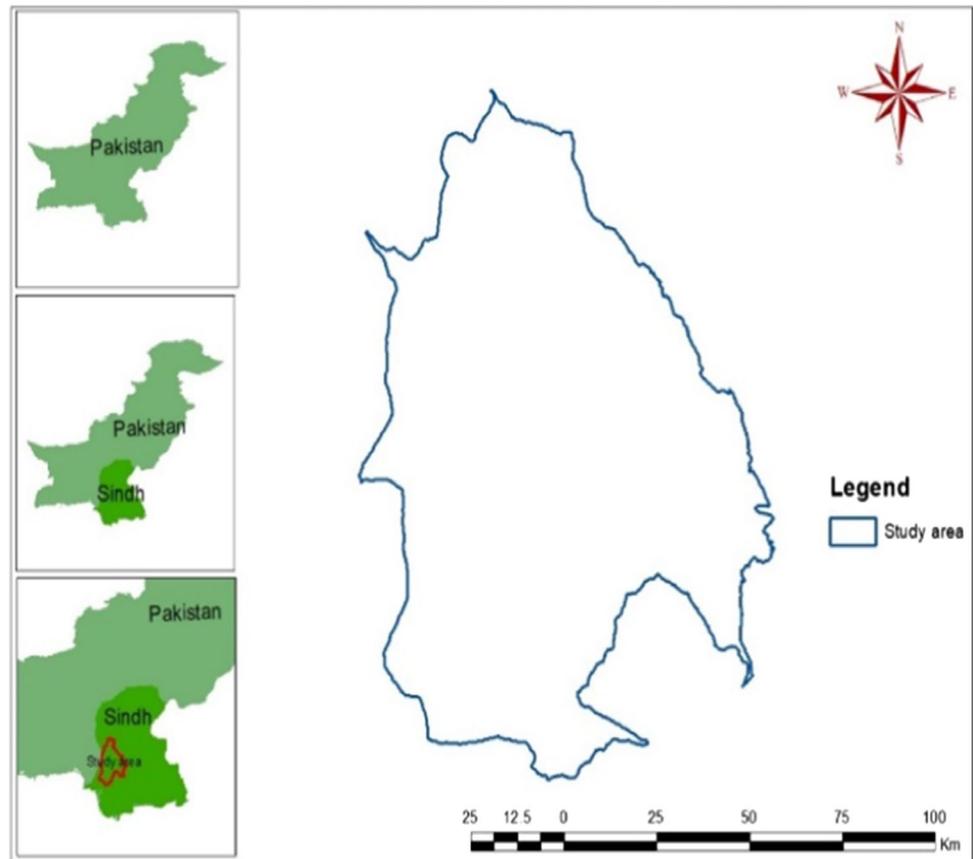
groundwater-based irrigation. Over and unjustified pumping of groundwater is a direct threat to the increasing population of the country. The supply and recharge of groundwater are the major challenging issues (Qureshi 2018).

The southwest hilly tract of Jamshoro district, Sindh province, is one of the significant drought prone and arid regions, well known is lower Kohistan. The area experienced historic extreme floods and droughts. The agriculture and livelihood depend on the hill torrents (Spate, Ephemeral River, rod-K, Nai) that carry the maximum runoff and feed the lands for agricultural purposes, which is the primary recharging source of shallow aquifers.

Since the last two decades, globally, numerous research experimentations have been carried out to estimate the groundwater yield (Muchingami et al. 2012; Redhaounia et al. 2016; Singh et al. 2019; Kalhor et al. 2019; Varua et al. 2018). Remote sensing (RS) and Geographic Information System (GIS) are capable techniques for sustainable development of underlying groundwater assets, particularly for the countries which lack data. A GIS can be utilized successfully for this reason to consolidate extraordinary hydrogeological themes and can examine the groundwater potential in the specific cluster or on a regional scale (Shahid et al. 2000). The use of GIS tool in investigating and quantifying groundwater potential has extended rapidly since the mid-1990s.

Chi and Lee (1994) studied the exploration of groundwater potential using an integrated application of GIS environment. Shahid et al. 2000 delineated the potential groundwater regions in the mountainous areas, using RS/GIS integration application, employing a hydrogeological database of the study area and also suggested that GIS technique is a powerful tool for existing wells and pumping tests. Srinivasa and Jugran (2010) conducted a study aimed to identify the potential groundwater zones in the mountainous regions by employing RS/GIS applications in Andhra Pradesh, India. They delineated and classified thematic maps of water quality for the livelihood. The results showed that 1.64% of the study area is of extreme groundwater potential with good to moderate quality of water. Though, 31.68% high groundwater potential, 31% suitable to the moderate potential of groundwater with the ideal to average quality. Ashraf et al. (2009) conducted a study to investigate the potential of artesian aquifers in the arid mountainous regions of DI Khan, Pakistan. They employed the GIS as an integrated tool to examine the potential of groundwater and picture the spatial data in two dimension and three dimension (2D and 3D). Subsurface lithology, stratigraphic data, and primary database of the aquifers have been recorded through the survey and observations. Various underground profile layers focused on examining the underground conditions of the study area employing Rockwork, the GIS integrated software. They discovered the extractable quantity

of about 1700 MCM of water potential was available in the study area. Using geoprocessing techniques for lithological modeling identified three different aquifers based on subsurface stratigraphic information. Two aquifers were discovered semi-confined to confine with the depth of about 200 m. Confined aquifer-1 was found at a depth of 118–113 m, aquifer-2 discovered at a depth of 182–195 m. Chowdhury et al. (2010) investigated the groundwater potential using RS/GIS and GPS ground truthing techniques in West Bengal, India. They delineated the thematic layers of lithology, land use, drainage, recharge, soil, slope, and water bodies using IRS-ID imagery and assigned weight on each thematic layers. The study discovered three types of groundwater potential zones categorized as good (15%), moderate (55%), and poor (30%) of the total study area. Machiwal et al. (2011) researched by employing GIS and MCDM techniques to assess groundwater potential in a semiarid region of Rajasthan, India. The study initiated by delineated thematic layers of topography, slope, geomorphology, geology, soil, groundwater level before and after the monsoon, annual recharge, yearly precipitation, and water bodies and estimated their impact on groundwater recharge by principal component analysis technique. The results showed four potential groundwater zones, i.e., good, moderate, poor, and very poor, based on groundwater potential index values. They have achieved consequences verified with the yield of 39 existing wells in the study area. Muchingami et al. (2012) have utilized a methodology of RS and GIS for groundwater investigation and identified the potential. GIS is additionally considered for multi-criteria examination in resource assessment. Jasrotia et al. (2013) employed the RS and GIS environment application to discover the groundwater and quality for sustainable development. The hydrological data, i.e., slope, soil, drainage, the status of aquifer level before and after rainfall in western Doon Valley, Himalaya, India. Ahmed et al. (2015) examined the groundwater potential using catastrophe theory-based evaluation method in the arid regions of lower Baluchistan, province of Pakistan. The groundwater system was categorized into five subsystems, i.e., geology, soil, drainage, density, slope, and precipitation. Furthermore, seventeen subsystem indicators were employed in the proposed modeling integrated with GIS approach. Thematic maps of subsystems and combined layers of groundwater potential zones were estimated using the weights of indicators. The consequences of confirmed by present wells in the study area. The results showed high-populated wells found in the highest potential zone of groundwater and proved the satisfactory results of the catastrophe theory for assessment of potential groundwater zone. Nasir et al. (2018) delineated groundwater potential using GIS and multi-influence factor (MIF) technique in Swat, Khyber Pakhtunkhwa province of Pakistan. Seven factors, i.e., slope, drainage density, geology, rainfall, soil, land use/land cover,

Fig. 1 Layout map of study area

and lineament density, were employed. The MIF method was used to calculate the weight and score of each effect. The study was designed by main and slight influences on the recharge and groundwater potential and eventually estimated the groundwater potential by the development of thematic maps weight overlaying using GIS. Four groundwater potentials were identified as poor, good, high, and very high potential. Sadaf et al. (2019) identified the potential sites of groundwater using remote sensing and GIS approach in the river basins of Pakistan. They delineated thematic maps of geology, drainage, and hydrological structural did overlay analysis and assessed the potential groundwater zones. Shahab et al. (2019) applied the DRASTIC method in GIS to estimate susceptibility of shallow aquifers.

A few examinations have been made to distinguish the groundwater potential utilizing hydrological and geological procedures. Reconciliation of RS and GIS gives a favorable position of approaching the extensive area of study. These strategies are fast and are useful techniques in delivering valuable information on geographical and geomorphological parameters that assist in interpreting groundwater potential.

In the previous investigations, the delineation of groundwater potential was investigated by employing GIS technology with lithological layer information. Thus, the groundwater potential was estimated by developing individual layers

of each rock, and the data of porosity have been interpreted with the rock volume of each layer (Saya 2010; Michael 2011).

The current study was conducted in the Kohistan region of Jamshoro district of Sindh province of Pakistan. The study aimed to estimate the groundwater potential using GIS modeling.

Study area

Location and accessibility

The study area is in the lower part of the Kohistan region, which stretches in the north with district Dadu, eastern district Hyderabad, southern Thatta district, southwest Malir district, and in the west Kalat district of Baluchistan. The study area is of 9057 km². Hyderabad, Thatta, and metropolitan city Karachi are nearest to the study area located at the distance of 70, 85, and 95 km, respectively (Figs. 1, 2).

Climatology

The study area is climatically arid; afterward, evaporation increased precipitation. The mean annual rainfall is in the range of 150–250 mm, and annual evaporation is in the

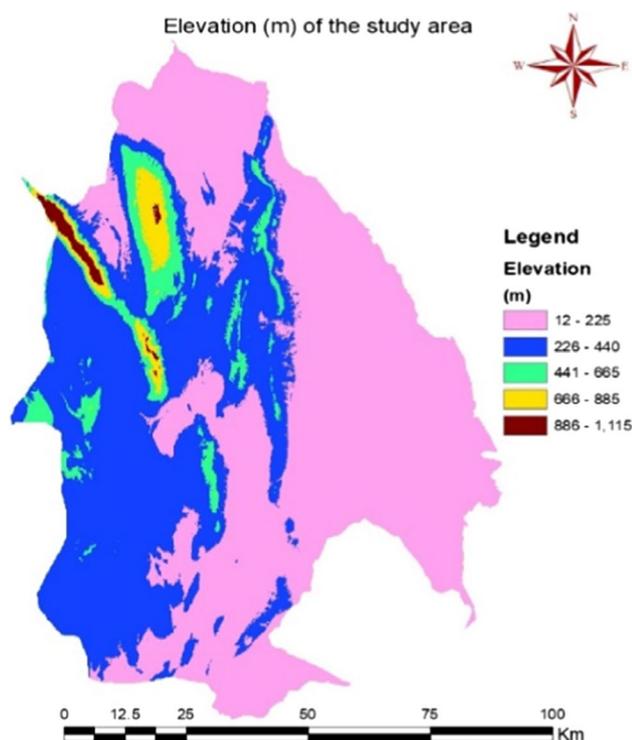


Fig. 2 Elevation of the study area

range of 1800–3000 mm. The moist weather of the south area becomes increasingly hot and dry as trips north in summer. Due to low and uneven rainfall, the rainwater harvesting practice is not possible in various areas. June is the warmest month, with an average maximum temperature exceeding 48.00 °C. Due to little moisture (Mirjat et al. 2011). The weather in the study area is intensively warm during summer and cold in the winter season. It has a large variation in the temperature from north to south region of the research area. However, the southern area consists of Mole, Sari, and Thana Boula Khan, and the main areas of the upper location of the area experience moderate climate. The precipitation ratio is of about 80% in the monsoon season (July–September). The short duration of rainfall percolates in the soil layer.

Hydrology

Precipitation is the primary source of torrential runoff. Most of the floods generate in the mountainous regions, which travel to the low-lying area of the study area that receives fresh floods; local people reduce the velocity of runoff by diverting a quantity to the lands for agricultural activities. The natural streams (hill torrents) are locally known as ‘Nai’ and well known as Ephemeral River, Spate, and Rod-Kohi. The primary ‘Nais’ in the region area are Desoi, Dawoo,

Darwat, Baran, Sari, Mole, and Kalu. Some specific reservoirs, ponds constructed in the depression of beds, are available to capture floodwater (Mirjat et al. 2011).

The watershed drainage slope of hill torrents is divided into two segments; one is the area toward Nai Baran, which is the prominent one, collecting the excess flood over Desoi, Dawoo, Darwat, and additional torrents throughout severe floods to divert into Indus River. The second segment of drainage slope receives intensive quantity of runoff through Mole, Sari, and Rahuja and finally diverts the torrential flood into the Kalu Nai and stores in the most prominent natural Lake Keenjhar located at the foot hills of district Thatta. Lake Keenjhar is the primary source of fresh water supply to Karachi metropolitan city.

Hydrogeology

The study area depends on shallow and deep aquifers. The primary sources for recharging aquifers are hill torrents depending on the precipitation pattern. The study area has unique phenomena for groundwater recharge depending on the various factors, i.e., infiltration, precipitation pattern, temperature. The precipitation produces runoff, which is naturally collected into the ditches, ponds due to rough conditions of the land surface. Water is absorbed in the unsaturated zones before recharging aquifers. The appropriate quantity of torrential water is diverted to the dry lands for harvesting purposes. The diverted water is stored for about 15–45 days; water gradually percolates beneath the surface by moistening the upper layer of dry lands to facilitate the land ready for cultivation. Monsoon is a significant period of recharging aquifers in the study area (PARC 2010). Thus, the groundwater is exploited in large quantities, which assumes a significant role in food, fiber, and drinking water supply in the region (Khan and Mian 2000; FAO 2002). Adequate knowledge of the occurrence and accessibility of groundwater is consequently essential for the sustainable management of groundwater resources to supply sufficient water to increasing population. Numerous components influence the event and development of groundwater, which incorporates geology, lithology, topographical structures, and scale of climate, porosity, land grade, infiltration rate, land use, land cover (Jaiswal et al. 2003).

Geology

The study area is situated in the Laki region, located in the hilly southern part of the lower Indus basin, and varies in age from Eocene to Pleistocene. Lithologically, the area comprises of alluvium, clay, silt, sand, gravel, limestone, sandstone, shale, conglomerates, etc. The main physical formation is Laki formation of primary Eocene age that covers the dominant lithology of limes along with shales,

Table 1 Geological formations of the study area

Name of formation	Age of formation
Manchar formation	Upper Miocene
Nari formation	Oligocene
Kirthar formation	Upper Eocene
Tiyon formation	Middle Eocene
Laki formation	Lower Eocene

sandstones, and lateritic clay. The Laki formation in mid of the Eocene period, Tiyon formation is showing in the investigation region, which contains primary lithology, i.e., shales, marls, and limestone. In the Oligocene age, the Nari development is visible, which covers mostly sandstone and marl store. In Miocene age, Gaj formation revealed and comprised limestone, sandstone, shales, and slight aggregates. In the mid of Miocene age, Manchar formation includes sandstones, shales, mud with subordinate aggregate beds, and in the Pliocene period, Dada combination is uncovered and generally contains mix, rock, and cobble (Table 1). A large portion of the fresh water aquifer exposes the hill torrents, mostly at the shallow depth (Soomro et al. 2019) (Fig. 3).

The basic stratigraphic setup of the study area from bottom to top is shown in Fig. 3.

Material and method

The assessment of the groundwater potential is based on densely populated cluster area of dug wells and tube wells; the rest of the area is not counted for the estimation of groundwater potential. Due to the large and scattered study area, the study was conducted only on a cluster basis. Fourteen dug wells and eleven tube wells scattered on the area of about 2359.64 km², and 847.61 km², respectively, selected as a high dense area of wells from the total study area. The principle criteria chose in this examination for the evaluation of groundwater potential are lithology, soil, layer thickness, and altitude level of each rock layer concerning mean sea level.

Data acquisition

The database has been developed by conducting interviews and filling questionnaires from the farmers of Jamshoro district. The survey contained detail and location-wise identification of farm, their owner's name, and type of well. The current location of groundwater sources has been recorded using GPS as a field truth. The collected GPS latitude and longitude coordinates were verified by Google Earth using GIS tool to convert point shape file into Kml format. The groundwater level was examined form a long period from October 2016 to April 2019 along with seasonal variation. Corresponding data of each aquifer (tube well and dug

well) are recorded and formulated in Table 2. The lithology (rock information) is collected from MICON Engineering Consultants.

Table 3 reveals the range of both aquifers, i.e., shallow aquifer ranges from 17.82 to 33.87 m, while the deep aquifer ranges from 95.44 to 188.4 m in the study area as per the data collection. The location of dug well and tube well is shown in Fig. 4. Thus, Table 4 gives estimated porosity in the Laki formation.

Table 4 gives porosity values and mentions their source through the literature. The italic values are taken and used for the estimation of groundwater potential in the study area. The rest of the values are not incorporated in the flow of groundwater due to very low permeability, so not considered as potential groundwater zone. Further details of the porosity are discussed at the end of the section of results and discussions.

The precipitation pattern of the study area is unreliable and experienced erratic droughts and extreme flood events in the previous history. The existing situation of water resources is critical in the region, especially the groundwater potential for sustainable development. The integration of different thematic layers is possible in a GIS environment, which helps in the rapid analysis of spatial data. To estimate the groundwater potential, a geospatial technique is beneficial. The groundwater potential is determined by developing individual layers of each rock, and the porosity data have been interpreted with a layer of each rock. The specific yield of water potential has been calculated from the layers of rock.

A CSV spreadsheet format of the database was used in ArcGIS version 10.5.1. The dataset is georeferenced; the thematic layers of existing dug wells and tube wells, surface hydrology, infrastructure, and landforms were developed in GIS using UTM (Universal Transverse Mercator) projection system (ESRI 1988). Subsequently, it is converted into a spatial layer, which allows for the analysis of lithological spatial layers. The layers developed based on surface and subsurface lithology and the variation (top and bottom layer values) of each thickness and zoning of aquifers into dug well and tube well. The point shape file is converted into raster layers using Topo to Raster, a tool of 3D Analyst. Later, depth of each layer is converted to a raster format by subtracting the bottom layer from the top layer using the raster calculator of the Spatial Analyst Tool. Finally, the volume of each layer was estimated using surface volume 3D Analyst tool and recorded in csv spreadsheet format. From the csv results, the size of each rock layer, corresponding cluster-wise area of dug wells and tube wells were quantified. The potential of groundwater was estimated by multiplying average porosity of each layer of stratigraphy volume of rock (Fig. 5).

The detailed methodology of the current study is described in Fig. 5.

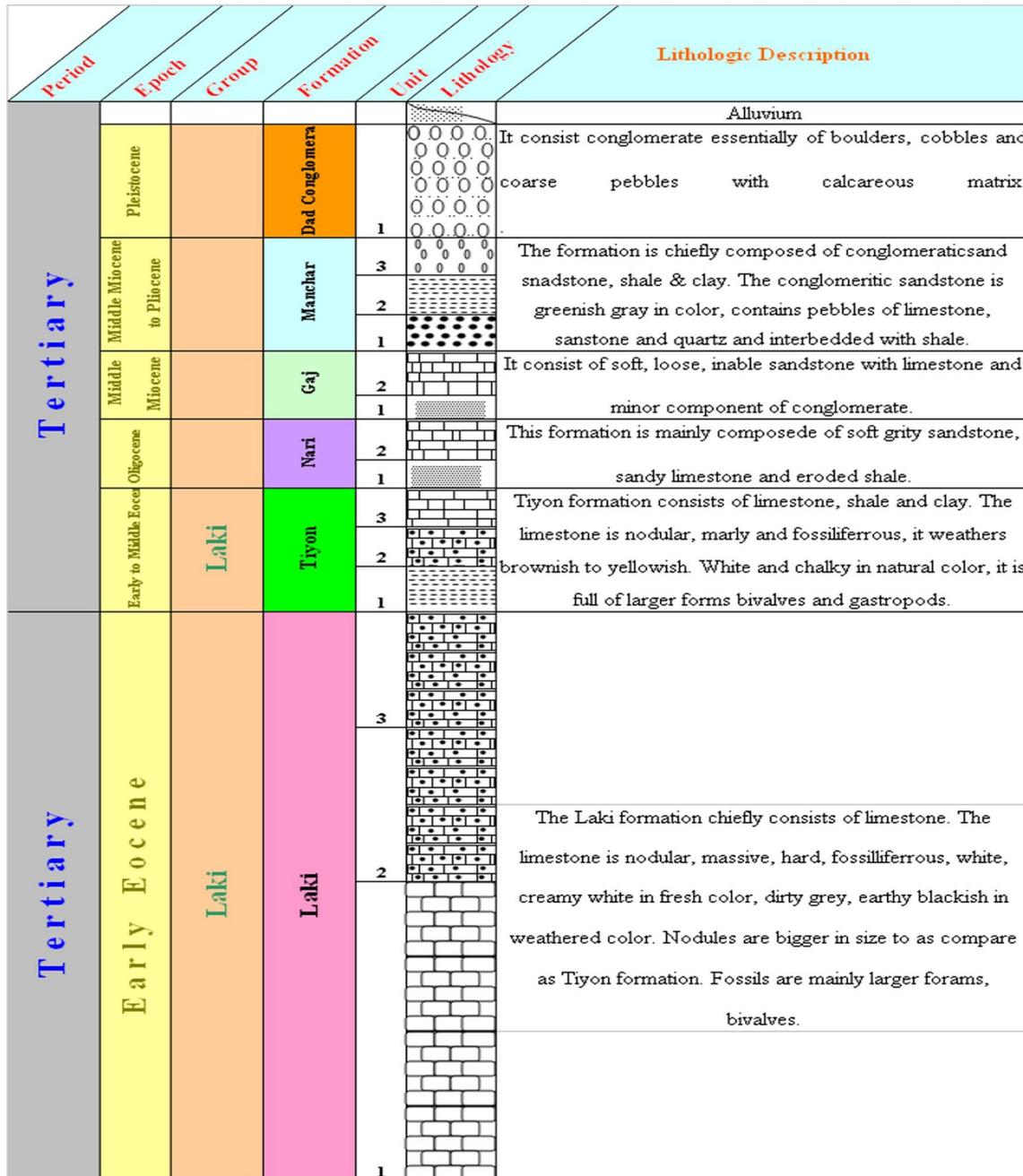


Fig. 3 Stratigraphical chart of surveyed area showing lithological log. Source: Soomro et al. (2019)

Results and discussions

The dense bunch of two aquifers zones of dug wells and tube wells have been selected containing an area of 2359.64 and 847.61 km², respectively. However, the total study area spread over 9057 km² (Table 5).

Mostly, Laki formation of Eocene age present in the study area contains limestone, sandstone, and clay lithology. Two aquifers zones are current, shallow, and deep,

dug wells are the source of shallow, and tube wells are the sources of tube well aquifer zones. The dug well lithology from top to bottom is alluvial, silt sand gravel, clay, shell, sand, and limestone layers, while the tube well lithology is alluvial, silt, sand, gravel, clay, shell, silt, sand, and limestone layer. The lithology of both aquifer zones is the same. Thus, the silt lithology layer is found again at the below section of lithology between the shell and sand layers in the tube wells. The lithology layers are sub-divided into

Table 2 Database of the Study area

Name Farmr.	Well_ID	Village	X	Y	ElevinM	Max Level	Min Level	Base Lev.	AHrznTop	AHrznBtm	SpnFmTop	SpnFmBtm	LdLsTop	LdLsBtm	PotFmTop	PotFmBtm	Bsmnt Top	Bsmnt Btm
G Ali	JTW1	Dad Bhai	67.62464	25.19962	209	170	167.5	78	209	193.76	193.76	163.7	163.7	148.04	148.04	78	78	42
Juma Bikak I	JTW2	Dad Bhai	67.62863	25.20493	202	110	105.5	49.6	202	194.38	194.38	180.66	180.66	170	170	49.6	49.6	13.6
Juma Bikak II	JTW3	Habkan	67.62864	25.20483	202	133	130.53	49.6	202	194.38	194.38	157.8	157.8	150.18	150.18	49.6	49.6	45.6
Rahu Shod	JTW4	Dad Bhai	67.53168	25.39641	299	235	227	192	299	298.09	298.09	267.61	267.61	264.56	264.56	228.9	228.9	192
M Barat	JTW5	Mole	67.54644	25.36981	289	227.88	226.55	167	289	273.76	273.76	243.28	243.28	228.04	228.04	189.94	189.94	167
TC TBK I	JTW6	TBK	67.8257	25.36953	128	25.89	23.79	1.5	128	118.86	118.86	83.8	83.8	60.94	60.94	1.5	1.5	-2.5
TC TBK II	JTW7	TBK	67.82462	25.36919	129	82.78	81.75	37.5	129	121.38	121.38	83.28	83.28	68.04	68.04	37.56	37.56	33.56
Ali Bux	JTW8	Morai Nai	67.67444	25.45884	242	182.55	181.05	120	242	238.95	238.95	220.66	220.66	206.75	206.75	130.75	130.75	120
Allah Bux	JTW9	Ular Nai	67.60877	25.42775	237	207.25	204.68	112	237	233.91	233.91	217.19	217.19	206.75	206.75	134.89	134.89	112
Ali Murad	JTW10	Ular Nai	67.62331	25.42728	234	205	203.1	112	234	217.24	217.24	179.14	179.14	165.42	165.42	119.7	119.7	112
Vrym Burft	JTW11	Sari	67.64762	25.4118	220	98	96.85	56	220	217.56	217.56	167.27	167.27	160.26	160.26	84.06	84.06	56
Nzar Mohammad	JDW1	Sari	67.56941	25.30644	254	247.3	246.5	230.8	254	249.8	249.8	245.5	245.5	244.2	244.2	233.5	233.5	230.8
Khskhli farm	JDW2	Sari	67.5715	25.29055	245	239.8	239.5	223.66	245	243.5	243.5	241.5	241.5	238.5	238.5	229.5	229.5	223.66
G Mohammad	JDW3	TBK	67.82553	25.37329	129	127.89	127.65	110.7	129	122.91	122.91	119.86	119.86	119.25	119.25	112.5	112.5	110.7
Pir Bux	JDW4	TAK	67.79294	25.43115	139	136.01	135.89	124	139	135.96	135.96	131.5	131.5	126.5	126.5	124	124	120.2
Sindhlni village	JDW5	TAK	67.79345	25.43054	140	136.99	136.65	125.98	140	137.45	137.45	132.5	132.5	131.78	131.78	125.98	125.98	122.18
Sanyal Khskhli	JDW6	TBK	67.82367	25.3696	129	125	124.56	109.19	129	126.75	126.75	123.19	123.19	121.19	121.19	113.5	113.5	109.19
Gul Muhammad	JDW7	TBK	67.82388	25.36964	139	134	133.01	117.05	139	136.8	136.8	133.8	133.8	132.1	132.1	120.8	120.8	117.05
A Latif	JDW8	TAK	67.84538	25.39528	131	123.79	123.35	97.5	131	125.5	125.5	120.78	120.78	115.5	115.5	97.5	97.5	93.7
Muhammad Ayoub	JDW9	TBK	67.84707	25.39705	132	125.98	125.79	106.1	132	129.25	129.25	121.78	121.78	114.45	114.45	106.1	106.1	102.1
H Nisir Plri	JDW10	TBK	67.84708	25.3997	132	124.95	124.65	102.13	132	129.78	129.78	121.78	121.78	119.5	119.5	102.13	102.13	98.13
Mtyo K I	JDW11	TAK	67.85583	25.41247	134	131.95	131.15	120.29	134	132.5	132.5	129.2	129.2	127.5	127.5	120.29	120.29	116.49
Mtyo K II	JDW12	TBK	67.85599	25.41218	134	128.8	128.35	114.19	134	132.5	132.5	128.8	128.8	127.1	127.1	115.8	115.8	114.19
M Yousif I	JDW13	Kotri	68.31263	25.57989	56	49.89	49.47	31.62	56	53.8	53.8	50.8	50.8	49.1	49.1	33.5	33.5	31.62
M Yousif II	JDW14	Kotri	68.31256	25.57996	56	49.78	49.01	29.18	56	54.5	54.5	50.2	50.2	44.5	44.5	29.18	29.18	27.38

Table 3 Groundwater sources and their corresponding depth in the study area

Aquifer	Jamshoro	
	Minimum	Maximum
	Depth (m)	
Dug well (shallow aquifer)	17.82	33.87
Tube well (deep aquifer)	95.44	188.4

Table 4 Average porosity for groundwater lithology in the study area. *Source:* Khokhar et al. (2016), Manger (1963)

Porosity of layer		
Alluvial	30.2%	0.30
Silt	53.2%	0.53
Sand	30%	0.30
Gravel	40%	0.40
Clay	42.4%	0.42
Shale	17%	0.17
Limestone	15%	0.15

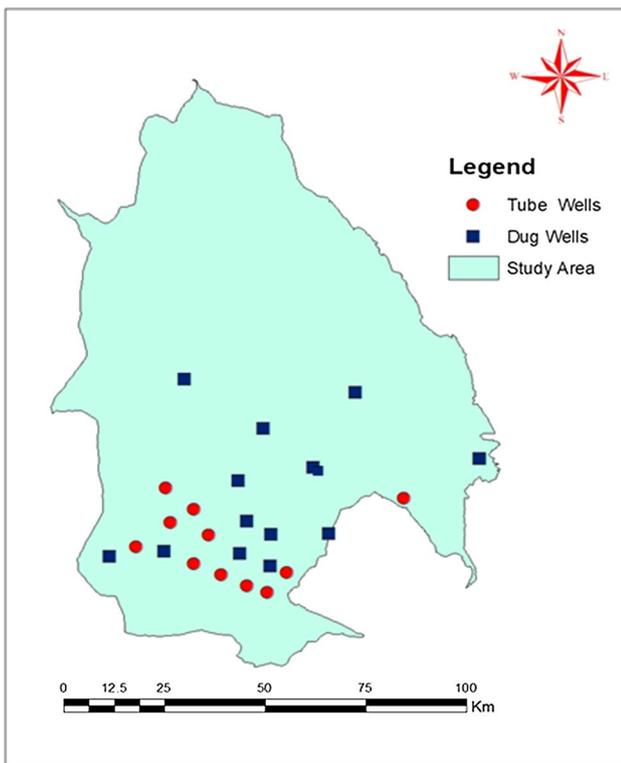


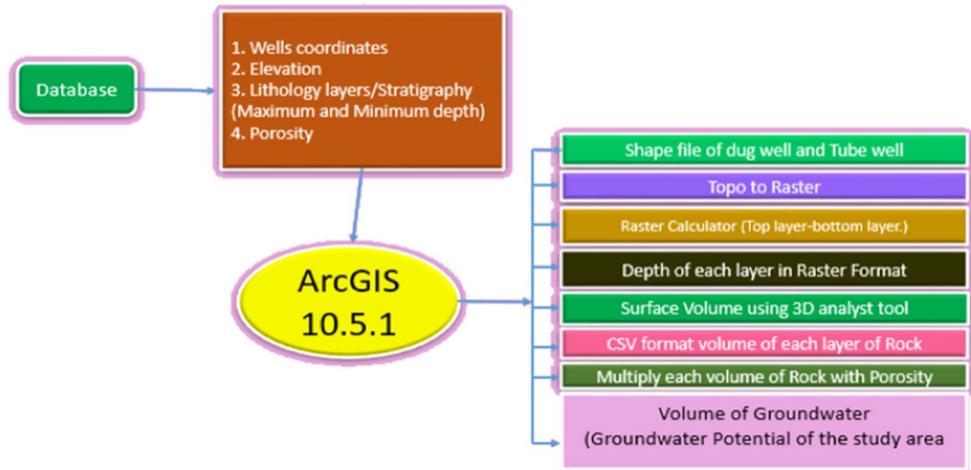
Fig. 4 Layout map of tube well and dug well

five layers of stratigraphy: (1) A-Horizon, (2) Spergen Fm, (3) Leadville Ls, (4) Potosi Fm, (5) Basement (shown on the left side of Tables 6 and 7). However, Table 6 shows the potential of groundwater in a cluster of the dug well (shallow) zone of the aquifer in the study area. The estimated figures represent the sub-divided portion of A-Horizon rock volume 7,556,882,845.34 m³, and Leadville Ls 9,686,109,673.62 m³, which contains zero water sustainability due to alluvial soil at the top part in the lithology, and, subsequently, silt, and dry sand particles that contain zero porosity.

Though the stratigraphy portion of Spergen Fm has rock volume 8,474,871,965.27 m³, the porosity of the gravel layer is 0.40 percent and the potential of water is 3389.95 million cubic meter of (MCM). The Potosi Fm rock volume is 16,971,426,450.43 m³, the porosity of sand layer is 0.30 percent, and the potential of water is 5091.43 MCM. However, the basement contains limestone rock and volume is 6,068,723,430.21 m³, the porosity of limestone is 0.15 percent, and the possibility of water is 910.31 MCM. The total potential of water in a cluster of dug well (shallow) aquifer zone contains 9391.69 MCM in the study area.

Table 7 describes in detail the potential of groundwater in a cluster of tube wells (deep) zone of the aquifer in the study area. The assessed scenario was expressing the stratigraphic

Fig. 5 Flowchart of methodology



Flow Chart, Estimation of Groundwater Potential using ArcGIS

Table 5 Cluster-wise area of groundwater zones

District	Study area (km ²)	Aquifer zone	No. of wells	Cluster-wise area (km ²)
Jamshoro	9057	Dug well	14	2359.64
		Tube well	11	847.61
Total			25	3207.25

setup, which detailed as sub-divided. The section of A-Horizon rock contains volume 7,265,891,994.48 m³, and Lead-Ville Ls 7,859,254,535.48 m³, which includes alluvial soil at the top, consequently silt and dry sand layer, with zero porosity of water potential. However, the stratigraphy slice of Spergen Fm rock volume is 15,974,791,401.32 m³, the porosity of the Gravel layer is 0.40 percent, and the potential of water is 6389.92 million cubic meter of (MCM). The Potosi Fm rock volume is 37,489,096,451.50 m³, the

Table 6 Groundwater potential of a cluster of dug well zone

Soil Type	Lithology	Stratigraphy	Mean Elevation (m)	Volume of Rock (Cubic Meter)	Porosity	Volume of Water (Cubic Meter)	Volume of Water (MCM)
Alluvial		AHrznTop	131.00	7556882845.34	0.00	0.00	0.00
Silt							
Sand		AHrznBtm	125.50	8474871965.27	0.40	3389948786.11	3389.95
Gravel		SpnFmTop	125.50				
Clay		SpnFmBtm	120.78	9686109673.62	0.00	0.00	0.00
		LdLsTop	120.78				
Shell		LdLsBtm	115.50	16971426450.43	0.30	5091427935.13	5091.43
Sand		PotFmTop	115.50				
Lime Stone		PotFmBtm	97.50	6068723430.21	0.15	910308514.53	910.31
		BsmntTop	97.50				
		BsmntBtm	93.70				
Groundwater Potential							9391.69

Table 7 Groundwater potential of a cluster of tube well zone

Jamshoro Tube well Cluster wise Lithology of Kirthar Formation, Volume of Ground water									
Soil Type	Lithology	Stratigraphy	Mean Elevation (m)	Volume of Rock (Cubic Meter)	Porosity	Volume of Water (Cubic Meter)	Volume of Water (MCM)		
Alluvial		AHrznTop	209.00						
Silt				7265891994.48	0.00	0.00	0.00		
Sand		AHrznBtm	193.76						
Gravel		SpnFmTop	193.76	15974791401.32	0.40	6389916560.53	6389.92		
Clay		SpnFmBtm	163.70						
Shell		LdLsTop	163.70						
Silt				7859254535.48	0.00	0.00	0.00		
Sand		LdLsBtm	148.04						
		PotFmTop	148.04	37489096451.50	0.30	11246728935.45	11246.73		
		PotFmBtm	78.00						
Lime Stone		BsmntTop	78.00	15345959236.91	0.15	2301893885.54	2301.89		
		BsmntBtm	42.00						
Groundwater Potential							19938.54		

porosity of sand layer is 0.30 percent, and the potential of water is 11,246.73 MCM, whereas the basement contains limestone rock volume of 15,345,959,236.91 m³, the porosity of limestone is 0.15, and the potential of water is 2301.89 MCM. The total potential of water in the cluster of the tube well (deep) aquifer zone contains 19,938.54 MCM in the study area.

However, the upper portion of stratigraphy contains alluvial, silt, and sand, and the porosity of the layer is zero. The top layer is unsaturated and assumed as surficial deposits not contributing to groundwater storage. The creation in coarse-textured soil leads to weak soil aggregation and consists of a free drainage profile with low storage and water repellent capability due to avoiding percolation rate beneath the soil layer subsequent more enormous surface water losses. The other layer LeadVille Ls stratigraphy (containing shale, clay, and silt) is confined under pressure acting as an aquitards that restrict the flow of groundwater due to very low permeability, so not considered as potential groundwater zone (soil quality). In the literature, it is known as the cap rocks. These layers allow the water to percolate downward and refill aquifers the barrier for groundwater. The clay has much water holding capacity with dense clay subsoil as well the perched water stored above the layer.

Table 8 describes that the study area is comprised of two aquifer zones, shallow and deep. Fourteen dug wells occupied a cluster area of 2359.64 km² and the estimated

Table 8 Total groundwater volumes (MCM) in the study area

District	Aquifer	No. of wells	Cluster-wise area (km ²)	Cluster-wise groundwater volume (MCM)
Jamshoro	Dug well	14	2359.64	9391.69
	Tube well	11	847.61	19938.54
Total		25	3207.25	29330.23

Table 9 Groundwater volume (MCM)/km² in the study area

District	Area under monitoring (km ²)	Total volume of groundwater potential (MCM)	Groundwater volume (MCM)/km ²
Jamshoro	3207.25	29330.22	9.14

potential of water is 9391.69 MCM in shallow zone. Nevertheless, 11 tubes wells occupied an area of 847.61 km² and the projected potential of water is 19,938.54 MCM in the deep zone of the aquifer. A total of 25 wells were selected in the study area that occupied total area of 3207.25 km², and the estimated potential of water is 29,330.23 MCM in the chosen cluster zones of the study area.

Table 9 represents the monitoring area of 3207.25 km², estimated total groundwater potential of 29,330.22 MCM,

and consequently, occupied groundwater potential volume of 9.14 MCM/km² in both the clusters of aquifer zones.

The GIS tool is advanced and useful to estimate the groundwater potential; the assessment process depends on the method applied. Some times study needs more data that complicate the research, and results may produce many errors against the accuracy of the results. Some times GIS assessment is not possible due to technical and procedural complications. Every methodology of GIS has its advantages and disadvantages during the processing.

Conclusions

The study area relies on the precipitation, which is only the recharge source of groundwater aquifers. The current situation of water resources is critical in the region, especially the groundwater potential for sustainable development. In the present investigation, the groundwater potential of shallow and deep aquifer zones was estimated in the study area. GIS has been employed as a preservative tool to determine groundwater potential. Groundwater aquifers were identified, which contains the lithology of Laki formation and consists of sand gravel and limestone rock layers in the shallow and deep aquifers. The estimated groundwater potential is 29,330.22 MCM under the monitoring area of 3207.25 km²; thus, the volume of groundwater is 9.14 MCM/km² in the selected cluster zone of the study area. The lithological information and geoprocessing techniques identify the potential groundwater sites of the study area. It is realized that using the GIS tool combats the susceptible situation and provided authenticated results of groundwater potential. Estimation of groundwater potential studies leads to the uncertain condition of the area for sustainable development.

Acknowledgements

Authors are thankful to MICON Engineering Consultants; USPCAS-W/MUET-Jamshoro is gratefully acknowledged for providing access to groundwater lithology and use of GIS laboratory to accomplish the goal of current research. Funding for the research study is not provided by any source.

References

- Ahmed K, Shahid S, bin Harun S, Ismail T, Nawaz N, Shamsudin S (2015) Assessment of groundwater potential zones in an arid region based on catastrophe theory. *Earth Sci Inf* 8(3):539–549
- Ashraf A, Iqbal MB, Mustafa N (2009) Investigating potential artesian aquifers in Rod-Kohi area of DI Khan NWFP using GIS and geo-processing techniques. Mehran University of Engineering and Technology, Jamshoro
- Chi KH, Lee BJ (1994) Extracting potential groundwater area using remotely-sensed data and GIS techniques. Paper presented at the Regional Seminar on Integrated Applications of Remote Sensing and GIS for Land and Water Resources Management, Bangkok (ESCAPE)
- Chowdhury A, Jha MK, Chowdary VM, Mal BC (2010) Integrated remote sensing and GIS-based approach for assessing groundwater potential in West Medinipur district West Bengal, India. *Int J Remote Sens* 30(1):231–250. <https://doi.org/10.1080/0143160802270131>
- ESRI Environmental Systems Research Institute, Inc. Redlands, CA 92373-8100 U.S.A. (1988)
- FAO (2002) Water Management in Baluchistan Pakistan. Food and Agriculture Organization of the United Nations
- Fetter CW (1980) Applied hydrogeology. Merrill Publication, Columbus, p 488
- Gardner-Outlaw T, Engelman R (2019) Revised data for the population action international report, sustaining water: population and the future of renewable water supplies, pp 21
<http://soilquality.org.au/factsheets/water-availability>
- Jaiswal RK, Mukherjee S, Krishnamurthy J, Saxena R (2003) Role of remote sensing and GIS techniques for generation of groundwater prospect zones towards rural development—an approach. *Int J Remote Sens* 24:993–1008
- Jasrotia AS, Bhagat BD, Kumar A, Kumar R (2013) Remote sensing and GIS approach for delineation of groundwater potential and groundwater quality zones of western Doon Valley, Uttarakhand, India *J Indian Soc Remote Sens* 41:365–377. <https://doi.org/10.1007/s12524-012-0220-9>
- Kalhor K et al (2019) Assessment of groundwater quality and remediation in karst aquifers: a review. *Groundw Sustain Dev* 8:104–121
- Khan AS, Mian BA (2000) Groundwater development issues of Baluchistan. Paper presented at the In Proceedings of the global water partnership seminar on regional groundwater management. Islamabad
- Khokhar QD, Hakro AAAD, Solangi SH, Siddiqui A (2016) Textural evaluation of nari formation, Laki Range, Southern Indus Basin. *Pak Sindh Univ Res J (Sci Ser)* 48(3):633–638
- Machiwal D, Jha MK, Mal BC (2011) Assessment of groundwater potential in a semi-arid region of India using remote sensing, GIS and MCDM techniques. *Water Resour Manage* 25(5):1359–1386
- Manger GE (1963) Porosity and bulk density of sedimentary rocks. Geological Survey Bulletin 1144-E. Developed by U.S. Atomic Energy Commission
- Michael CPBS (2011) New methods for quantifying and modeling estimates of anthropogenic and natural recharge: a case study for the barton springs segment of the edwards aquifer, Austin, Texas. A Thesis document, Master of Science in Geological Sciences. The University of Texas at Austin May 2011
- Mirjat MS, Soomro AG, Mirjat KH, Mirjat MU, Chandio AS (2011) 1 Potential of Hill-torrent spate irrigation in the Kohistan areas of Sindh “A Case Study”. *Pak J Agric Agril Eng Vet Sci* 27(2):100–114
- Muchingami I, Hlatywayo DJ, Chuma C (2012) Electrical resistivity survey for groundwater investigations and shallow subsurface evaluation of the basaltic-greenstone formation of the urban Bulawayo Aquifer. *Phys Chem Earth* 50(52):44–51. <https://doi.org/10.1016/j.pce.2012.08.014>
- Murasingh S, Jha R (2013) Conference paper: identification of groundwater potential zones using remote sensing and GIS in a mine area of Odisha. Conference Paper <https://doi.org/10.13140/rg.2.1.3374.6644>. <https://www.researchgate.net/publication/279715243>
- Nasir MJ, Khan S, Zahid H, Khan A (2018) Delineation of groundwater potential zones using GIS and multi influence factor (MIF) techniques: a study of district Swat, Khyber Pakhtunkhwa Pakistan. *Environ Earth Sci* 77(10):367
- Pakistan Agricultural Research Council (PARC) (2010) NRDP on water management of spate irrigation project-Dadu component. Completion report of Spate Project 2005–2010

- Qureshi AS (2018) Challenges and opportunities of groundwater management in Pakistan. In: *Groundwater of South Asia*. Springer, Singapore, pp 735–757
- Redhaounia B, Bedir M, Gabtni H, Ilondo B, Dhaoui M, Chabaane A, Khomsi S (2016) Hydro-geophysical characterization for groundwater resources potential of fractured limestone reservoirs in Amdoun Monts (North-western Tunisia). *J Appl Geophys* 128:150–162
- Sadaf R, Mahar GA, Younes I (2019) Appraisal of ground water potential through remote sensing in River Basin, Pakistan. *Int J Econ Environ Geol*:25–32
- Saya A (2010) Water volume calculation of hill country trinity Aquifer Blanco, Hays, and Travis Counties, Central Texas. GIS and GPS Applications in Earth Science. pp 22, <http://www.twdb.state.tx.us/GwRD/GMA/PDF/TrinityAquifer.pdf>
- Shahab A, Shihua Q, Rad S, Keita S, Khan M, Adnan S (2019) Groundwater vulnerability assessment using GIS-based DRAS-TIC method in the irrigated and coastal region of Sindh province Pakistan. *Hydrol Res* 50(1):319–338
- Shahid S, Nath SK, Roy J (2000) Groundwater potential modelling in a soft rock area using a GIS. *Int J Remote Sens* 21:1919–1924. <https://doi.org/10.1080/014311600209823>
- Singh A, Panda SN, Uzokwe VN, Krause P (2019) An assessment of groundwater recharge estimation techniques for sustainable resource management. *Groundw Sustain Dev* 9:100218
- Soomro A, Qureshi AL, Jamali MA, Ashraf A (2019) Groundwater investigation through vertical electrical sounding at hilly area from Nooriabad toward Karachi. *Acta Geophys* 67:247–261. <https://doi.org/10.1007/s11600-019-00247-9>
- Srinivasa RY, Jugran DK (2010) Delineation of groundwater potential zones and zones of groundwater quality suitable for domestic purposes using remote sensing and GIS. *Hydrol Sci J* 48:821–833. <https://doi.org/10.1623/hysj.48.5.821.51452>
- Todd DK, Mays LW (2004) *Groundwater Hydrology*, 3rd edn. Wiley, Hoboken (**Incorporated**)
- Varua ME et al (2018) Groundwater management and gender inequalities: the case of two watersheds in rural India. *Groundw Sustain Dev* 6:93–100



Novel hybrid approaches based on evolutionary strategy for streamflow forecasting in the Chellif River, Algeria

Moussaab Zakhrouf¹ · Hamid Bouchelkia¹ · Madani Stamboul² · Sungwon Kim³

Received: 11 July 2019 / Accepted: 22 October 2019 / Published online: 30 October 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

In this study, the feedforward neural networks (FFNNs) were proposed to forecast the multi-day-ahead streamflow. The parameters of FFNNs model were optimized utilizing genetic algorithm (GA). Moreover, discrete wavelet transform was utilized to enhance the accuracy of FFNNs model's forecasting. Therefore, the wavelet-based feedforward neural networks (WFFNNs-GA) model was developed for the multi-day-ahead streamflow forecasting based on three evolutionary strategies [i.e., multi-input multi-output (MIMO), multi-input single-output (MISO), and multi-input several multi-output (MISMO)]. In addition, the developed models were evaluated utilizing five different statistical indices including root mean squared error, signal-to-noise ratio, correlation coefficient, Nash–Sutcliffe efficiency, and peak flow criteria. Results provided that the statistical values of WFFNNs-GA model based on MISMO evolutionary strategy were superior to those of WFFNNs-GA model based on MISO and MIMO evolutionary strategies for the multi-day-ahead streamflow forecasting. Results indicated that the performance of WFFNNs-GA model based on MISMO evolutionary strategy provided the best accuracy. Results also explained that the hybrid model suggested better performance compared with stand-alone model based on the corresponding evolutionary strategies. Therefore, the hybrid model can be an efficient and robust implement to forecast the multi-day-ahead streamflow in the Chellif River, Algeria.

Keywords Daily streamflow forecasting · Wavelet transform · Feed forward neural networks · Multi-input multi-output · Multi-input single-output · Multi-input several multi-output · Chellif River

Introduction

Accurate estimates of streamflow can be utilized in several water engineering problems such as designing flood protection works for urban areas and agricultural land, and optimizing the water allocation for different purposes including agriculture, municipalities, and hydropower generation (Rezaie-Balf et al. 2019; Seo et al. 2018; Samsudin et al. 2011). The complexity of natural processes and the lack of data available for modeling streamflow require the utilization of specific implements for nonlinear and non-stationary natural phenomenon (Chu et al. 2016; Shoaib et al. 2015). Time series forecasting is one of the most and important methodologies utilized in streamflow modeling (Rezaie-Balf and Kisi 2017a; Seo et al. 2015; Krishna et al. 2011). The problem complexity rises when the models are applied for days/months forecasting in advance. Streamflow forecasting utilizing available multi-time-ahead series is a common task (Ghorbani et al. 2018; Karimi et al. 2018; Wang et al. 2009).

✉ Sungwon Kim
swkim1968@dyu.ac.kr

Moussaab Zakhrouf
mousaab@gmail.com

Hamid Bouchelkia
bouchelkiahamid@gmail.com

Madani Stamboul
stamboul_m@yahoo.fr

- ¹ URMER Laboratory, Department of Hydraulic, Faculty of Technology, University of Tlemcen, Tlemcen, Algeria
- ² Research Laboratory of Water Resources, Soil and Environment, Department of Civil Engineering, Faculty of Architecture and Civil Engineering, Amar Telidji University, Laghouat, Algeria
- ³ Department of Railroad Construction and Safety Engineering, Dongyang University, Yeongju 36040, South Korea

In recent years, the field of computational intelligence has promoted revolutionary changes to forecast the streamflow of complex and non-stationary time series in the development of non-conventional techniques (Delafrouz et al. 2018; Yu et al. 2018; Shafaei and Kisi 2016). In addition, various data pre-processing techniques have been utilized to enhance the accuracy of streamflow forecasting. The underlying techniques include principal component analysis (PCA) (Ravikumar and Somashekar 2017; Hu et al. 2007), continuous wavelet transform (CWT) (Deo et al. 2017; Rezaie-Balf et al. 2017; Sang et al. 2013), moving average (MA) (Yuan et al. 2017), wavelet multi-resolution analysis (WMRA) (Zakhrouf et al. 2018), maximum entropy spectral analysis (MESA) (Benedetto et al. 2015) and singular spectrum analysis (SSA) (Wu and Huang 2009; Baydaroglu et al. 2018). Recently, the combination of discrete wavelet transform (DWT) and artificial neural networks (ANNs) approaches has been accomplished as the successful alternative for hydrological modeling and forecasting from the previous literatures (Abdollahi et al. 2017; Ravansalar et al. 2016, 2017; Zakhrouf et al. 2016).

In many engineering problems, since streamflow forecasting utilizing simple time series (e.g., one-day-ahead) may not provide enough information, the multi-day-ahead (e.g., 2, 3, and 4-day-ahead etc.) forecasting is required to understand the physical processes of streamflow time series clearly (Chang et al. 2007). For this purpose, this research is designed to investigate the efficiency and capability of the hybrid model to forecast the multi-day-ahead streamflow. The wavelet-based feedforward neural networks (WFFNNs) model optimized by genetic algorithm (GA) is developed to forecast the multi-day-ahead streamflow in the Chellif River, Algeria. Also, this study applies the reconstruction of three evolutionary strategies [i.e., multi-input multi-output (MIMO), multi-input single-output (MISO), and multi-input several multi-output (MISMO)] utilizing the WFFNNs-GA model.

The contributions of this research can be classified as two parts. First, the combinational approaches (i.e., DWT, FFNNs, and GA) based on three evolutionary strategies (i.e., MIMO, MISO, and MISMO) are proposed, respectively. Second, the authors present the hybrid model for forecasting the multi-day-ahead streamflow of the Chellif River, Algeria. The performance of developed models is evaluated utilizing five statistical indices and diagnostic plots. Finally, the conclusion and future research are given.

Methodology description

Feedforward neural networks (FFNNs) model

FFNNs model typically consists of input, hidden, and output layers, respectively. Utilizing randomly generated connection

weights, the information has to be feed-forwarded from input to the hidden neurons and from the previous hidden to the next hidden neurons. Each hidden neuron utilizes a transfer function to calculate the outputs of hidden layer. The processed values have to be summed up at each hidden neuron, and information has to be passed on to the output neuron through connecting weights. The final values at the output layer are compared with the target values. The difference in errors between observed and calculated values is evaluated utilizing a backpropagation (BP) algorithm until the connection weights are optimized to obtain minimum error between observed and calculated values (Gowda and Mayya 2014). The detailed information and concept of FFNNs model can be seen in Bishop (1995) and Haykin (1999).

Genetic algorithm (GA)

GA is a heuristic and stochastic optimization technique based on the evolutionary and genetic theory (Kim and Kim 2008; Holland 1992). The working procedure of GA usually starts with random strings representing design or decision variables. Later, each string is evaluated to allocate the fitness value based on checking objective and constraint conditions. Then termination condition is verified in the algorithm (Gowda and Mayya 2014). GA starts with a population of individuals including selection, crossover, and mutation. These procedures are repeated from one generation to the next with the objective of reaching the global optimal solution.

Therefore, this paper focuses on optimizing the FFNNs and WFFNNs models' structure utilizing GA implementation. The first step in this approach is to choose the GA chromosome. Each individual in the population represents a possible configuration for the FFNNs and WFFNNs models' structure.

Discrete wavelet transform (DWT)

Wavelet transform (WT), one of data pre-processing techniques, has been utilized for hydrological time series analysis (Ghaemi et al. 2019; Tayyab et al. 2019; Quilty and Adamowski 2018; Uysal and Sorman 2017; Seo and Kim 2016; Parmar and Bhardwaj 2015). WT is a mathematical function which is capable of decomposing the continuous or discrete time signal into the higher- and low-frequency components of a signal (Mallat 1989). There are two main categories of wavelet transforms [e.g., continuous wavelet transform (CWT) and discrete wavelet transform (DWT)], respectively. CWT method treats with continuous functions and can be applied for discrete functions or time series (Seo et al. 2015, 2018; Nourani et al. 2009; Mallat 1989). In brief, CWT method is time-consuming and requires large

resources, while DWT method can be applied than CWT method easily (Nourani et al. 2009; Mallat 1989).

DWT method consists of two decompositions (e.g., low-pass and high-pass) and reconstructions (e.g., low-pass and high-pass), respectively (Seo et al. 2015, 2018; Mallat 1989). For practical application of DWT method, two filters (e.g., low-pass and high-pass) are utilized rather than two wavelets (e.g., father and mother) (Seo et al. 2015). The low-pass filter allows for the analysis of low-frequency components, while the high-pass filter allows for the analysis of high-frequency components (Seo et al. 2015; González-Audicana et al. 2005). The multi-resolution approach utilizing DWT method is a process to draw ‘approximations (show a conventional trend of original signal)’ and ‘details (indicate the high-frequency components)’ for a given signal (Seo et al. 2015, 2018; Mallat 1989). The feature report for DWT method can be collected from Nason (2008).

In this study, the authors focused on DWT method for decomposition and reconstruction of streamflow time series. By utilizing DWT method, the original time series was decomposed into various sub-signals of the low-scale, high-frequency components named ‘details’ (D), and the high-scale, low-frequency components named ‘approximations’ (A) at different resolution levels. The decomposed details and approximation of each input data were applied as new inputs to the FFNNs and WFFNNs models’ structure.

Evolutionary strategy

For a given time series $\{Q_1, Q_2, \dots, Q_t\}$ to perform H step ahead streamflow forecasting, independent method consists of estimating a set of H forecasting networks, each

returning a direct forecasting of Q_{t+h} with $h \in \{1, \dots, H\}$. Joint approach replaces the H networks of direct approach with one multi-output networks $\{Q_{t+1}, Q_{t+2}, \dots, Q_{t+H}\}$

MISMO strategy can be expressed as a combination between two approaches. The independent and joint approach can be indeed seen as two distinct instances of the same forecasting approach. In the independent case, the number of forecasting tasks is equal to the size of the horizon H , and the size of the outputs is 1. In the joint case, the number of forecasting tasks is equal to one, and the size of the output is H , respectively. Intermediate configurations can be considered by transforming the original task into $n = H/s$ forecasting tasks, each with multiple outputs of size s , where $s \in \{1, \dots, H\}$. MISMO trades off the property of preserving the stochastic dependency between future values with a greater flexibility of the predictor (Ben Taieb et al. 2010).

In this study, for the multi-day-ahead streamflow forecasting utilizing the FFNNs-GA and WFFNNs-GA models, three evolutionary strategies, including multi-input multi-output (MIMO), multi-input single-output (MISO), and multi-input several multi-output (MISMO), were applied. In this category, the MISO strategy was employed utilizing independent method (Fig. 1a), and the MIMO and MISMO strategies were suggested utilizing joint method (Fig. 1b) (Kline 2004).

FFNNs-GA and WFFNNs-GA models

Evolutionary algorithms [e.g., genetic algorithm (GA), evolutionary strategy (ES), and programming (EP)] are a class of stochastic searches and optimization techniques (Ding et al. 2013). This study aims to take advantage of GA

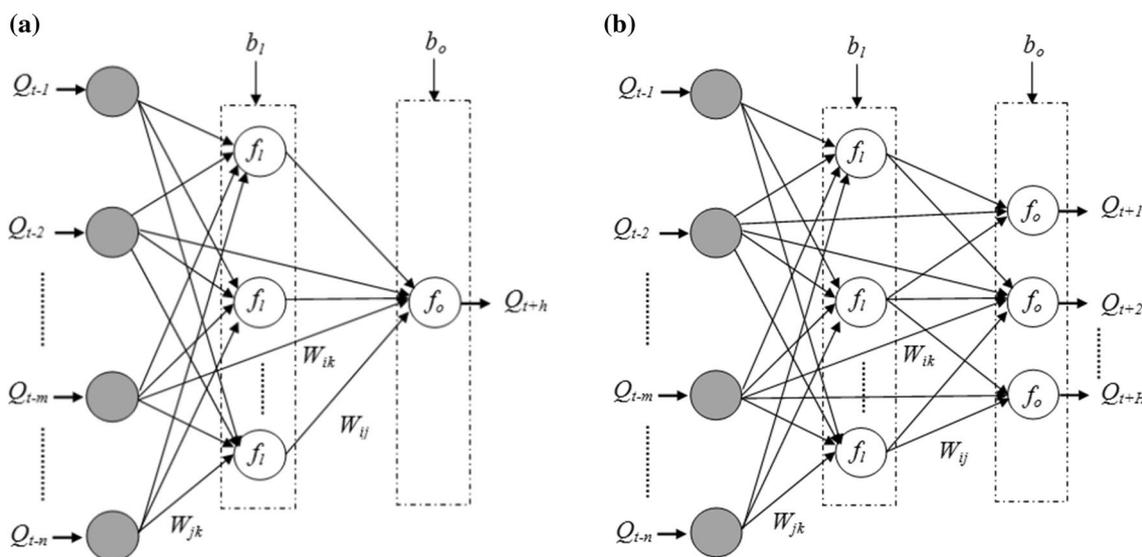


Fig. 1 Conventional FFNNs architecture a independent method, b joint method

properties including selection, crossover, and mutation to find the global minimum errors (Kim and Kim 2008; Jain and Srinivasulu 2004; Holland 1992; Goldberg and Deb 1991; Goldberg and Holland 1988).

Therefore, a method for designing the FFNNs and WFFNNs models' structures utilizing GA was proposed to construct the best models. The GA and BP algorithms were applied to optimize the FFNNs and WFFNNs model. In addition, the ability of BP algorithm can find local minimum errors. For this purpose, the steps can be described as follows:

Step 1 Encoding

The different parameters of WFFNNs model are coded by a chromosome model developed from a series of genes as shown in Fig. 2. Each gene represents the input delay (D), type of the mother wavelet (TMW), decomposition levels (L), number of neurons in the hidden layers (NHL1 and NHL2), activation functions in hidden and output layers (AFHL1, AFHL2, and AFOL), and the initial connection weights and bias coefficients (IWB), respectively.

This study compared the effects of 20 selected wavelet functions from the most frequently utilized mother wavelets [e.g., Haar (Har), Daubechies (Db), Coiflets (Coif), Symmlets (Sym), and Biorthogonal (Bior)]. For analysis of DWT method, Daubechies, Symmlets, and Coiflets wavelets have been commonly utilized as mother wavelets in wavelet-based hydrologic studies (Seo et al. 2015; Evrendilek 2014; Santos et al. 2014; Adamowski and Sun 2010; Tiwari and Chatterjee 2010). In addition, the effects of activation functions for the accuracy of WFFNNs-GA model were demonstrated utilizing the four activation functions, including linear transfer function (Purelin), symmetric saturating linear transfer function (Satlins), log-sigmoid transfer function (Logsig), and hyperbolic tangent sigmoid transfer function (Tansig), respectively.

Step 2 Initializing random individual topology (chromosomes).

Step 3 Training each individual by BP algorithms utilizing k-fold cross-validation.

Step 4 Calculating the fitness values

With regard to the fitness function, it is based on the mean squared error (MSE) over a training dataset. With k-fold cross-validation method, datasets are divided into *k* parts. Note that each dataset should be divided into two parts of the calibration and validation phases. In each run, one fold of dataset is allocated for the validation phase, and *k-1* folds are allocated for calibrating the model. This process is repeated *k* times. Real error of this model can be estimated by averaging the error of *k* runs of the model, which is represented by Eq. (1)

$$E = \frac{1}{K} \sum_{i=1}^k MSE_i \tag{1}$$

where MSE_i is the mean squared error for each $i = 1, 2, \dots, k$. In case of this study, the number of folds is equal to 10.

Step 5 Selecting the best individual mechanism.

Step 6 Crossover and mutation operators.

Step 7 Replacing the current population by the newly generated offsprings.

Step 8 Stopping criteria; otherwise, go to step 3.

Case study and data

Description of study watershed

The selection of appropriate meteorological data is one of the most addressed issues for streamflow forecasting (Ghorbani et al. 2018; Karimi et al. 2018; Seo et al. 2018;

Fig. 2 Chromosome encoding for WFFNNs model

Genes	Description	Values
g ₁	D	{1,2, ..., 14}
g ₂	TMW	{Har, Db(4...9), Sym(3... 8), Coif(2... 5), Bior(1.3, 1.5, 2.2)}
g ₃	L	{1,2, ..., 6}
g ₄	NHL1	{2, 3, ..., 10}
g ₅	NHL2	{2, 3, ..., 10}
g ₆	AFHL1	{Satlins, Purelin, Logsig, Tansig}
g ₇	AFHL2	{Satlins, Purelin, Logsig, Tansig}
g ₈	AFOL	{Satlins, Purelin, Logsig, Tansig}
g ₉	IWB	[-1 1]
g ₁₀		[-1 1]
⋮		⋮
g _n		[-1 1]

Zakhrouf et al. 2018; Dariane and Azimi 2016; Kalteh 2015; Seo et al. 2015; Kisi et al. 2014; Nourani et al. 2014; Sahay and Srivastava 2014; Asadi et al. 2013; Badrzadeh et al. 2013; Danandeh Mehr et al. 2013; Nourani et al. 2013; Kisi and Shiri 2012; Guo et al. 2011). Installing the standard meteorological stations can collect the optimal meteorological data (Jajarmizadeh et al. 2016).

The Chellif River basin (latitude $33^{\circ} 50' - 36^{\circ} 50' N$ and longitude $0^{\circ} 00' - 3^{\circ} 50' E$) is located in the northwestern of Algeria which area is $43,700 \text{ km}^2$. The elevation is characterized from 20 to 1983 m altitude, and the perimeter is 1383 km. In addition, the compactness index is 1.85, and the length of its principal stream is 759 m as shown in Fig. 3. The meteorological characteristics in the Chellif River basin can be expressed as Mediterranean climate with summer (e.g., hot and dry) and winter (e.g., cool and tempered) with wind speed less than 10 km/h and a high average sunshine corresponding to 60–80% of the day duration (Bouchelkia et al. 2014).

The datasets (14 years) were collected from Ponteba Defluent station (01 22 03) of the National Agency of Water Resources. The first 10-year data (70% of the whole dataset) were employed for the model calibration, and the remaining 4 years (30% of the whole dataset) were applied for the

model validation in this study (Fig. 4). The size and rapid slope of the Chellif River basin can restrict the ranges of multi-time-ahead streamflow as 1-, 2-, 3-, and 4-day-ahead forecasting. Also, the authors applied three evolutionary strategies as MISO (four model), MISMO (two models), and MIMO (one model), respectively.

Measures of accuracy

To assess the performance of the stand-alone and hybrid models for streamflow forecasting based on the different strategies during the calibration and validation phases, five statistical indices are presented in Table 1, including root mean squared error (RMSE), signal-to-noise ratio (SNR), correlation coefficient (CC), Nash–Sutcliffe efficiency (NSE), and peak flow criteria (PFC).

In forecasting models, the discrepancy between observed and forecasted values can be shown utilizing RMSE index. This metric would vary from zero to a large value which presents perfect forecasting by the difference between observed and forecasted values. NSE index is taken into account to evaluate the ability of hydrological models. The higher value of it demonstrates a perfect fit between observed and forecasted streamflow (Nash and Sutcliffe 1970).

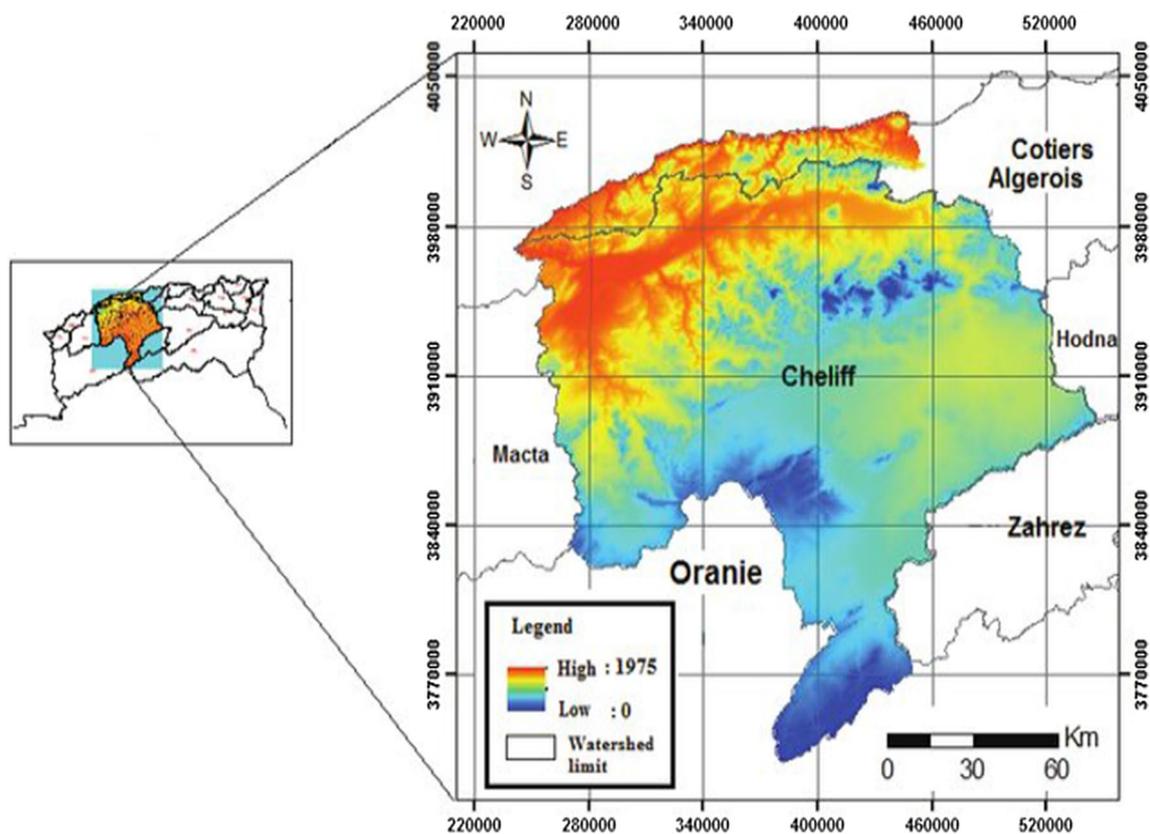


Fig. 3 The Chellif River basin (Malika et al. 2018)

Fig. 4 Observed streamflow hydrograph (14 years)

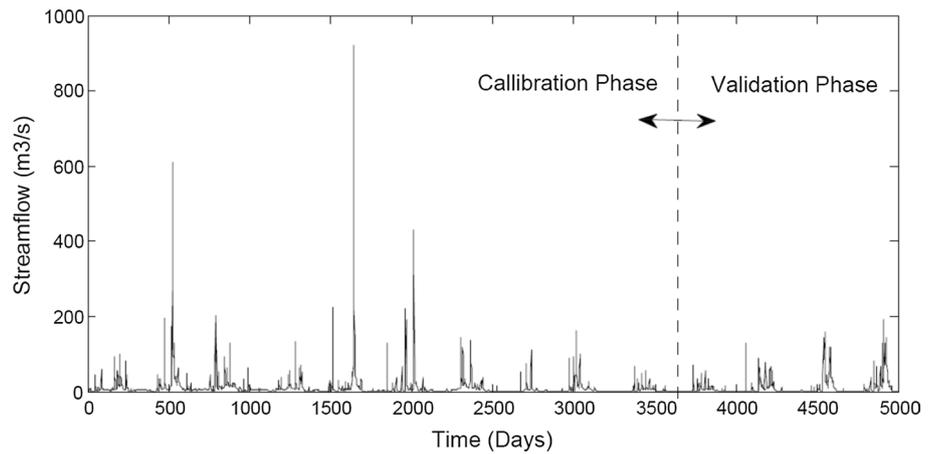


Table 1 Performance evaluation indices

Evaluation indices	Definition
Root mean squared error	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [Q_{t_i} - \hat{Q}_{t_i}]^2}$
Signal-to-noise ratio	$SNR = \frac{\sqrt{\frac{1}{\gamma} \sum_{i=1}^N [Q_{t_i} - \hat{Q}_{t_i}]^2}}{\delta}$
Correlation coefficient	$CC = \frac{\sum_{i=1}^N [Q_{t_i} - \bar{Q}_t][\hat{Q}_{t_i} - \bar{\hat{Q}}_t]}{\sqrt{\sum_{i=1}^N [Q_{t_i} - \bar{Q}_t]^2 \sum_{i=1}^N [\hat{Q}_{t_i} - \bar{\hat{Q}}_t]^2}}$
Nash–Sutcliffe efficiency	$NSE = \left(1 - \frac{\sum_{i=1}^N [Q_{t_i} - \hat{Q}_{t_i}]^2}{\sum_{i=1}^N [Q_{t_i} - \bar{Q}_t]^2} \right) \times 100$
Peak flow criteria	$PFC = \frac{\sqrt[4]{\sum_{i=1}^{Tp} [Q_{t_i} - \hat{Q}_{t_i}]^2 \cdot Q_{t_i}^2}}{\sqrt{\sum_{i=1}^{Tp} Q_{t_i}^2}}$

Q_{t_i} =the observed streamflow value, \hat{Q}_{t_i} =the forecasted streamflow value, \bar{Q}_t =the mean observed streamflow, $\bar{\hat{Q}}_t$ =the mean forecasted streamflow, N =the number of data, γ =the number of degrees of freedom, and δ =the standard deviation of the observed streamflow, Tp =the number of peak streamflow greater than one-third of the observed mean peak streamflow

SNR index can be defined as the ratio of meaningful information to the unwanted one. It enables the comparison of modeling uncertainty to the effects of hydrologic scenarios. If SNR index is equal to zero, it indicates the perfect performance between observed and forecasted streamflow. However, the higher value of it explains the unacceptable performance (Bormann 2005).

CC index is a measure of the accuracy of a hydrologic modeling and is utilized for comparisons of alternative models. A high CC value addresses a good model performance, and vice-versa. A perfect match between the observed and forecasted streamflow yields $CC=1.0$ (Kim and Kim 2008).

NSE index can range from negative infinite ($-\infty$) to 100%. If NSE equal to 100%, it corresponds to a perfect fits of forecasted streamflow. If NSE index is equal to zero, it has no more accurate forecasting, and the model performance cannot be acceptable. If NSE can range from minus

100% to zero, it indicates the unacceptable performance. The addressed statistical indices (e.g., RMSE, SNR, CC, and NSE) may not illustrate the models’ performance in case of the extreme events (e.g., flood and drought) of streamflow. That is, one model may forecast the mean values of streamflow accurately, but cannot forecast the high and low values of streamflow. As a result, the statistical indices such as NSE, RMSE, Willmott’s index of agreement (WI) (Willmott 1984), and Legates–McCabe’s index (LMI) (Legates and McCabe 1999) cannot present an adequate diagram. Therefore, it is essential to assess and monitor the extreme values utilizing PFC index for forecasting the extreme events. PFC index plays a significant role in monitoring the extreme events including flood to achieve the efficient model. It is noticeable to say that PFC equal to zero represents a perfect fit of model.

Results and discussion

Table 2 presents the optimal structure of WFFNNs-GA model based on the MISO strategy in this study. It can be found from Table 2 that the variables of WFFNNs-GA model were demonstrated for the multi-day-ahead (e.g., 1-, 2-, 3-, and 4-day-ahead) forecasting, respectively. For example, the variables of WFFNNs-GA model with 1-day-ahead streamflow forecasting were determined as follows; input delay = 12, the type of mother wavelet = Db9, decomposition level = 4, number of neuron in the first hidden layer = 9, number of neuron in the second hidden layer = 5, activation function in the first hidden layer = log-sigmoid transfer function, activation function in the second hidden layer = linear transfer function, and activation function in output layer = linear transfer function, respectively.

Table 3 proposes the optimal structure of WFFNNs-GA model based on the MISMO strategy in this study. It can be seen from Table 3 that the variables of WFFNNs-GA model with 1- and 2-day-ahead streamflow forecasting were

Table 2 The optimal structure for the WFFNNs-GA model based on the MISO evolutionary strategy

Variables	Model ($t+1$)	Model ($t+2$)	Model ($t+3$)	Model ($t+4$)
D	12	11	12	11
MWT	Db9	Coif5	Db6	Db9
L	4	5	4	5
NHL1	9	6	6	10
NHL2	5	5	4	8
AFHL1	Logsig	Tansig	Logsig	Tansig
AFHL2	Purelin	Logsig	Satlins	Logsig
AFOL	Purelin	Purelin	Purelin	Purelin

provided as follows: input delay = 12, the type of mother wavelet = Db9, decomposition level = 5, number of neurons in the first hidden layer = 9, number of neurons in the second hidden layer = 2, activation function in the first hidden layer = symmetric saturating linear transfer function, activation function in the second hidden layer = symmetric saturating linear transfer function, and activation function in output layer = symmetric saturating linear transfer function, respectively.

Table 4 provides the optimal structure of WFFNNs-GA model based on the MIMO strategy in this study. It can be judged from Table 4 that the variables of WFFNNs-GA model with 1-, 2-, 3-, and 4-day-ahead streamflow forecasting were suggested as follows: input delay = 10, the type of mother wavelet = Sym7, decomposition level = 4, number of neuron in first hidden layer = 4, number of neuron in second hidden layer = 6, activation function in the first hidden layer = log-sigmoid transfer function, activation function in the second hidden layer = symmetric saturating linear transfer function, and activation function in output layer = symmetric saturating linear transfer function, respectively. Tables 2, 3, and 4 explain that the wavelet transform utilizing Daubechies 9 (Db9) contributed the selection of optimal model's structure based on three evolutionary strategies (i.e., MISO, MIMO, and MISMO) and improved the

Table 3 The optimal structure for the WFFNNs-GA model based on the MISMO evolutionary strategy

Variables	Model ($t+1, t+2$)	Model ($t+3, t+4$)
D	12	11
MWT	Db9	Db9
L	5	6
NHL1	9	2
NHL2	2	6
AFHL1	Satlins	Logsig
AFHL2	Satlins	Satlins
AFOL	Satlins	Satlins

Table 4 The optimal structure for the WFFNNs-GA model based on the MIMO evolutionary strategy

Variables	Model ($t+1, t+2, t+3, t+4$)
D	10
MWT	Sym7
L	4
NHL1	4
NHL2	6
AFHL1	Logsig
AFHL2	Satlins
AFOL	Satlins

efficiency of the FFNNs model consequently. This finding followed the recent literature of Seo et al. (2015) closely. Also, GA determined all of parameters including wavelet transform that could yield the best performance without the intervention of human activity.

Table 5 suggests the performance of WFFNNs-GA model based on the three evolutionary strategies (i.e., MISO, MIMO, and MISMO) for the multi-day-ahead (e.g., 1-, 2-, 3-, and 4-day-ahead) streamflow forecasting. For the multi-day-ahead streamflow forecasting, the calibration and validation phases of WFFNNs-GA model, the forecasting results of the 1-day-ahead streamflow yielded the best performance based on RMSE, SNR, CC, NSE, and PFC indices for three evolutionary strategies (i.e., MISO, MIMO, and MISMO), respectively. As the multi-time-ahead streamflow forecasting for three evolutionary strategies (i.e., MISO, MIMO, and MISMO) were varied from 1- to 4-day-ahead one by one, it can be seen that the model accuracy decreased definitely. Also, it can be concluded that the 1-day-ahead streamflow forecasting can produce the least errors than the other multi-day-ahead (e.g., 2-, 3-, and 4-day-ahead) in this study. Table 5 produces that the values of RMSE (e.g., 1.550, 4.659, 6.414, and 6.707 m³/sec) and SNR (e.g., 0.066, 0.198, 0.273, and 0.285) indices for the WFFNNs-GA model based on the MISMO evolutionary strategy were lower than those of the WFFNNs-GA model based on the MISO and MIMO evolutionary strategies for the validation phase. In addition, the values of NSE and CC indices for the WFFNNs-GA model based on the MISMO evolutionary strategy were higher than those of the WFFNNs-GA model based on the MISO and MIMO evolutionary strategies. It can be judged that the results of the WFFNNs-GA model based on the MISMO evolutionary strategy were superior to those of the WFFNNs-GA model based on the MISO and MIMO evolutionary strategies definitely. Since the structure of MISMO evolutionary strategy is more complex and complicated than that of MISO and MIMO evolutionary strategies, it is difficult to converge the global minimum errors utilizing few iterations numbers of MISMO evolutionary strategy for the

Table 5 Statistical results of WFFNNs-GA model based on the MISO, MIMO, and MISMO evolutionary strategies in the (a) calibration and (b) validation phases

		RMSE (m ³ /s)	SNR	CC	NSE (%)	PFC
<i>(a) Calibration phase</i>						
MISO	<i>t</i> +1	1.491	0.047	0.999	99.783	0.053
	<i>t</i> +2	3.687	0.115	0.994	98.672	0.186
	<i>t</i> +3	6.584	0.206	0.979	95.766	0.201
	<i>t</i> +4	5.222	0.163	0.987	97.337	0.052
MIMO	<i>t</i> +1	3.389	0.106	0.995	98.878	0.206
	<i>t</i> +2	6.232	0.195	0.981	96.206	0.229
	<i>t</i> +3	7.441	0.233	0.973	94.592	0.245
	<i>t</i> +4	6.520	0.204	0.979	95.847	0.169
MISMO	<i>t</i> +1	2.047	0.064	0.998	99.591	0.131
	<i>t</i> +2	6.805	0.213	0.978	95.477	0.179
	<i>t</i> +3	5.573	0.174	0.985	96.967	0.199
	<i>t</i> +4	5.913	0.185	0.983	96.585	0.136
<i>(b) Validation phase</i>						
MISO	<i>t</i> +1	1.576	0.067	0.998	99.550	0.072
	<i>t</i> +2	6.820	0.290	0.959	91.579	0.199
	<i>t</i> +3	7.553	0.321	0.947	89.670	0.190
	<i>t</i> +4	24.355	1.036	0.714	−7.398	0.315
MIMO	<i>t</i> +1	3.448	0.147	0.989	97.848	0.121
	<i>t</i> +2	6.236	0.265	0.966	92.958	0.179
	<i>t</i> +3	7.660	0.326	0.953	89.376	0.198
	<i>t</i> +4	11.017	0.469	0.883	78.023	0.250
MISMO	<i>t</i> +1	1.550	0.066	0.998	99.565	0.081
	<i>t</i> +2	4.659	0.198	0.981	96.070	0.135
	<i>t</i> +3	6.414	0.273	0.962	92.552	0.172
	<i>t</i> +4	6.707	0.285	0.958	91.856	0.164

calibration phase. However, the calibration phase utilizing the sufficient iterations number can improve the accuracy of WFFNNs-GA model utilizing MISMO strategy in this study.

Table 6 shows the performance of FFNNs-GA models based on the three evolutionary strategies (i.e., MISO, MIMO, and MISMO) for the multi-day-ahead (e.g., 1-, 2-, 3-, and 4-day-ahead) streamflow forecasting. For the calibration and validation phases of FFNNs-GA model based on the multi-day-ahead forecasting, the results of FFNNs-GA model based on the MISO evolutionary strategy were superior to those of FFNNs-GA model based on the MIMO and MISMO evolutionary strategies for 1- and 2-day-ahead forecasting. However, the results of FFNNs-GA model based on the MISMO evolutionary strategy were superior to those of FFNNs-GA model based on the MISO and MIMO evolutionary strategies for 3- and 4-day-ahead forecasting. In addition, since the values of NSE index in the FFNNs-GA model based on the MIMO evolutionary strategy produced the negative values, they presented the unacceptable performance in the validation phase. Comparison between Tables 5 and 6 revealed that the results of WFFNNs-GA model were superior to those of FFNNs-GA for corresponding evolutionary strategy, respectively. In addition, because

sub-time series captured the high variation that existed into the original series of inputs data, it can be seen that utilizing sub-time series decomposed by DWT method as input data of WFFNNs-GA model can improve the performance of FFNNs-GA model clearly.

Figure 5 presents the scatter diagrams for the WFFNNs-GA model based on the evolutionary MISO, MIMO, and MISMO strategies, respectively. The straight line in Fig. 5 represents best-fit line equation. It was clearly seen from the best-fit line and R^2 values that the WFFNNs-GA model with 1-day-ahead based on the three evolutionary strategies (i.e., MISO, MIMO, and MISMO) could forecast the daily streamflow better than other multi-step-ahead (e.g., 2-, 3-, and 4-day-ahead) forecasting clearly. It can be also judged that the WFFNNs-GA model based on the MISMO evolutionary strategy can forecast the daily streamflow better than the corresponding WFFNNs-GA model based on the MISO and MIMO evolutionary strategies from the viewpoint of the best-fit line and R^2 values.

Figure 6 produces the scatter diagrams for the FFNNs-GA model based on the evolutionary MISO, MIMO, and MISMO strategies, respectively. It can be seen from the best-fit line and R^2 values that the FFNNs-GA model with 1- and

Table 6 Statistical results of FFNNs-GA model based on the MISO, MIMO, and MISMO evolutionary strategies in the (a) calibration and (b) validation phases

		RMSE (m ³ /s)	SNR	CC	NSE (%)	PFC
<i>(a) Calibration phase</i>						
MISO	<i>t</i> +1	17.113	0.535	0.846	71.395	0.446
	<i>t</i> +2	21.302	0.666	0.747	55.676	0.506
	<i>t</i> +3	27.481	0.859	0.513	26.235	0.657
	<i>t</i> +4	22.592	0.706	0.715	50.149	0.451
MIMO	<i>t</i> +1	17.819	0.602	0.800	63.986	0.404
	<i>t</i> +2	19.279	0.602	0.800	63.695	0.348
	<i>t</i> +3	24.518	0.766	0.647	41.284	0.555
	<i>t</i> +4	25.872	0.808	0.596	34.623	0.505
MISMO	<i>t</i> +1	17.389	0.543	0.841	70.463	0.395
	<i>t</i> +2	24.370	0.762	0.648	41.990	0.582
	<i>t</i> +3	27.508	0.860	0.511	26.090	0.651
	<i>t</i> +4	27.809	0.869	0.496	24.465	0.642
<i>(b) Validation phase</i>						
MISO	<i>t</i> +1	11.486	0.489	0.879	76.113	0.243
	<i>t</i> +2	15.301	0.651	0.760	57.610	0.290
	<i>t</i> +3	17.727	0.754	0.685	43.100	0.298
	<i>t</i> +4	39.643	1.686	0.322	−184.5	0.314
MIMO	<i>t</i> +1	59.780	2.543	0.593	−547.0	0.540
	<i>t</i> +2	68.370	2.908	0.458	−746.3	0.525
	<i>t</i> +3	37.009	1.574	0.488	−147.9	0.377
	<i>t</i> +4	45.122	1.919	0.435	−268.6	0.389
MISMO	<i>t</i> +1	22.847	0.972	0.814	5.485	0.350
	<i>t</i> +2	20.763	0.883	0.717	21.941	0.311
	<i>t</i> +3	17.106	0.728	0.693	47.018	0.300
	<i>t</i> +4	17.736	0.754	0.657	43.043	0.305

2-day-ahead based on the MISO evolutionary strategy can forecast the daily streamflow better than the corresponding FFNNs models based on the MIMO and MISMO evolutionary strategies. It can be also derived, however, that FFNNs-GA model with 3- and 4-day-ahead utilizing the MISMO evolutionary strategy can forecast the daily streamflow better than the corresponding FFNNs-GA models based on the MISO and MIMO evolutionary strategies from the viewpoint of the best-fit line and R^2 values, respectively.

In the current research, the authors developed a heuristic model combined a data pre-processing technique and an optimization algorithm based on three evolutionary strategies (i.e., MISO, MIMO, and MISMO). It can be found that this combination can be applied the complex and non-stationary natural phenomenon including the non-linear hydrologic time series (e.g., rainfall, groundwater, evaporation, and water stage etc.). In addition, an addressed attempt to forecast the daily streamflow based on three evolutionary strategies (i.e., MIMO, MISO, and MISMO) is unique process compared to the previous researches following the category of multi-day-ahead streamflow forecasting (Seo et al. 2018; Zakhrouf et al. 2018; Abdollahi et al. 2017; Ravansalar et al. 2016, 2017; Zakhrouf et al.

2016; Kalteh 2015; Seo et al. 2015; Nourani et al. 2014; Sahay and Srivastava 2014). The forecasted results can be accepted based on the diverse statistical indices and scatter diagrams. In addition, the various heuristic models (e.g., multilayer perceptron (MLP), generalized regression neural networks (GRNNs), support vector machines (SVMs), extreme learning machines (ELMs), multivariate adaptive regression spline (MARS), and genetic programming (GP) etc.) combined data pre-processing techniques and optimization algorithms based on evolutionary strategies can be proposed to confirm the accuracy and efficiency of multi-step-ahead streamflow forecasting for the further researches.

Conclusion

This study suggests the wavelet-based feed forward neural networks (WFFNNs) model optimized utilizing genetic algorithm (GA) based on the three evolutionary strategies [i.e., multi-input multi-output (MIMO), multi-input single-output (MISO), and multi-input several multi-output (MISMO)], for streamflow forecasting in the Chellif River basin, Algeria. The WFFNNs-GA model based on the three

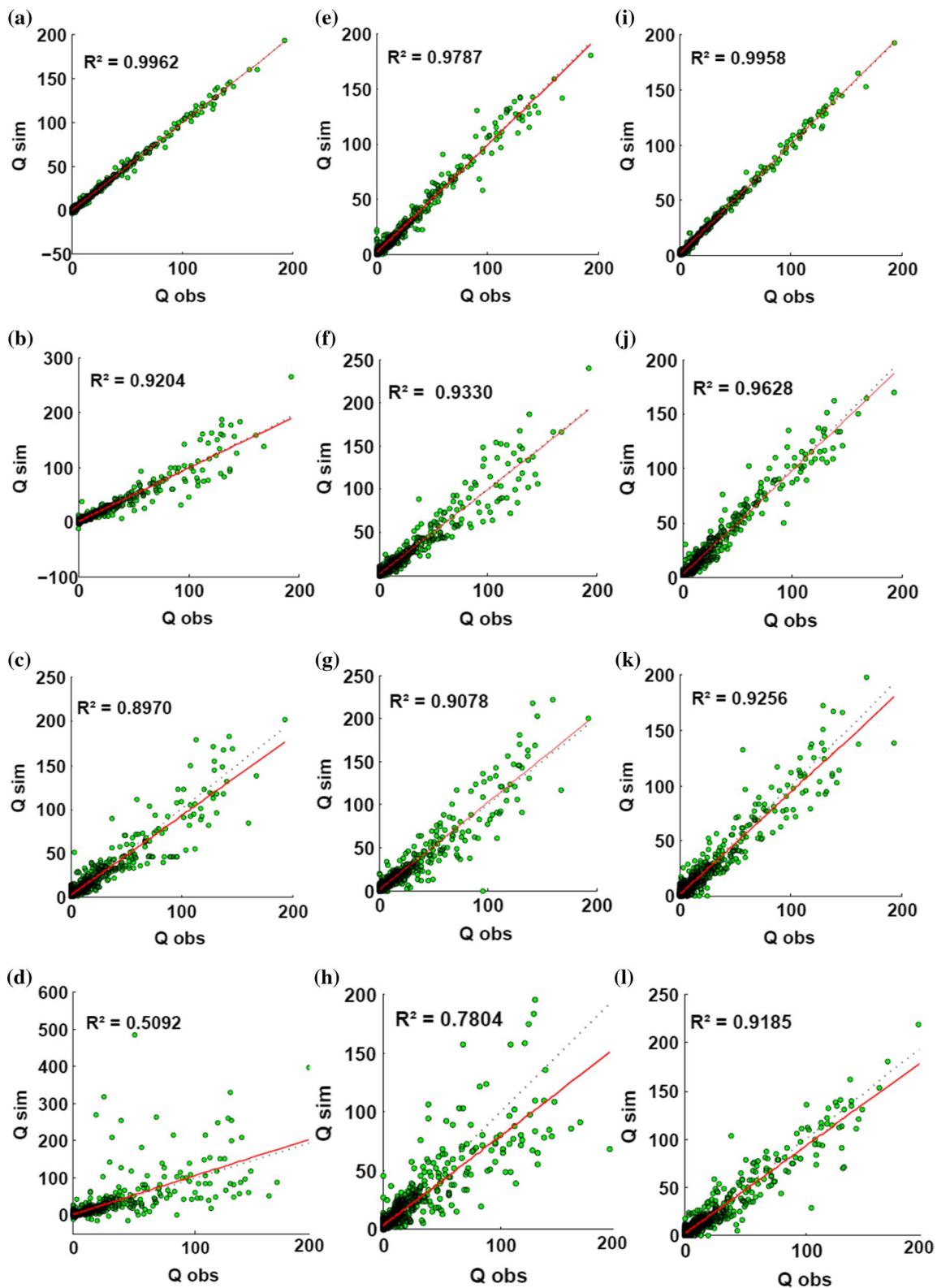


Fig. 5 Scatter diagrams for the WFNNs-GA model in the validation phase **a** MISO ($t+1$), **b** MISO ($t+2$), **c** MISO ($t+3$), **d** MISO ($t+4$), **e** MIMO ($t+1$), **f** MIMO ($t+2$), **g** MIMO ($t+3$), **h** MIMO ($t+4$), **i** MISMO ($t+1$), **j** MISMO ($t+2$), **k** MISMO ($t+3$), **l** MISMO ($t+4$)

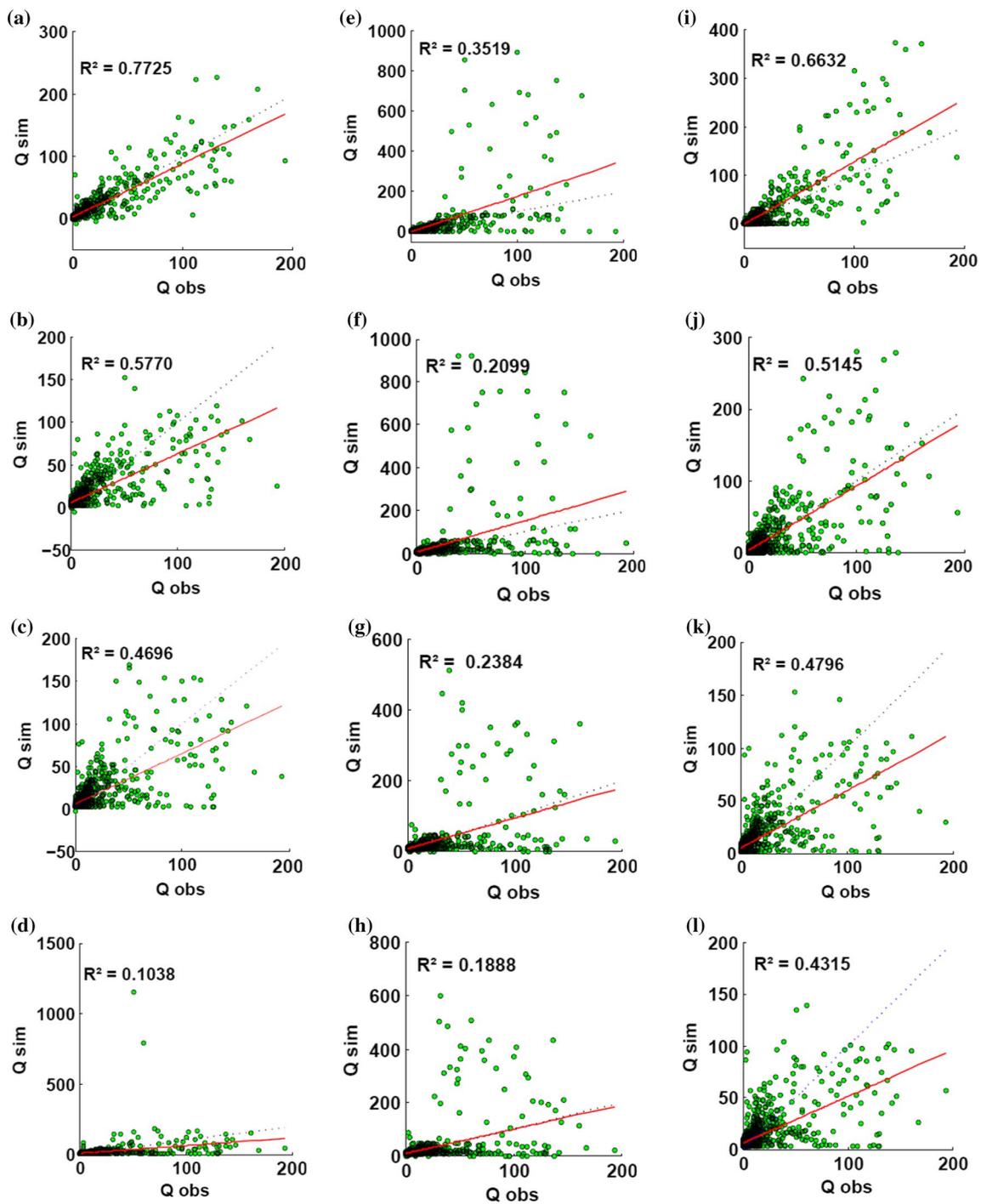


Fig. 6 Scatter diagrams for the FFNNs-GA model in the validation phase **a** MISO ($t+1$), **b** MISO ($t+2$), **c** MISO ($t+3$), **d** MISO ($t+4$), **e** MIMO ($t+1$), **f** MIMO ($t+2$), **g** MIMO ($t+3$), **h** MIMO ($t+4$), **i** MISMO ($t+1$), **j** MISMO ($t+2$), **k** MISMO ($t+3$), **l** MISMO ($t+4$)

evolutionary strategies evaluates the forecasting accuracy for multi-step-ahead (e.g., one-, two-, three-, and four-step-ahead) streamflow utilizing five statistical indices including root mean squared error (RMSE), signal-to-noise ratio (SNR), correlation coefficient (CC), coefficient of efficiency (NSE), and peak flow criteria (PFC).

The optimal structure of WFFNNs-GA model is categorized utilizing the input delay, the type of mother wavelet, decomposition level, number of neuron in hidden layers, and activation functions in the hidden, and output layers, respectively. Also, the optimal values of each variable are determined utilizing the calibration dataset.

It can be concluded that the WFFNNs-GA model based on the three evolutionary strategies produces the best results with 1-day-ahead streamflow forecasting. Within the category of 1-day-ahead streamflow forecasting, the WFFNNs-GA model based on the MISO evolutionary strategy provides the values of RMSE = 1.576 (m³/sec), SNR = 0.067, CC = 0.998, NSE = 99.550 (%), and PFC = 0.072 for the validation phase. In the MIMO evolutionary strategy, it gives the values of RMSE = 3.443 (m³/sec), SNR = 0.146, CC = 0.990, NSE = 97.853 (%), and PFC = 0.119 for the validation phase. In addition, it also yields the values of RMSE = 1.550 (m³/sec), SNR = 0.066, CC = 0.998, NSE = 99.565 (%), and PFC = 0.081 within the MISMO evolutionary strategy for the validation phase. It can be suggested from the WFFNNs-GA model that the results of MISMO evolutionary strategy are superior to those of MISO and MIMO evolutionary strategies based on the statistical indices and scatter diagrams. The WFFNNs-GA model based on the MISMO evolutionary strategy can forecast the accurate and efficient streamflow compared with the other WFFNNs-GA models based on the MISO and MIMO evolutionary strategies in this study.

Results also indicate that the application of sub-time series decomposed by DWT as the input data of WFFNNs-GA model can improve the performance of FFNNs-GA models, and forecast the accurate streamflow in Chellif River basin, Algeria.

Acknowledgements The authors would like to reveal our extreme appreciation and gratitude to the National Agency of Water Resources (NAWR), Algeria. This is with regard to providing the metrological information.

Compliance with ethical standard

Conflict of interest The authors declare that we have no conflict of interest.

References

- Abdollahi S, Raeisi J, Khalilianpour M, Ahmadi F, Kisi O (2017) Daily mean streamflow prediction in perennial and non-perennial rivers using four data driven techniques. *Water Resour Manage* 31(15):4855–4874. <https://doi.org/10.1007/s11269-017-1782-7>
- Adamowski J, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J Hydrol* 390(1–2):85–91. <https://doi.org/10.1016/j.jhydrol.2010.06.033>
- Asadi S, Shahrazi J, Abbaszadeh P, Tabanmehr S (2013) A new hybrid artificial neural networks for rainfall–runoff process modeling. *Neurocomputing* 121:470–480. <https://doi.org/10.1016/j.neucom.2013.05.023>
- Badrzadeh H, Sarukkalige R, Jayawardena AW (2013) Impact of multi-resolution analysis of artificial intelligence models inputs on multi-step ahead river flow forecasting. *J Hydrol* 507:75–85. <https://doi.org/10.1016/j.jhydrol.2013.10.017>
- Baydaroğlu Ö, Koçak K, Duran K (2018) River flow prediction using hybrid models of support vector regression with the wavelet transform, singular spectrum analysis and chaotic approach. *Meteorol Atmos Phys* 130(3):349–359. <https://doi.org/10.1007/s00703-017-0518-9>
- Ben Taieb S, Sorjamaa A, Bontempi G (2010) Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing* 73(10–12):1950–1957. <https://doi.org/10.1016/j.neucom.2009.11.030>
- Benedetto F, Giunta G, Mastroeni L (2015) A maximum entropy method to assess the predictability of financial and commodity prices. *Digit Signal Proc* 46:19–31. <https://doi.org/10.1016/j.dsp.2015.08.001>
- Bishop CM (1995) *Neural networks for pattern recognition*. Oxford University Press, Oxford, pp 116–160
- Bormann H (2005) Evaluation of hydrological models for scenario analyses: signal-to-noise-ratio between scenario effects and model uncertainty. *Adv Geosci* 5:43–48
- Bouchelkia H, Belarbi F, Remini B (2014) Quantification of suspended sediment load by double correlation in the watershed of Chellif (Algeria). *J Water Land Dev* 21:39–46. <https://doi.org/10.2478/jwld-2014-0012>
- Chang FJ, Chiang YM, Chang LC (2007) Multi-step-ahead neural networks for flood forecasting. *Hydrol Sci J* 52(1):114–130. <https://doi.org/10.1623/hysj.52.1.114>
- Chu H, Wei J, Li T, Jia K (2016) Application of support vector regression for mid- and long-term runoff forecasting in “Yellow river headwater” region. *Proc Eng* 154:1251–1257. <https://doi.org/10.1016/j.proeng.2016.07.452>
- Danandeh Mehr A, Kahya E, Olyaei E (2013) Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. *J Hydrol* 505:240–249. <https://doi.org/10.1016/j.jhydrol.2013.10.003>
- Darlane AB, Azimi S (2016) Forecasting streamflow by combination of a genetic input selection algorithm and wavelet transforms using ANFIS models. *Hydrol Sci J* 61(3):585–600. <https://doi.org/10.1080/02626667.2014.988155>
- Delafrouz H, Ghaehri A, Ghorbani MA (2018) A novel hybrid neural network based on phase space reconstruction technique for daily river flow prediction. *Soft Comput* 22(7):2205–2215. <https://doi.org/10.1007/s00500-016-2480-8>
- Deo RC, Tiwari MK, Adamowski JF, Quilty JM (2017) Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. *Stoch Env Res Risk Assess* 31(5):1211–1240. <https://doi.org/10.1007/s00477-016-1265-z>
- Ding S, Li H, Su C, Yu J, Jin F (2013) Evolutionary artificial neural networks: a review. *Artif Intell Rev* 39(3):251–260. <https://doi.org/10.1007/s10462-011-9270-6>
- Evrendilek F (2014) Assessing neural networks with wavelet denoising and regression models in predicting diel dynamics of eddy covariance-measured latent and sensible heat fluxes and evapotranspiration. *Neural Comput Appl* 24(2):327–337. <https://doi.org/10.1007/s00521-012-1240-7>
- Ghaemi A, Rezaie-Balf M, Adamowski J, Kisi O, Quilty J (2019) On the applicability of maximum overlap discrete wavelet transform integrated with MARS and M5 model tree for monthly pan evaporation prediction. *Agric Meteorol* 278:107647. <https://doi.org/10.1016/j.agrformet.2019.107647>
- Ghorbani MA, Khatibi R, Mehr AD, Asadi H (2018) Chaos-based multigene genetic programming: a new hybrid strategy for river flow forecasting. *J Hydrol* 562:455–467. <https://doi.org/10.1016/j.jhydrol.2018.04.054>
- Goldberg DE, Deb K (1991) A comparative analysis of selection schemes used in genetic algorithms. *Found Genet Algorithms* 1:69–93. <https://doi.org/10.1016/B978-0-08-050684-5.50008-2>

- Goldberg DE, Holland JH (1988) Genetic algorithms and machine learning. *Mach Learn* 3(2–3):95–99. <https://doi.org/10.1023/A:1022602019183>
- Gowda CC, Mayya SG (2014) Comparison of back propagation neural network and genetic algorithm neural network for stream flow prediction. *J Comput Environ Sci* 2014:290127. <https://doi.org/10.1155/2014/290127>
- Guo J, Zhou J, Qin H, Zou Q, Li Q (2011) Monthly streamflow forecasting based on improved support vector machine model. *Expert Syst Appl* 38(10):13073–13081. <https://doi.org/10.1016/j.eswa.2011.04.114>
- Haykin S (1999) *Neural networks: a comprehensive foundation*. Prentice-Hall, Upper Saddle River, pp 178–274
- Holland JH (1992) Genetic algorithms. *Sci Am* 267(1):66–73
- Hu T, Wu F, Zhang X (2007) Rainfall–runoff modeling using principal component analysis and neural network. *Hydrol Res* 38(3):235–248. <https://doi.org/10.2166/nh.2007.010>
- Jain A, Srinivasulu S (2004) Development of effective and efficient rainfall–runoff models using integration of deterministic, real-coded genetic algorithms and artificial neural network techniques. *Water Resour Res* 40(4):W04302. <https://doi.org/10.1029/2003WR002355>
- Jajarmizadeh M, Sidek LM, Mirzai M, Alaghmand S, Harun S, Majid MR (2016) Prediction of surface flow by forcing of climate forecast system reanalysis data. *Water Resour Manage* 30(8):2627–2640. <https://doi.org/10.1007/s11269-016-1303-0>
- Kalteh AM (2015) Wavelet genetic algorithm–support vector regression (wavelet GA–SVR) for monthly flow forecasting. *Water Resour Manage* 29(4):1283–1293. <https://doi.org/10.1007/s11269-014-0873-y>
- Karimi S, Shiri J, Kisi O, Xu T (2018) Forecasting daily streamflow values: assessing heuristic models. *Hydrol Res* 49(3):658–669. <https://doi.org/10.2166/nh.2017.111>
- Kim S, Kim HS (2008) Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling. *J Hydrol* 351(3–4):299–317. <https://doi.org/10.1016/j.jhydrol.2007.12.014>
- Kisi O, Shiri J (2012) Reply to discussion of “Precipitation forecasting using wavelet–genetic programming and wavelet–neuro–fuzzy conjunction models”. *Water Resour Manage* 26(12):3663–3665. <https://doi.org/10.1007/s11269-012-0060-y>
- Kisi O, Latifoğlu L, Latifoğlu F (2014) Investigation of empirical mode decomposition in forecasting of hydrological time series. *Water Resour Manage* 28(12):4045–4057. <https://doi.org/10.1007/s11269-014-0726-8>
- Kline DM (2004) Methods for multi-step time series forecasting with neural networks. In: *Neural networks in business forecasting*, IGI Global, pp 226–250. <https://doi.org/10.4018/978-1-59140-176-6.ch012>
- Krishna B, Satyaji Rao YR, Nayak PC (2011) Time series modeling of river flow using wavelet neural networks. *J Water Resour Prot* 3:50–59. <https://doi.org/10.4236/jwarp.2011.31006>
- Legates DR, McCabe GJ (1999) Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 35(1):233–241. <https://doi.org/10.1029/1998WR900018>
- Malika A, Abderrahman H, Aicha L, Laounia N, Habib M (2018) Use of high spatial resolution satellite data for monitoring and characterization of drought conditions in the northwestern Algeria. *Min Sci* 25:85–113. <https://doi.org/10.5277/msc182507>
- Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 11(7):674–693. <https://doi.org/10.1109/34.192463>
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models I: a discussion of principles. *J Hydrol* 10:282–290
- Nason GP (2008) *Wavelet methods in statistics with R*. Springer Science and Business Media, New York, pp 15–78. <https://doi.org/10.1007/978-0-387-75961-6>
- Nourani V, Alami MT, Aminfar MH (2009) A combined neural–wavelet model for prediction of Ligvanchai watershed precipitation. *Eng Appl Artif Intell* 22(3):466–472. <https://doi.org/10.1016/j.engappai.2008.09.003>
- Nourani V, Hosseini BA, Adamowski J, Gebremicheal M (2013) Using self-organizing maps and wavelet transforms for space–time preprocessing of satellite precipitation and runoff data in neural network based rainfall–runoff modeling. *J Hydrol* 476:228–243. <https://doi.org/10.1016/j.jhydrol.2012.10.054>
- Nourani V, Baghanam AH, Adamowski J, Kisi O (2014) Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. *J Hydrol* 514:358–377. <https://doi.org/10.1016/j.jhydrol.2014.03.057>
- Parmar KS, Bhardwaj R (2015) River water prediction modeling using neural networks, fuzzy and wavelet coupled model. *Water Resour Manage* 29(1):17–33. <https://doi.org/10.1007/s11269-014-0824-7>
- Quilty J, Adamowski J (2018) Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. *J Hydrol* 563:336–353. <https://doi.org/10.1016/j.jhydrol.2018.05.003>
- Ravansalar M, Rajae T, Zounemat-Kermani M (2016) A wavelet–linear genetic programming model for sodium (Na+) concentration forecasting in rivers. *J Hydrol* 537:398–407. <https://doi.org/10.1016/j.jhydrol.2016.03.062>
- Ravansalar M, Rajae T, Kisi O (2017) Wavelet–linear genetic programming: a new approach for modeling monthly streamflow. *J Hydrol* 549:461–475. <https://doi.org/10.1016/j.jhydrol.2017.04.018>
- Ravikumar P, Somashekar RK (2017) Principal component analysis and hydrochemical facies characterization to evaluate groundwater quality in Varahi river basin, Karnataka state, India. *Appl Water Sci* 7(2):745–755. <https://doi.org/10.1007/s13201-015-0287-x>
- Rezaie-Balf M, Kisi O (2017) New formulation for forecasting streamflow: evolutionary polynomial regression versus extreme learning machine. *Hydrol Res* 49(3):939–953. <https://doi.org/10.2166/nh.2017.283>
- Rezaie-Balf M, Zahmatkesh Z, Kim S (2017) Soft computing techniques for rainfall–runoff simulation: local non–parametric paradigm versus model classification methods. *Water Resour Manage* 31(12):3843–3865. <https://doi.org/10.1007/s11269-017-1711-9>
- Rezaie-Balf M, Kim S, Fallah H, Alaghmand S (2019) Daily river flow forecasting using ensemble empirical mode decomposition based heuristic regression models: application on the perennial rivers in Iran and South Korea. *J Hydrol* 572:470–485. <https://doi.org/10.1016/j.jhydrol.2019.03.046>
- Sahay RR, Srivastava A (2014) Predicting monsoon floods in rivers embedding wavelet transform, genetic algorithm and neural network. *Water Resour Manage* 28(2):301–317. <https://doi.org/10.1007/s11269-013-0446-5>
- Samsudin R, Saad P, Shabri A (2011) River flow time series using least squares support vector machines. *Hydrol Earth Syst Sci* 15:1835–1852. <https://doi.org/10.5194/hess-15-1835-2011>
- Sang YF, Wang Z, Liu C (2013) Discrete wavelet-based trend identification in hydrologic time series. *Hydrol Process* 27(14):2021–2031. <https://doi.org/10.1002/hyp.9356>
- Santos CAG, Freire PKMM, Silva GBL, Silva RM (2014) Discrete wavelet transform coupled with ANN for daily discharge forecasting into Três Marias reservoir. In: *Proceedings of the international association of hydrological sciences*, Bologna, Italy, pp 100–105. <https://doi.org/10.5194/piahs-364-100-2014>
- Seo Y, Kim S (2016) River stage forecasting using wavelet packet decomposition and data-driven Models. *Proc Eng* 154:1225–1230. <https://doi.org/10.1016/j.proeng.2016.07.439>

- Seo Y, Kim S, Kisi O, Singh VP (2015) Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *J Hydrol* 520:224–243. <https://doi.org/10.1016/j.jhydrol.2014.11.050>
- Seo Y, Kim S, Singh V (2018) Machine learning models coupled with variational mode decomposition: a new approach for modeling daily rainfall-runoff. *Atmosphere* 9(7):251. <https://doi.org/10.3390/atmos9070251>
- Shafaei M, Kisi O (2016) Lake level forecasting using wavelet-SVR, wavelet-ANFIS and wavelet-ARMA conjunction models. *Water Resour Manage* 30(1):79–97. <https://doi.org/10.1007/s11269-015-1147-z>
- Shoaib M, Shamseldin AY, Melville BW, Khan MM (2015) Runoff forecasting using hybrid wavelet gene expression programming (WGEP) approach. *J Hydrol* 527:326–344. <https://doi.org/10.1016/j.jhydrol.2015.04.072>
- Tayyab M, Zhou J, Dong X, Ahmad I, Sun N (2019) Rainfall-runoff modeling at Jinsha River basin by integrated neural network with discrete wavelet transform. *Meteorol Atmos Phys* 131(1):115–125. <https://doi.org/10.1007/s00703-017-0546-5>
- Tiwari MK, Chatterjee C (2010) Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. *J Hydrol* 394(3–4):458–470. <https://doi.org/10.1016/j.jhydrol.2010.10.001>
- Uysal G, Şorman AU (2017) Monthly streamflow estimation using wavelet-artificial neural network model: a case study on Çamlidere dam basin, Turkey. *Proc Comput Sci* 120:237–244. <https://doi.org/10.1016/j.procs.2017.11.234>
- Wang WC, Chau KW, Cheng CT, Qiu L (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J Hydrol* 374(3–4):294–306. <https://doi.org/10.1016/j.jhydrol.2009.06.019>
- Willmott CJ (1984) On the evaluation of model performance in physical geography. *Spatial statistics and models*. Springer, Dordrecht, pp 443–460
- Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv Adapt Data Anal* 1(1):1–41. <https://doi.org/10.1142/S1793536909000047>
- Yu H, Wen X, Feng Q, Deo RC, Si J, Wu M (2018) Comparative study of hybrid-wavelet artificial intelligence models for monthly groundwater depth forecasting in extreme arid regions, Northwest China. *Water Resour Manage* 32(1):301–323. <https://doi.org/10.1007/s11269-017-1811-6>
- Yuan X, Tan Q, Lei X, Yuan Y, Wu X (2017) Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine. *Energy* 129:122–137. <https://doi.org/10.1016/j.energy.2017.04.094>
- Zakhrouf M, Bouchelkia H, Stamboul M (2016) Neuro-wavelet (WNN) and neuro-fuzzy (ANFIS) systems for modeling hydrological time series in arid areas. A case study: the catchment of Ain Hadjadj (Algeria). *Desalin Water Treat* 57(37):17182–18194. <https://doi.org/10.1080/19443994.2015.1085908>
- Zakhrouf M, Bouchelkia H, Stamboul M, Kim S, Heddad S (2018) Time series forecasting of river flow using an integrated approach of wavelet multi-resolution analysis and evolutionary data-driven models. A case study: sebaou River (Algeria). *Phys Geogr* 39(6):506–522. <https://doi.org/10.1080/02723646.2018.1429245>



A new mathematical model to calculate the equilibrium scour depth around a pier

Ainal Hoque Gazi¹ · Mohammad Saud Afzal¹

Received: 12 October 2019 / Accepted: 25 October 2019 / Published online: 4 November 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

This paper sheds light on the formulation of a new equilibrium local scour depth equation around a pier. The total bed materials removed from the scour hole due to the force exerted by the flowing fluid after colliding with the pier in the flow field are estimated. At the equilibrium condition, the shape of the scour hole around the pier may take any form, viz. linear, circular, parabolic, triangular, or combination of different shapes. To consider that, two functions are assumed at the stoss and the lee sides of the pier. The total volume of bed materials removed from the scour hole of an arbitrary shape at the stoss and the lee sides of the pier is obtained by integrating the two functions. The equilibrium scour depth is formed by applying the energy balance theorem. An example problem is illustrated and the results are compared with the equations presented by Melville and Coleman (Bridge scour. Water Resources Publication, Colorado, 2000) and HEC-18 (Richardson and Davis in Evaluating scour at bridges, HEC-18. Technical report no. FHWA NHI, 2001).

Keywords Analytical solution · Scour · Energy balance · Sediment transport

Introduction

The construction of a bridge involves an enormous amount of cost, and the losses associated with the failure of a bridge are unrecoverable. Harik et al. (1990) did a failure study of bridges in the USA between 1951 to 1988 by collecting the relevant data from various sources, including engineering journals and nationally and state-wise distributed newspapers. Chang (1988) reported the failure of 46 bridges, from 1961 to 1976, because of the scour. Wardhana and Hadipriono (2003) did failure studies of 503 bridges from 1989 to 2001 with the data collected from the New York State Department of Transportation (NYSDOT). Scour accounts for nearly 60% of the total bridge failures in the USA since 1950 (Kattell and Eriksson 1998; Melville and Coleman 2000; Lagasse 2007). The AASHTO (2010) LRFD Bridge Design Specification found that scour is one of the vital reasons behind the bridge failure. In 1993, it was reported that during the flood event at Upper Mississippi, 28 highway

bridges failed, of which 22 were because of scour (Kamoj-jala et al. 1994), which amounts to almost 80%.

Hydraulic Engineering Circular No. 18 (HEC-18) (Richardson and Davis 2001) and Arneson et al. (2012) defined scour as “Erosion of a streambed or bank material due to flowing water; often considered as being localized.” National Highway Institute (NHI), Washington, DC, listed four reasons behind the occurrence of the scour: general scour (natural movement of the sediments resulting in the reduction in the streambed elevation), contraction scour (the acceleration of the incoming flow, due to the narrowing of the channel at certain sections, leading to a higher rate of erosion), local scour (flow turbulence and vorticity, due to the presence of an obstacle in the channel, resulting in the movement of the sediment particles away from the neighborhood of the obstruction), and lateral stream migration (occurring when the presence of hydraulic structures forces the channel to change the flow path). Among these, local scour is the most crucial one. Dey (1997) presented a detailed review of the scour studies up to 1997 with large captivating pieces of information.

A review on the mathematical modeling of the scour was also given by Mutlu Sumer (2007). They, however, mentioned that recently, more advanced computational techniques like computational fluid dynamics (CFD) code are

✉ Mohammad Saud Afzal
saud@civil.iitkgp.ac.in

¹ Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal 721302, India

being used to model phenomenon of scour under steady current. Some of the examples of such studies included but not limited to are Olsen and Melaaen (1993), Nurtjahyo et al. (2002), Salaheldin et al. (2004), Roulund et al. (2005), Bouratsis et al. (2017), Yang et al. (2018), Ahmad et al. (2017), and Afzal et al. (2015).

Chabert and Engeldinger (1956) classified local scour into two categories, i.e., the clear water scour and the live-bed scour. Clear water scour occurs when the flow, responsible for causing the scour, is not fed with sediment load, whereas the live-bed scour accounts when the flow is already fed with sediment particles. The equilibrium scour depth (d_s) is defined as the vertical distance between the deepest point in the scour hole and the stable bed level. In the live-bed scour, the formation of the equilibrium scour depth is rapid, and d_s oscillates with time because of the formation of the bed forms. In clear water scour, the idea of the equilibrium scour is still a mystery (Lança et al. 2013). Kothiyari et al. (2007) argued that the time required of the clear water scour to reach the equilibrium depth is a function of finite random variable, whereas Sheppard et al. (2004) reported the same as a function of an infinite random variable.

Due to the presence of a vertical cylindrical pier, the flow velocity slows down at the vicinity of the piers, and a boundary layer is formed. At the stagnation point, S_1 , in Fig. 1, the velocity is zero (due to the no-slip boundary condition), and the pressure is maximum. The pressure decreases toward the poles, P_1 and P_2 , in Fig. 1 and increases at the rear half of the pier. Thus, a pressure gradient is developed within the boundary layer. A decrease in pressure along the direction of the motion is considered favorable as it helps in the movement of the fluid. However, an increase in pressure along the flow direction (termed as adverse pressure gradient) leads to either cessation of the motion of fluid or flow reversal, which in turn causes the sediment particles to move away from the pier surface.

Sediment particles move away from the pier when the shear stress on the sediment particles exceeds the critical shear stress. Under the steady currents, the most responsible parameters for the development of the scour around a pier,

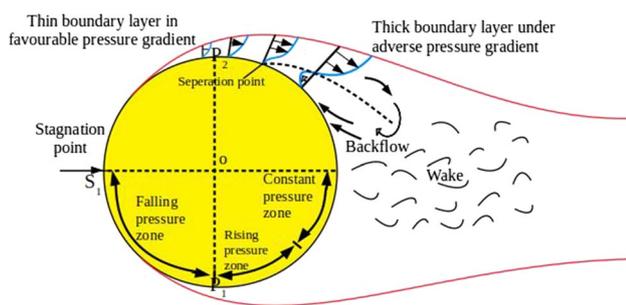


Fig. 1 Top view of the flow field around a pier

identified in the scientific literature, are the downflow, horseshoe vortex (HSV), bow wave, and wake vortices (Fig. 2). The direction of the rotation of the bow wave and the HSV is opposite in a manner, as shown in Fig. 2.

Dey et al. (1992a, b) noticed the development of the HSV because of the motion of the fluid into the scour hole. The authors observed, at the stoss side of the pier, the bed materials are detached by the activity of the lee side flow and consequently are fetched up toward the stoss side of the scour hole by the motion of the HSV. At the quasi-equilibrium condition, the mean slope, also known as the dynamic angle of repose (ϕ_d), of the scour hole at the frontal face of the pier was observed to be 10–20% more than the threshold angle of the sediment. This was in agreement with the result presented by Melville and Raudkivi (1977).

Most of the manuals, providing the guidelines for the design of the bridges, consider the maximum equilibrium scour depth as a reference depth against the extreme flood conditions having return periods in the range of 100–200 years (Melville and Coleman 2000). However, in many cases, this equilibrium scour depth is found to be lesser than the actual scour depth under the flood condition, and the bridge fails. One example is the case study of South Carolina (Shatanawi et al. 2008). Thus, a lot of uncertainty related to the scour calculation is involved in the design of the bridges.

The sediment particles are removed from the vicinity of the pier by the fluid-induced force, which is a combined effect of the bed shear stress, the turbulent properties, and the fluctuations due to the horseshoe vortex as delineated by Dey (1997). Based on the fundamental equation of scour under clear water, Dey (1996) estimated the sediment pickup rate around a cylindrical pier.

Some of the notable studies where flow was measured inside the scour hole are Dey et al. (1995), Dey (1995), Graf and Istiarto (2002), Dey and Raikar (2007), and Raikar and Dey (2005a, b, 2008). Dey et al. (1995) used a five-hole pitot sphere, whereas Graf and Istiarto (2002) used an acoustic

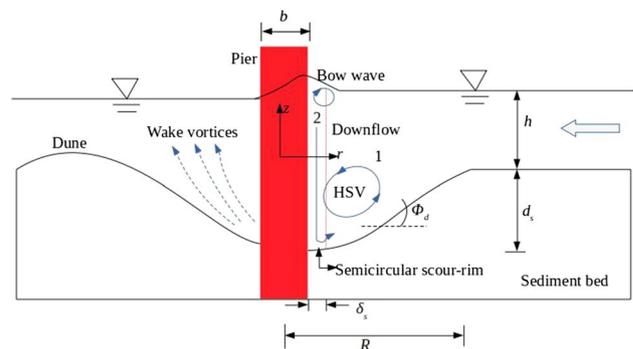


Fig. 2 Typical geometry of a scour hole and the components of the flow field around a pier

Doppler velocity (ADV) meter for the measurements. Dey and Raikar (2007) and Raikar and Dey (2008) also used ADV to calculate the discharge and gave detailed information on the formation of the turbulent horseshoe vortex in the vicinity of a developing scour hole around a cylindrical pier. The effect of the upstream interference toward the downstream pier in terms of the local scour was investigated by Khaple et al. (2017).

A substantial amount of research has been performed to find out the maximum scour depth around a pier or pile base. Most of the researchers developed scour depth formulae by the means of empirical methods, based on dimensional analysis and data fitting, which establish a functional relationship between the non-dimensional parameters. This has been discussed in detail in Melville and Coleman (2000) and Ettema et al. (2011). Recently, Manes and Brocchini (2015) calculated the equilibrium scour depth by the phenomenological concept of turbulence under steady-state conditions. They derived a simple analytical formula for the clear water scour and the live-bed scour with the aid of Kolmogorov theory of turbulence, which was an extended work of Gioia and Bombardelli (2005) and Gioia and Chakraborty (2006). At equilibrium, Manes and Brocchini (2015) assumed that the characteristic length scale of energetic eddies in the scour hole is equal to the depth of the scour hole and presented two equations for the equilibrium scour depth (d_s).

Dey et al. (1995) presented a kinematic model considering horseshoe vortex (HSV) motion in the scour hole. They divided the upstream section of the pier into two zones: zone1 (vertically upward from the inclined bed) and zone2 (vertically upward from the planner bed and close to the pier) (Fig. 2). They expressed the tangential velocity (u_θ) in terms of the power law at the bed, the radial velocity (u_r) as parabolic in the direction of the radius r , and the vertical velocity (u_z) obtained by integrating the continuity equation. A theoretical model of the flow field around a circular pier describing the velocity distribution was reported by Dey and Bose (1994) using aforementioned equations for u_θ , u_r , and u_z . They integrated the Navier–Stokes equation based on the one-seventh power law, analyzed the turbulent boundary layers, and presented a clear water scour model in terms of the bed shear stress. Most of the time, the bed shear stress remains lesser than the critical shear stress for a developing scour hole at an intermediate phase or for a scour hole that has already reached its equilibrium depth (Dey and Bose 1994).

Recently, Hafez (2016) used the energy balance concept to calculate equilibrium scour depth of a scour hole assuming a triangular shape. However, the scour hole pattern around a pier may take any form after the equilibrium condition is reached. It is observed that most of the time, the shape is considered as triangular or parabolic, at both the stoss and the lee sides of the pier. In the present study, an arbitrary

shape is considered, and a general scour depth equation is developed considering two functions for the upstream and the downstream sections. An example problem is formulated and solved, and the results obtained are compared with the existing results.

Present mathematical model

A local scour equation is obtained in this section using a mathematical approach presented by Hafez (2016). The objective, herein, is not to introduce a new equation but to highlight a method for the estimation of the scour depth, capable of producing the nonlinear variations of the scouring pattern. The present mathematical model considers the energy balance concept as described by Hafez (2016). The model provides a significant improvement over the model developed by Hafez (2016) as it can incorporate the scour hole of any shape at the upstream and the downstream of the pier.

The energy balance concept states that the work done by the incoming flow at the stoss side of the pier, due to sudden blockage of water, is the reason behind the removal of the bed materials from the pier surface. When all the energy is utilized for the removal of the bed materials, an equilibrium is reached, and a maximum scour depth (d_s) is established.

The hydrodynamic momentum force (F) can be written as $F = \rho Qu_c$, where ρ is the fluid density, Q is the discharge of the fluid, and u_c is the streamwise flow velocity. The force exerted due to the presence of the pier, F_b , can be expressed by Eq. (1) as

$$F_b = \rho u_c^2 hb \quad (1)$$

A schematic view of the pier at the initial stage when the scour has not started is shown in Fig. 3a, whereas Fig. 3b represents the scour hole shape when the equilibrium condition is reached. Here, h is the water depth, b is diameter of the pier (pier width in case of a rectangular pier), and d_s is the depth of the scour at the equilibrium. As the channel is

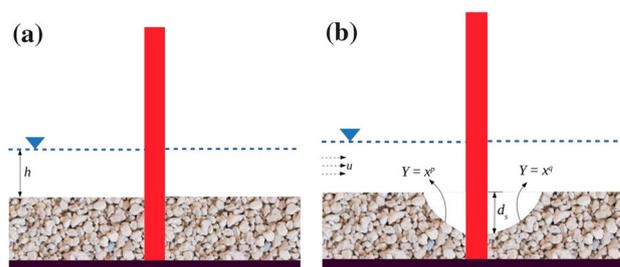


Fig. 3 A schematic view of the development of the scour around a pier: **a** before scouring, **b** after formation of the scour

open to the atmosphere, the velocity u_c at a vertical distance y from the bed can be expressed using the power law as $u_c \propto y^{1/m}$, where m is a real number. Thus, F_b takes the form $F_b \approx y^{\frac{2+m}{m}}$. The point ($y_{c.g}$) at which this force (F_b) acts from the free surface can be calculated by $y_{c.g} = \frac{\int_0^h y F_b dy}{\int_0^h F_b dy}$, which yields Eq. (2).

$$y_{c.g} = \frac{2 + 2m}{2 + 3m} h \tag{2}$$

The displacement of the sediment due to F_b is given by Hafez (2016) as $y_{c.g} + \frac{d_s}{2}$. The work done by the approaching fluid is a dot product of the force and the displacement, which can be denoted as $W_{in} = F_b(y_{c.g} + d_s/2)$, which yields Eq. (3).

$$W_{in} = (\rho u_c^2 h b) \left(\frac{2 + 2m}{2 + 3m} h + \frac{d_s}{2} \right) \tag{3}$$

At the equilibrium condition, the scour hole pattern at the upstream and the downstream may take any form, viz. linear, circular, parabolic, triangular, or combination of different shapes. The top view of a scour hole is circular, and it varies nonlinearly from the bottom most point of the pier to the topmost point at the boundary of the scour hole. To take into account this nonlinearity, y can be expressed by Eqs. (4) and (5), respectively (Fig. 4).

$$y = x^p \tag{4}$$

$$y = x^q \tag{5}$$

Here, $y = x^p$ is at the upstream side and $y = x^q$ is at the downstream side and p and q can be assume any real value. As discussed earlier, the force developed due to the presence of the pier or any other obstruction in the flow field is responsible for the removal of the total bed materials. Total volume of the bed materials removed from the scour hole

can be split into V_1 and V_2 , where V_1 and V_2 are the volume of bed materials removed from the front side and the back side of the pier, respectively. The radius of the scour hole at the upstream and the downstream sides can be obtained from Eqs. (4) and (5) as $r_1 = x = y^{1/p}$ and $r_2 = x = y^{1/q}$, respectively. Now the volume of the scour hole V_1 and V_2 at the front and the back side of the pier can be calculated using Eqs. (6) and (7), respectively, as

$$V_1 = \int_0^{d_s} \pi r_1^2 dy = \frac{\pi p}{2 + p} d_s^{\frac{2+p}{p}} \tag{6}$$

$$V_2 = \int_0^{d_s} \pi r_2^2 dy = \frac{\pi q}{2 + q} d_s^{\frac{2+q}{q}} \tag{7}$$

After the rearrangement of Eqs. (4) and (5), total volume, V_T ($= V_1 + V_2$), takes the form given by Eq. (8).

$$V_T = \frac{\pi p}{2 + p} d_s^{\frac{2+p}{p}} + \frac{\pi q}{2 + q} d_s^{\frac{2+q}{q}} \tag{8}$$

Total sediment particles, having porosity θ (generally considered 0.4 for sand), removed from the scour hole are $V(1 - \theta)$. Hence, the total submerged weight of the removed bed materials from the scour hole assumes the form $V(1 - \theta)(\gamma_s - \gamma)$, where γ_s is unit weight of the bed materials and γ is the unit weight of the fluid. Thus, the work done by the water on the sediment particles to move them away from the scour hole can be expressed as Eq. (9).

$$W_{out} = \left(\frac{\pi p}{2 + p} d_s^{\frac{2+p}{p}} + \frac{\pi q}{2 + q} d_s^{\frac{2+q}{q}} \right) (1 - \theta)(\gamma_s - \gamma) \tag{9}$$

Now, applying the energy balance concept ($W_{in} = W_{out}$) depth of scour can be expressed by Eq. (10) [considering Eqs. (3) and (9)].

$$\begin{aligned} & (\rho u_c^2 h b) \left(\frac{2 + 2m}{2 + 3m} h + \frac{d_s}{2} \right) \\ & = \left(\frac{\pi p}{2 + p} d_s^{\frac{2+p}{p}} + \frac{\pi q}{2 + q} d_s^{\frac{2+q}{q}} \right) (1 - \theta)(\gamma_s - \gamma) \end{aligned} \tag{10}$$

using $A = \left(\frac{\pi p}{2+p} \right) (1 - \theta)(\gamma_s - \gamma)$; $C = \frac{\rho u_c^2 h b}{2}$; $B = \left(\frac{\pi q}{2+q} \right) (1 - \theta)(\gamma_s - \gamma)$; and $E = (\rho u_c^2 h b) \left(\frac{2+2m}{2+3m} h \right)$. Equation (10) can be expressed as Eq. (11).

$$A d_s^{\frac{2+p}{p}} + B d_s^{\frac{2+q}{q}} - C d_s - E = 0 \tag{11}$$

Equation (11) is a nonlinear equation in terms of the scour depth d_s , where the corresponding coefficients A , B , C , and E can be obtained from the given field conditions. The above equation can be solved numerically with the implementation

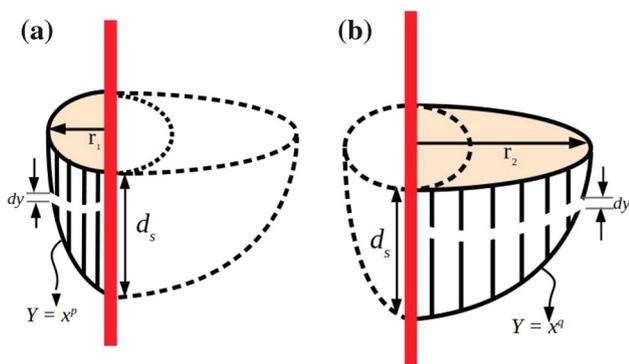


Fig. 4 Large view of the scour shape (Fig. 3b) around the pier: **a** scour at the stoss side of the pier, **b** scour at the lee side of the pier

of a simple FORTRAN or MATLAB code. An example calculation has been done below showing the implementation of presented formulation against methods of Melville and Coleman (2000) and HEC-18 (Richardson and Davis 2001). However, the solution becomes complex as the value of p and q increases.

Illustrative example

The problem is solved for the given data:

- Pier width/diameter, $b = 1.5$ m
- Water depth, $h = 5$ m
- Water discharge intensity, $q = 10$ m² s⁻¹
- Sediment size, $d_{50} = 0.8$ mm,
- Flow skewness, $\alpha = 15^0$
- Unit weight of the bed materials, $\gamma_s = 19,000$ N m⁻³
- Unit weight of the fluid, $\gamma = 9810$ N m⁻³
- Density of the fluid, $\rho = 1000$ kg m⁻³
- Porosity of sand, $\theta = 0.4$

The equilibrium scour depth calculated using the equations presented by Melville and Coleman (2000), HEC-18 (Richardson and Davis 2001), and the present model is given in Table 1. The procedure behind the calculations is given in “Appendix.” For the given data, the scour depths calculated by methods of Melville and Coleman (2000) and HEC-18 (Richardson and Davis 2001) are 5.65 m and 2.2 m, respectively, whereas by the present equation it is 6.88 m for a linear scour hole at the stoss and the lee sides, 3.72 m for a linear scour hole at the stoss side and a parabolic scour hole at the lee side, and 2.93 m when the scour hole is parabolic at both the stoss and the lee sides of the pier. The shape of the scour hole can be varied by choosing any real value of p and q in Eq. 11 formulated in this paper. The differences in the values of the scour depth, given in Table 1, are due to the fact that the assumptions and the factors considered while developing these individual equations are different.

Table 1 Scour depth calculated from the equation given by various authors and the present equation

Author	d_s (m)
Melville and Coleman (2000)	5.65
HEC-18 (Richardson and Davis 2001)	2.2
<i>Present equation</i>	
$p = 1, q = 1$	6.88
$p = 1, q = 2$	3.72
$p = 2, q = 1$	3.72
$p = 2, q = 2$	2.93

Conclusion

An equation for the equilibrium scour depth around a pier is presented in this paper. The equation is developed applying the energy balance concept at both the stoss and the lee sides of the pier for an arbitrary scour hole pattern. The applicability of the equation is checked by solving an example problem utilizing the present equation. The solution is compared with the equations given by Melville and Coleman (2000) and HEC-18 (Richardson and Davis 2001). A good agreement is obtained. As future work, the presented equation can be extended by incorporating the bed slope and the threshold of sediment motion.

Acknowledgements This work was carried out as part of the Institute Scheme for Innovative Research and Development (ISIRD) titled “3D CFD Modeling of the Hydrodynamics and Local Scour Around Offshore Structures Under Combined Action of Current and Waves” from IIT Kharagpur.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest in the current paper.

Appendix: Calculation of scour depth

The detailed calculation procedure of the scour depth given by Melville and Coleman (2000) and HEC-18 (Richardson and Davis 2001) can be found in Dey (2014).

Melville and Coleman

Velocity of incoming flow, $u = \frac{q}{h} = \frac{10}{5} = 2$ ms⁻¹

Threshold shear velocity, u_{*c} , and incoming flow, u_{cr} , are estimated as follows:

$$u_{*c} \quad (0.1 \leq d_{50} < 1 \text{ mm}) = 0.0115 + 0.0125 d_{50}^{1.4} = 0.02 \text{ ms}^{-1}$$

$$u_{cr} = u_{*c} 5.75 \log \left(5.53 \frac{h}{d_{50}} \right) = 0.52 \text{ ms}^{-1}$$

For uniform sediment, $u_a = u_{cr}$.

k -factor

$$\text{For } \frac{b}{h} = \frac{1.5}{5} = 0.3 < 0.7, k_h = 2.4b = 3.6 \text{ m}$$

$$\text{For } \frac{h - u_a - \bar{u}_{cr}}{u_{cr}} = 3.84 > 1, k_l = 1$$

$$\text{For } \frac{b}{d_{50}} = 1875 > 1, k_d = 1$$

For a circular pier, $k_s = 1$

$$\text{For the projected width, } b_p = L \sin \alpha + b \cos \alpha = 3 \text{ m}$$

$$k_\alpha = \left(\frac{b_p}{b} \right)^{0.65} = 1.569$$

For equilibrium scour depth, $(t = t_e)$, $k_t = 1$

$$\text{Now, scour depth } d_s = k_h k_l k_d k_s k_\alpha k_t = 5.65 \text{ m}$$

HEC-18

For a circular pier, $k_s = 1$

For $\frac{L}{b} = 4$ and $\alpha = 15^\circ$, $k_\alpha = 1.5$

Assuming small dunal bed form and Froude number,

$$F_r = \frac{2}{\sqrt{9.8 \times 5}} = 0.28, k_{\text{bed}} = 1.1$$

For $d_{50} < 2 \text{ mm}$, $k_a = 1$

Now, scour depth

$$d_s = bk_s k_\alpha k_{\text{bed}} k_a \left(\frac{h}{b}\right)^{0.35} F_r^{0.43} = 2.2 \text{ m}$$

Present equation

Considering one-sixth power law, ($m = 6$)

For $\mathbf{p} = 1, \mathbf{q} = 1$

$$A = \left(\frac{\pi p}{2 + p}\right)(1 - \theta)(\gamma_s - \gamma) = 5771.32$$

$$B = \left(\frac{\pi q}{2 + q}\right)(1 - \theta)(\gamma_s - \gamma) = 5771.32$$

$$C = \frac{\rho u_c^2 h b}{2} = 15,000$$

$$E = (\rho u_c^2 h b) \left(\frac{2 + 2m}{2 + 3m} h\right) = 105,000$$

$$5771.32d_s^{1.5} + 5771.32d_s^{1.5} - 15,000d_s - 105,000 = 0,$$

$$d_s = 6.88 \text{ m}$$

Similarly, for $\mathbf{p} = 1, \mathbf{q} = 2$

$$A = 5771.32, B = 8656.98, C = 15,000, \text{ and } E = 105,000;$$

$$d_s = 3.72 \text{ m}$$

For $\mathbf{p} = 2, \mathbf{q} = 1$

$$A = 8656.98, B = 5771.32, C = 15,000, \text{ and } E = 105,000;$$

$$d_s = 3.72 \text{ m}$$

For $\mathbf{p} = 2, \mathbf{q} = 2$

$$A = 8656.98, B = 8656.98, C = 15,000, \text{ and } E = 105,000;$$

$$d_s = 2.93 \text{ m}$$

References

- AASHTO, LRFD (2010) Bridge design specifications
- Afzal MS, Bihs H, Kamath A, Arntsen ØA (2015) Three-dimensional numerical modeling of pier scour under current and waves using level-set method. *J Offshore Mech Arct Eng* 137(3):032001. <https://doi.org/10.1115/1.4029999>
- Ahmad N, Bihs H, Chella MA, Arntsen ØA, Aggarwal A et al (2017) Numerical modelling of arctic coastal erosion due to breaking waves impact using REEF3D. In: The 27th international ocean and polar engineering conference. International Society of Offshore and Polar Engineers
- Arneson L, Zevenbergen L, Lagasse P, Clopper P (2012) Evaluating scour at bridges. HEC-18. Federal Highway Administration (FHWA)
- Bouratsis P, Diplas P, Dancey CL, Apsilidis N (2017) Quantitative spatio-temporal characterization of scour at the base of a cylinder. *Water* 9(3):227. <https://doi.org/10.3390/w9030227>
- Chabert J, Engeldinger P (1956) Study of scour around bridge piers. Technical report, prepared for the Laboratoire National d'Hydraulique
- Chang H (1988) Fluvial processes in river engineering. Wiley, New York
- Dey S (1995) Three-dimensional vortex flow field around a circular cylinder in a quasi-equilibrium scour hole. *Sadhana* 20(6):871–885. <https://doi.org/10.1007/BF02745871>
- Dey S (1996) Sediment pick-up for evolving scour near circular cylinders. *Appl Math Model* 20(7):534–539. [https://doi.org/10.1016/0307-904X\(95\)00172-G](https://doi.org/10.1016/0307-904X(95)00172-G)
- Dey S (1997) Local scour at piers, part I: a review of developments of research. *Int J Sediment Res* 12(2):23–46
- Dey S (2014) Scour. In: Fluvial hydrodynamics, Springer, Berlin, pp. 563–639. https://doi.org/10.1007/978-3-642-19062-9_10
- Dey S, Bose S, Sastry G (1992a) Clear water scour at circular piers, part I: flow model. In: Proceedings of 8th conference on IAHR Asian and Pacific Division, pp 69–80
- Dey S, Bose S, Sastry G (1992b) Clear water scour at circular piers, part II: design formula. In: Proceedings of 8th conference on IAHR Asian and Pacific Division, pp 81–92
- Dey S, Bose SK (1994) Bed shear in equilibrium scour around a circular cylinder embedded in a loose bed. *Appl Math Model* 18(5):265–273. [https://doi.org/10.1016/0307-904X\(94\)90334-4](https://doi.org/10.1016/0307-904X(94)90334-4)
- Dey S, Bose SK, Sastry GL (1995) Clear water scour at circular piers: a model. *J Hydraul Eng* 121(12):869–876. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1995\)121:12\(869\)](https://doi.org/10.1061/(ASCE)0733-9429(1995)121:12(869))
- Dey S, Raikar RV (2007) Characteristics of horseshoe vortex in developing scour holes at piers. *J Hydraul Eng* 133(4):399–413. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2007\)133:4\(399\)](https://doi.org/10.1061/(ASCE)0733-9429(2007)133:4(399))
- Ettema R, Melville BW, Constantinescu G (2011) Evaluation of bridge scour research: Pier scour processes and predictions. Citeseer
- Gioia G, Bombardelli FA (2005) Localized turbulent flows on scouring granular beds. *Phys Rev Lett* 95(1):014501
- Gioia G, Chakraborty P (2006) Turbulent friction in rough pipes and the energy spectrum of the phenomenological theory. *Phys Rev Lett* 96(4):044502
- Graf W, Istiarto I (2002) Flow pattern in the scour hole around a cylinder. *J Hydraul Res* 40(1):13–20. <https://doi.org/10.1080/00221680209499869>
- Hafez YI (2016) Mathematical modeling of local scour at slender and wide bridge piers. *J Fluids*. <https://doi.org/10.1155/2016/4835253>
- Harik I, Shaaban A, Gesund H, Valli G, Wang S (1990) United states bridge failures, 1951–1988. *J Perform Constr Facil* 4(4):272–277. [https://doi.org/10.1061/\(ASCE\)0887-3828\(1990\)4:4\(272\)](https://doi.org/10.1061/(ASCE)0887-3828(1990)4:4(272))
- Kamojjala S, Gattu N, Parola A, Hagerty D (1994) Analysis of 1993 upper Mississippi flood highway infrastructure damage. In: Water resources engineering, ASCE, pp 1061–1065
- Kattell J, Eriksson M (1998) Bridge scour evaluation: screening, analysis, and countermeasures. Technical report, USDA Forest Service, San Dimas Technology and Development Center
- Khaple S, Hanmaiahgari PR, Gaudio R, Dey S (2017) Interference of an upstream pier on local scour at downstream piers. *Acta Geophys* 65(1):29–46. <https://doi.org/10.1007/s11600-017-0004-2>
- Kothiyari UC, Hager WH, Oliveto G (2007) Generalized approach for clear-water scour at bridge foundation elements. *J Hydraul Eng* 133(11):1229–1240. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2007\)133:11\(1229\)](https://doi.org/10.1061/(ASCE)0733-9429(2007)133:11(1229))
- Lagasse PF (2007) Countermeasures to protect bridge piers from scour, vol 593, Transportation Research Board

- Lança RM, Fael CS, Maia RJ, Pêgo JP, Cardoso AH (2013) Clear-water scour at comparatively large cylindrical piers. *J Hydraul Eng* 139(11):1117–1125. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0000788](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000788)
- Manes C, Brocchini M (2015) Local scour around structures and the phenomenology of turbulence. *J Fluid Mech* 779:309–324. <https://doi.org/10.1017/jfm.2015.389>
- Melville BW, Coleman SE (2000) Bridge scour. Water Resources Publication, Colorado
- Melville BW, Raudkivi AJ (1977) Flow characteristics in local scour at bridge piers. *J Hydraul Res* 15(4):373–380. <https://doi.org/10.1080/00221687709499641>
- Mutlu Sumer B (2007) Mathematical modelling of scour: a review. *J Hydraul Res* 45(6):723–735. <https://doi.org/10.1080/00221686.2007.9521811>
- Nurtjahyo P, Chen H, Briaud J, Li Y, Wang J (2002) Bed shear stress around rectangular pier: numerical approach. In: First international conference on scour of foundations international society of soil mech and foundations
- Olsen NR, Melaaen MC (1993) Three-dimensional calculation of scour around cylinders. *J Hydraul Res* 119(9):1048–1054. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1993\)119:9\(1048\)](https://doi.org/10.1061/(ASCE)0733-9429(1993)119:9(1048))
- Raikaar R, Dey S (2005a) Scour of gravel beds at bridge piers and abutments, vol 158, Thomas Telford Ltd, pp 157–162
- Raikaar RV, Dey S (2005b) Clear-water scour at bridge piers in fine and medium gravel beds. *Can J Civ Eng* 32(4):775–781. <https://doi.org/10.1139/105-022>
- Raikaar RV, Dey S (2008) Kinematics of horseshoe vortex development in an evolving scour hole at a square cylinder. *J Hydraul Res* 46(2):247–264. <https://doi.org/10.1080/00221686.2008.9521859>
- Richardson E, Davis S (2001) Evaluating scour at bridges, HEC-18. Technical report, Rep. No. FHWA NHI
- Roulund A, Sumer BM, Fredsøe J, Michelsen J (2005) Numerical and experimental investigation of flow and scour around a circular pile. *J Fluid Mech* 534:351–401. <https://doi.org/10.1017/S0022112005004507>
- Salaheldin TM, Imran J, Chaudhry MH (2004) Numerical modeling of three-dimensional flow field around circular piers. *J Hydraul Eng* 130(2):91–100. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2004\)130:2\(91\)](https://doi.org/10.1061/(ASCE)0733-9429(2004)130:2(91))
- Shatanawi KM, Aziz NM, Khan AA (2008) Frequency of discharge causing abutment scour in South Carolina. *J Hydraul Eng* 134(10):1507–1512. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2008\)134:10\(1507\)](https://doi.org/10.1061/(ASCE)0733-9429(2008)134:10(1507))
- Sheppard DM, Odeh M, Glasser T (2004) Large scale clear-water local pier scour experiments. *J Hydraul Eng* 130(10):957–963. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2004\)130:10\(957\)](https://doi.org/10.1061/(ASCE)0733-9429(2004)130:10(957))
- Wardhana K, Hadipriono FC (2003) Analysis of recent bridge failures in the united states. *J Perform Constr Facil* 17(3):144–150. [https://doi.org/10.1061/\(ASCE\)0887-3828\(2003\)17:3\(144\)](https://doi.org/10.1061/(ASCE)0887-3828(2003)17:3(144))
- Yang Y, Qi M, Li J, Ma X (2018) Evolution of hydrodynamic characteristics with scour hole developing around a pile group. *Water* 10(11):1632



Understanding the hydrometeorological characteristics and relationships in the semiarid region of Maharashtra (western India): implications for water management

Rahul S. Todmal¹

Received: 23 July 2019 / Accepted: 1 November 2019 / Published online: 9 November 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

In spite of enormous investment for water harvesting in Maharashtra, the issue of water deficiency persist. Therefore, with the proper understanding of recent (1975–2014) hydrometeorological characteristics in the semiarid region of Maharashtra, the study endeavored to suggest management strategies to optimize the water resources. To understand the inherent characteristics of rainfall, runoff, dam storages and groundwater, statistical techniques including descriptive statistics, correlation, regression and Student's *t* test were applied. The monsoon rainfall has notable control over the water resources. As the study region characterized by significantly low water availability with high variability which is accountable for higher frequency and intensity of drought, it should be prioritized while formulating a judicial plan for water management. The major rivers of Maharashtra follow the unimodal pattern of rainfall. Whereas the study area displays a peculiar character of the bimodal and unimodal regimes of rainfall and runoff, respectively, this highlights the role of hydrological losses. However, it also indicates that there is a wide scope to harvest surface water, particularly during the months of September and October. Interestingly, the inter-annual variability in the extreme rainfall and runoff events over the study basins is observed to be higher in India. Most of the water (> 50%) received during flashy rainfall events drains out through flashy discharges. The impoundment of this water and modified irrigation schedule by considering the recent hydrological characteristics may diminish the variability and deficiency of surface and subsurface water. Furthermore, for the precise forecast of water availability in the study area, ENSO condition needs to incorporate, as it has a significant connection with rainfall and runoff.

Keywords Maharashtra · Semiarid region · Rainfall · Runoff · Groundwater · ENSO · Water management

Introduction

Scientific understanding of hydrometeorological characteristics of a region plays a crucial role in the management of water resources, particularly for the semiarid regions. The regional hydrological tendencies chiefly determine an allocation of water resources to the agricultural, hydropower generation, industrial and domestic sectors (Wang et al. 2003). In the areas where the agricultural practices are rain dependent, the availability of water resources dictates the cropping pattern (Sethi et al. 2006; Khot and Sabanna 2018). Therefore, the water resource planners emphasize the

inherent character of meteorological and hydrological variables to cope with disasters such as flood (Sanyal and Lu 2004), drought and famine.

The hydrology of Maharashtra is primarily depending on the monsoon rainfall received during the monsoon season (June to October). As about 83% of the land area of the state fall under the semiarid climate, water scarcity is a major concern (Kale et al. 2014), especially from the agricultural and domestic viewpoint. Although after the severe drought of 1972, the Drought-Prone Area Programme (DPAP) and Watershed Development Programme (WDP) were started with the objective of solving the agricultural problems of rainfed areas in Maharashtra (Kalamkar 2011), it has limited success (Kale et al. 2014) in the semiarid region, which has been reflected from the experience of severe droughts during the last few decades. The recent droughts in Maharashtra have occurred mainly due to mismanagement of available water resources (Purandare 2013) which is one of

✉ Rahul S. Todmal
todmalrahul@gmail.com

¹ Department of Geography, Vidya Pratishthan's ASC College
Baramati, Pune 413 133, India

the reasons for the agrarian distress and suicides of ~ 15,000 farmer since 1994 (Mishra 2006; Udmale et al. 2014; Suresh 2015). Apart from this, the World Bank (2008) and TERI (2012) have reported that the drought event in 2003 has collapsed the economic budget for irrigation, agriculture and rural development of Maharashtra for the period between 2002 and 2007. Negligence toward the appraisal of water resources in the scarcity zone of Maharashtra is noticeable from the limited scientific studies on the groundwater resource which irrigate > 50% of agricultural land (Pathak et al. 1999; GoM 2014). Israel (with < 500 mm rainfall) and two outstanding examples (Hiware Bazar and Ralegan Siddhi) from the drought-prone area of Maharashtra (Mehta and Satpathy 2011) have proved that the water crises can be overcome with the application of recent water management techniques by considering the recent regional hydro-climatic characteristics.

Generally, the water resources and irrigation management is based on the assumption of stationarity of hydrological variables (Milly et al. 2008). However, the climatic changes play a crucial role in violating this assumption (Pahl-Wostl, 2007; Milly et al. 2008). Under the recent warming scenario (since 1970), climatic changes over the semiarid region of Maharashtra are evident (Todmal and Kale 2016), which may have altered the regional hydrology. The seasonality (Guhathakurta and Saji 2013) and variability (Todmal and Kale 2016) of rainfall in this region are observed to be increasing, which may exacerbate the future water management challenge. Additionally, the groundwater in the Upper Bhima Basin cannot buffer the water scarcity during the rainfall drought of two or more consecutive years, as it has been overexploited during the last few years (Surinaidu et al. 2013; Pavelic et al. 2012).

Under such circumstances, it is time to shed the current approaches of water resources management in the water-scarce region of Maharashtra (Todmal and Kale 2016) and formulate a judicial plan. For this, it is worth necessary understanding the inherent and recent (post-1980) characteristics of the precipitation, surface and subsurface water resources. In the present investigation, therefore, an attempt has been made to decipher the spatiotemporal, seasonal and intra-seasonal characteristics of rainfall, runoff, dam storage and groundwater levels in the five semiarid river basins. The study also endeavors to understand the interrelationships between water resources as well as their dependence on ENSO (El Niño–Southern Oscillation). Additionally, the potential implications of findings emerged from the present investigation for the water resources management are discussed.

The semiarid region of Maharashtra

In the present study, five river basins, namely the Sina, Man, Karha, Yerala and Agrani, which drain the rain shadow zone of Maharashtra, are considered. All the five basins are situated between 16° 45" and 19° 25" N latitude and 73° 45" and 76° 25" E longitude (Fig. 1). The Sina, Karha and Man Basins form a part of the Bhima River which is the largest tributary of the Krishna River in Maharashtra. The Agrani and Yerala Rivers directly drain into the Krishna River. The selected basins cover about 32% area of the scarcity zone (agro-climatic zone) of Maharashtra. Among them, the Sina and Karha are the largest and smallest basins, respectively (Fig. 1).

The study basins are characterized by the semiarid tropical (Bsh) type of climate (IMD 2005). Agro-climatologically these basins are the part of scarcity zone (Deosthali 2002; Kalamkar 2011), where the annual water deficiency is between 900 and 1100 mm (Dikshit 1983; Todmal 2019). The mean daily temperature is generally above 22 °C, except during winter (18–22 °C). The mean maximum temperature varies between 30 and 40 °C from May to October but is generally below 32 °C in the remaining months of the year. The annual maximum temperature (> 40 °C) is normally recorded in the month of May (IMD 2005). The hydrology of the study area completely depends upon the monsoon rainfall (received from June to October). The study basins are underlain by erosion-resistant rock, which limits the base flow contribution to the river discharge. The lower- and middle-order streams completely become dry for about 8 months, and the flows in the higher-order streams become discontinuous (Kale et al. 2014). Therefore, for irrigation and drinking purposes, a large number of water storage structures have been constructed in the study basins during the last three decades (GoM 1999).

Methodology

In the present investigation, four types of hydrometeorological data including monsoon rainfall, monsoon runoff, dam storage and groundwater levels (pre- and post-monsoon) are used. Details of the data collection and locations of the selected stations are given in Table 1 and Fig. 1, respectively. The daily rainfall data of 40 well-distributed rain gauge stations were collected from the India Meteorological Department (IMD) and the Hydrological Data Users Group (HDUG) for about the last three decades. As the semiarid region of Maharashtra receives about 85–90% of annual precipitation during the monsoon

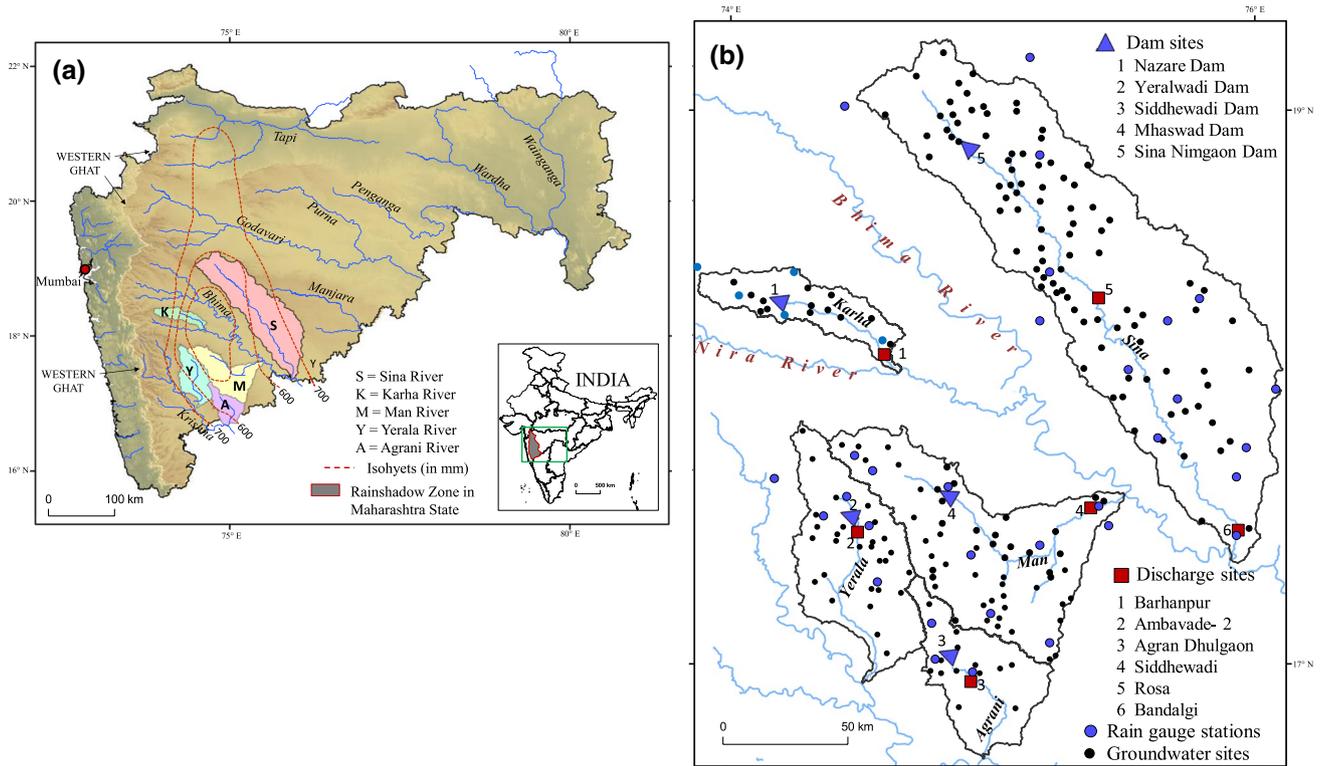


Fig. 1 a Location map of study basins and b distribution of selected hydrometeorological stations in the study area. Base map of Maharashtra State after Kale et al. (2014)

Table 1 Details of data used in the present study

Variables	Selected stations	Length of records	Source	Missing records filled
Monsoon rainfall	40	1981–2013 (33 years)	IMD, HDUG	8%
Monsoon discharge	6	1981–2010 (30 years)	HDUG	0%
Dam storage and spillway discharge	5	1981–2014 (34 years)	IDMS	0%
Post- and pre-monsoon groundwater levels	205	1975–2014 (40 years)	GSDA	7%
Satellite images	Landsat TM-5 for 1998 (wet year) in the month of November and December (post-monsoon season)			

months (Gadgil 2002), the present study has considered yearwise daily rainfall data for the monsoon period. The daily discharge data of six gauge discharge sites were obtained from HDUG for the period between 1981 and 2010. In each of the selected river basin, except the Sina, only one discharge gauging site was available. As the semi-arid rivers flow during the monsoon period (1 June to 31 October), the discharge data during this period are considered. The discharge gives the volume of water passing through a point in a given unit of time at a cross section (Muthreja 1986). The daily runoff values are calculated as

a product of daily discharge and the number of seconds in a day. In order to understand the role of surface water storages in determining the characteristics of monsoon runoff, the daily dam storage and spillway discharge data were acquired from the Irrigation Department of Maharashtra State (IDMS) for the available period (1981–2014). For this, five dams (one from each basin) on the trunk stream were selected (Fig. 1). The groundwater level data of 205 groundwater observations sites (GOS) for which the data were available for both the pre- and post-monsoon seasons were acquired from the Groundwater Survey and

Development Agency (GSDA). Such data were available for variable durations between 1975 and 2014. The missing values in the rainfall and groundwater level data (Table 1) were filled by the linear regression interpolation. The average basin rainfall over each of the selected basin was calculated by using the Thiessen polygon method. As the available discharge gauging sites in the Yerala and Agrani Basin are located at the middle of the basins, in such cases the basin area behind discharge gauging site is treated as upper (U) basin (domain). Similarly, in the case of Sina Basin, the basin areas behind Rosa and Bandalgi (downstream) stations are labeled as Sina (U) and Sina (Dw), respectively (Fig. 1).

To identify the hydrometeorological characteristics, parameters of central tendency and dispersion were obtained from the descriptive statistical analysis. By using these parameters, the coefficient of variation (CV) and normalized departures (Z score values) were calculated for all variables. In order to understand the rainfall and runoff regime in the study basins, the daily data series were derived by averaging the rainfall as well as runoff data of each day of the monsoon period for the gauge period (about three decades). These newly derived data series were used to construct the mean monsoon hydrographs, which represents the average rainfall and runoff characteristics during the monsoon season.

To understand the variability in high-magnitude events over the study basins, the FFMI, which is the standard deviation of annual peak discharges on the log scale, calculated for the selected discharge gauging sites. Based on the same methodology, to evaluate the annual variability in monsoon 1-day maximum rainfall events, the Monsoon Rainfall Magnitude Index (MRMI) developed by Kale (2012) was applied. With the use of linear regression technique, the relationship between FFMI and MRMI was established. To compare the rainfall and runoff over the study basins with other river basins in Maharashtra, Student's t test was applied. With the application of the same technique, stationwise pre- and post-monsoon groundwater levels were compared to understand the role of monsoon rainfall in recharging the groundwater. To explain the spatial variations in groundwater levels, surface water bodies were identified by using freely available satellite images for 1998 (wet monsoon year) (Fig. 1). For this, the Normalized Difference Vegetation Index (NDVI) values were calculated. Generally, the NDVI values between 0 and -0.40 represent surface water bodies. For the assessment of rainfall–runoff–groundwater relationship, the bivariate linear relationships between all the combinations of selected variables (Table 1) were developed. Additionally, the connection between hydrometeorological variables and the Southern Oscillation Index (SOI) were assessed. The annual average SOI values were obtained from the Web site of National Oceanic and Atmospheric Administration Published by New Zealand's Environment

Reporting Series. The data obtained from the same source were used to calculate the average yearly SOI values for the monsoon period.

Results

Rainfall distribution and variability

Figure 2 shows that all the basins under investigation receive monsoon rainfall between 293 and 864 mm, which is very scanty as compared to the rivers heading in the Western Ghats (Krishna, Koyna, Nira, Bhima, etc.). On account of the rain shadow effect of the Western Ghats, noteworthy variations in monsoon rainfall can be observed from east to west direction. Therefore, the western part of the Yerala and Karha Basins receives comparatively higher rainfall amount (> 750 mm). The remaining parts of the basins receive rainfall < 400 mm that results in higher spatial variability (~ 300 – 850 mm). Similarly, toward the eastern side of the Sina Basin, the monsoon rainfall gradually increases (ranging between 500 and 850 mm). The zone of minimum rainfall with maximum variability covers parts of the Karha, Man and Agrani River Basins. In this zone, the average monsoon rainfall and its variability are observed to be < 510 mm and $> 35\%$, respectively (Fig. 2). In the lower part of the Agrani Basin, rainfall is observed between 300 and 450 mm with the highest inter-annual variability (35–55%). More or less, a similar pattern can be observed in most of the Karha and Man Basins.

The box and whisker plot given in Fig. 3a reveals that almost all the basins and their upper parts receive average monsoon rainfall between 450 and 600 mm, with the solitary exception of the Agrani Basin. The lowest amount of monsoon rainfall is recorded in the Agrani Basin, followed by the Man Basin (313 and 453 mm, respectively). On the other hand, the Sina Basin experiences comparatively higher monsoon rainfall total (584 mm), followed by the Yerala and Karha Basins (502 and 490 mm, respectively). As mentioned earlier, the annual variation in the monsoon rainfall is the inherent character of this region. All the study basins reveal highly erratic behavior of annual monsoon rainfall, which varies between 151 and 1007 mm (Fig. 3a). The highest range of monsoon rainfall was observed in the Upper Agrani Basin (200–865 mm), followed by the Sina (353–1007 mm) and Karha (177–808 mm) Basins. This fact is also supported by the high coefficient of variation for the Upper Agrani, Sina and Karha Basins (31, 28 and 33%, respectively).

Notable intra-basin monsoon rainfall variation can be observed in the Agrani Basin, where the entire basin exhibits the lower average monsoon rainfall (313 mm) as compared to the upper domain (490 mm). It can be inferred from this fact that the lower Agrani Basin receives

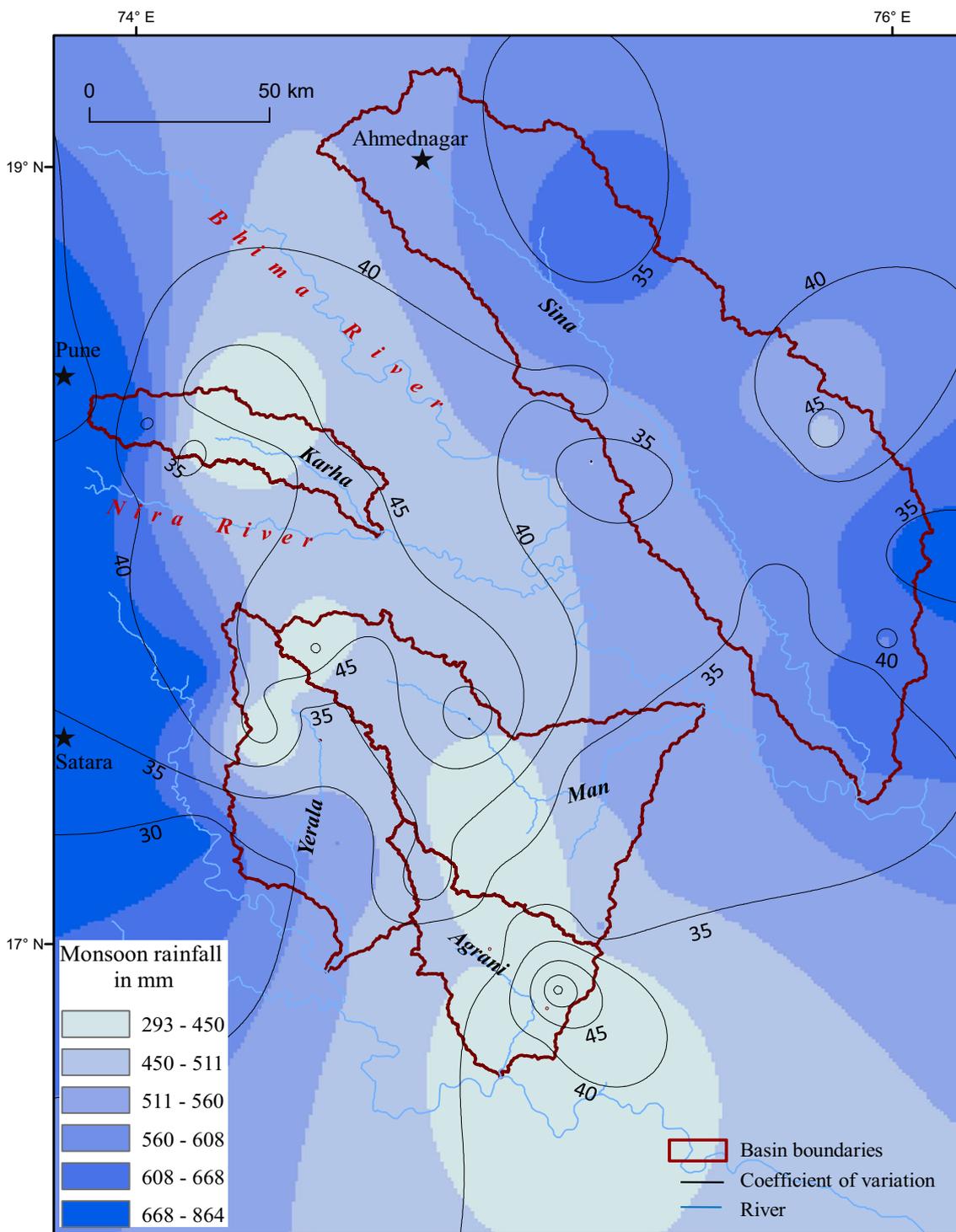
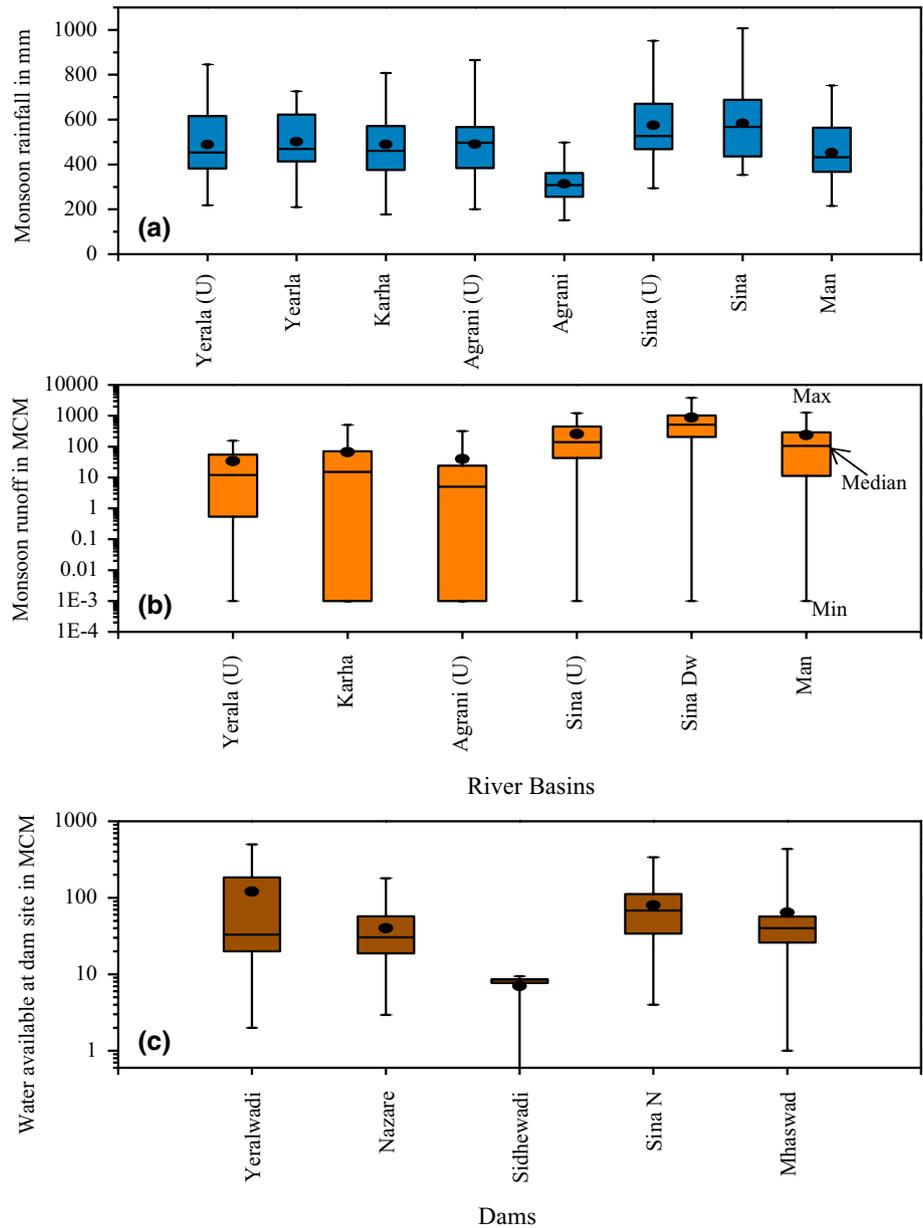


Fig. 2 Spatial distribution of monsoon rainfall and coefficient of variation over the selected study basins. Monsoon rainfall classes are based on natural breaks

monsoon rainfall < 313 mm. The upper domain of the basin depicts high year-to-year variations in the monsoon rainfall (between 200 and 865 mm) as compared to the entire basin (Fig. 3a). Owing to this reason, the Upper

Agrani Basin is characterized by higher rainfall variability (31%) than the entire basin (29%). In comparison, the Sina and Yerala Basins exhibit modest intra-basin variations (Fig. 3a).

Fig. 3 Box–whisker plot showing within- and inter-basin variations in **a** monsoon rainfall, **b** monsoon runoff and **c** water available at dam sites. Black dots denote the mean value, U = upper domain of basin, Max and Min denote maximum and minimum, MCM = million cubic meters. Sina N = Sina Nimgaon Dam, Sian Dw = Sina Basin upstream of the Bandalgi discharge gauging site; spillway discharges are not considered for the Siddhewadi Dam site



Rainfall regime

The southwest monsoon arrives over western Maharashtra by the end of first week of June. During the second and third weeks of June, the monsoon spreads over western and central Maharashtra including the study area. Figure 4 (left panel) depicts the bimodal distribution of monsoon rainfall over the study basins. The first and second peaks in monsoon rainfall are experienced during the months of June and September–October, respectively. It is evident that at the beginning of the monsoon season (second and third dectad of June), each of the selected basins receive about 3–5 mm/day rainfall. Further, the monsoon rainfall pattern shows a remarkable reduction in daily rainfall amounts during July

and August. This is possibly due to the absence of widespread wet spells. In comparison with other basins, the Yerala and Karha Basins display a noteworthy increase in the average rainfall during the month of July. At the end of August, all the basins, except Man, in the study area experience a phase of lowest rainfall during the monsoon period. There is a slight difference in the time of occurrence of this phase, which is characterized by < 2 mm daily average basin rainfall. However, in the Sina and Man Basins, during the second and fourth weeks of August, rainfall events with daily rainfall between 3 and 4 mm are observed. From the second dectad of September, the peak period of monsoon rainfall is experienced in all the study basins (Fig. 4). This monsoon period is characterized by daily average basin rainfall

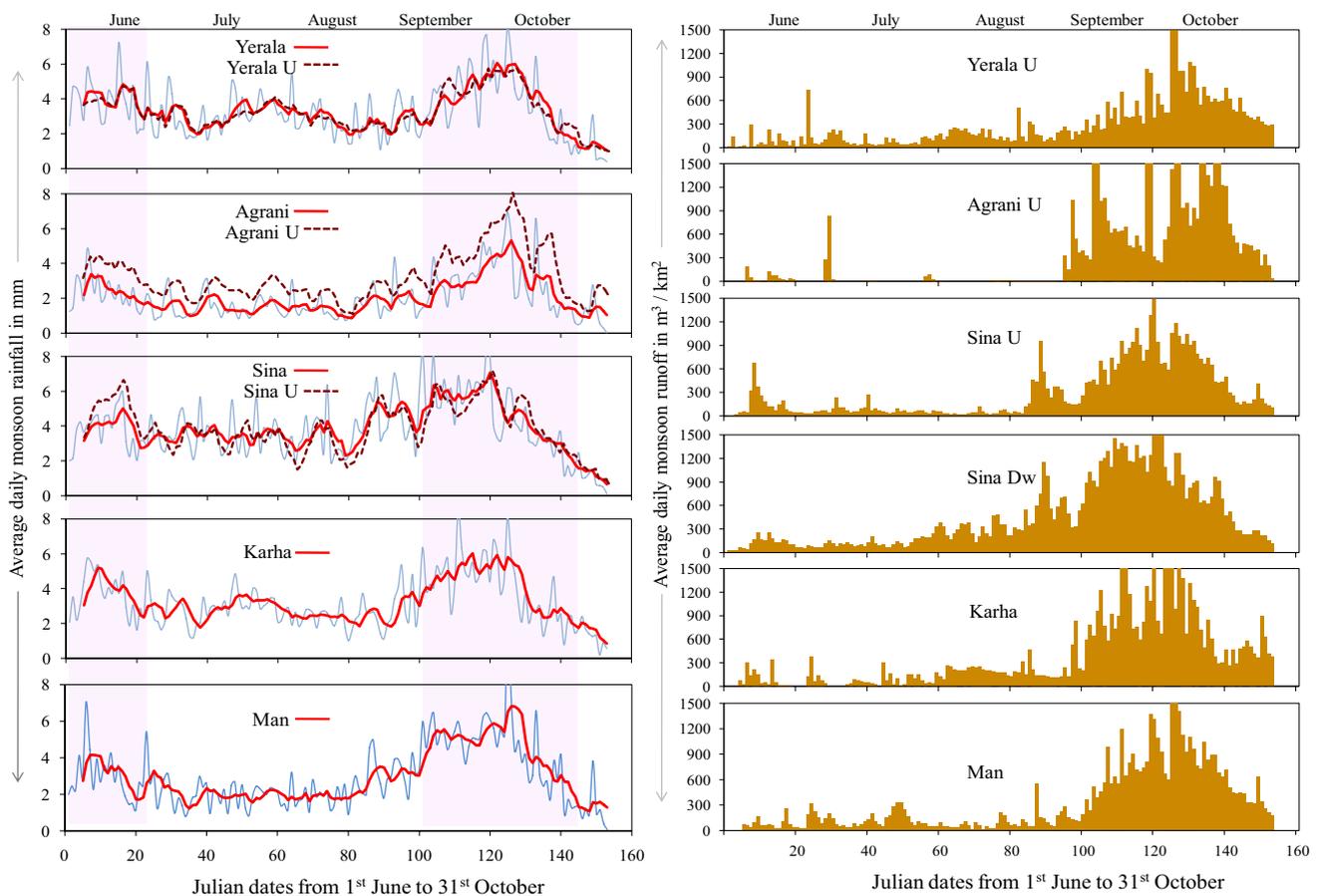


Fig. 4 Basinwise average daily monsoon rainfall (left panel) and runoff (right panel). The daily averages are based on the data between 1981 and 2013. Red and dashed lines represent 5-day moving average monsoon rainfall. Shaded periods indicate synchronized peak mon-

soon rainfall phases. The basin area above the gauge discharge station is considered to calculate unit runoff volume (per km^2). U=upper domain of the basin, Dw=basin area upstream of Bandalgi gauge discharge station

between 4 and 8 mm. This phase terminates by the second dectad of October with the gradual reduction in the rainfall. In general, the drought-prone area of Maharashtra receives about 50% of the total monsoon rainfall during these days. The monsoon ends with a marginal contribution of rainfall during the last two dectad of October.

Surface water distribution and variability

Figure 3 shows that there are remarkable spatial variations in the surface water availability over the study area. The hydrological losses (through evaporation) are, by and large, uniform over the semiarid region of western Maharashtra. Therefore, the basin area and rainfall distribution are the chief determinants for the surface runoff volume. This can be observed in the case of the Sina and Man Basins, which reveal higher average water availability (878 and 238 MCM, respectively) as compared to the Karha, Agrani and Yerala Basins (66, 40 and 34 MCM, respectively) (Fig. 3b). It can also be noticed that the study basins are

characterized by the dominance of below average runoff events (mean > median) during the gauge period. An almost similar picture is observed in the case of dam site water availability. Due to non-availability of spillway records, this fact cannot be ascertained for the Siddhewadi Dam (Fig. 3c). Being smaller in areal extent, the Agrani and Karha Basins exhibit a pronounced effect of the surface storages on the downstream average monsoon runoff. Therefore, these two basins recorded zero runoff events for about > 25% of the gauge period (Fig. 3b). However, such completely dry events are almost absent at the Nazare and Siddhewadi Dam sites (Fig. 3c).

In order to compare the inter-basin variations in the monsoon runoff over the study basins, the average runoff volumes were converted into per unit monsoon runoff volume (thousand cubic meters or TCM)/ km^2 (Table 2). The Sina Basin receives > 550 mm monsoon rainfall, thereby exhibiting the highest average monsoon runoff per unit area (71 TCM/ km^2), followed by the Agrani (U) and Karha Basins (64 and 58 TCM/ km^2 , respectively). The lowest monsoon

Table 2 Basinwise surface water characteristics

Basin	Basin area (km ²)	MRMI (rainfall)	FFMI (runoff)	Average monsoon runoff in TCM/km ²	CV of monsoon runoff	Average available water at dam site in TCM/km ²	CV of water availability at dam site
Yerala U	774	0.21 (2.82)	0.70 (5.6)	44	127	158 (Yeralwadi)	121
Agrani U	622	0.18 (1.74)	0.76 (10)	64	207	40 ^a (Siddhewadi)	47 ^a
Sina U	5654	0.16 (2.10)	0.43 (3)	45	115	50 (Sina N)	88
Sina Dw	12,365	0.15 (1.67)	0.47 (3.1)	71	112	–	–
Karha	1141	0.17 (2.03)	0.57 (5.4)	58	179	101 (Nazare)	87
Man	4626	0.17 (2.22)	0.48 (3.3)	51	180	52 (Mhaswad)	130

U upper domain of the basin, *Dw* basin area behind the downstream gauge discharge site. *CV* coefficient of variation in percent. *TCM* thousand cubic meters and *MCM* million cubic meters. Values in parentheses are the ratio of maximum (rainfall/runoff) during the gauge period to the average maximum (rainfall/runoff). Names of dam are given in parentheses. No dam on the trunk stream at downstream Sina River. *Sina N* Sina Nimgaon

^aDenotes the figures calculated without considering the spillway discharges

runoff volume is observed over the upper Sina and the Yerala Basins (45 and 44 TCM/km², respectively). It is mainly due to the fact that in the Yerala Basin, about 95% of the basin area above the discharge gauging site constitutes the catchment area of the Yeralwadi Dam. Therefore, the actual runoff availability of the Yerala Basin is observed over the Yeralwadi Dam catchment, which is around 158 TCM/km². Although the upper Sina Basin receives a higher amount of rainfall (576 mm), due to a large number of watershed management structures, the majority of runoff gets trapped. In addition to this, in the upper Sina Basin, about 25% of the basin area above the discharge gauging site (Rosa) forms the catchment area of the Sina Nimgaon Dam, which has a storage capacity of 65 MCM.

In the Sina, Karha and Man Basins, it is possible to examine the intra-basin runoff variations by considering the data of water volume at the dam site or another upstream discharge gauging station (Table 2). The larger per unit runoff volume can be noticed over the dam catchments of Yeralwadi and Nazare Dams (158 and 101 TCM/km², respectively) in comparison with the entire river basin. However, marginal runoff over the Sina Nimgaon and Mhaswad Dams is observed, which possibly indicates the impact of watershed management practices adopted within the upstream domain of the dam site.

Similar to the rainfall, the highest monsoon runoff variability (207%) is observed over the Agrani Basin, followed by the Man Basin (180%) (Table 2). It is important to mention here that these basins fall in the high-rainfall-variability (> 35%) zone. The annual runoff in the Sina Basin is observed with the least variability (112%). Being a part

of high-variability and low-rainfall zone, the Mhaswad and Yeralwadi Dam catchment areas show highly variable runoff (130 and 121%, respectively). On the other hand, the Sina Nimgaon and Nazare Dams depict considerably lower annual variations (about 88%) in the surface runoff (Table 2).

Runoff regime

Figure 4 shows that, although the monsoon rainfall displays a bimodal distribution, the monsoon runoff shows a unimodal pattern. The runoff generation begins with the onset of the monsoon season (Fig. 4) (right panel). Although the month of June contributes about 20% of the total monsoon rainfall, marginal surface runoff (< 5%) is generated over the study basins during this period. The similar situation prevails during July and August. The Sina, Karha and Yerala Basins display slightly higher runoff values during the month of August. The months of September and October contribute to the maximum amount of rainfall. As a result of this, the same period is characterized by peak monsoon runoff (between 300 and 1500 m³/km² per day). However, the time of the rise in the runoff (peak runoff) shows slight inter-basin variations.

About > 70% of the monsoon runoff in the study basins is observed in the last 2 months of the monsoon season. Particularly, in the case of the Agrani Basin, about 95% of the runoff occurs mainly during this period (Fig. 4). In other words, this basin remains almost dry during the early monsoon (June to August) period. Therefore, among the study basins, the Agrani Basin exhibits high intra-seasonal

variations in the average daily monsoon runoff. The abrupt rise in the daily runoff (during the first dectad of September) can be noticed over the Karha and Agrani Basins. Another interesting fact that can be observed is that the Yerala and Karha Basins reveal comparatively high runoff formation in the first two dectad of the month of October. As a result of this, in the Yerala Basin, the highest amount (about 55%) of the total runoff is observed during the month of October. An examination of the peaks in the monsoon hydrographs in Fig. 4 (right panel) provides some interesting facts. During the last two dectad of the October month, the retreat of the monsoon season is reflected by the falling limb. The daily average runoff amount during this period sharply decline, particularly in the Karha and Agrani Basins.

Variability of extreme rainfall and runoff

The basinwise MRMI and FFMI values are given in Table 2. The higher year-to-year variations in 1-day maximum rainfall are observed in the Yerala (U) Basin (0.21), followed by the Agrani Basin (0.18). The Karha, Man and the Sina (U) Basins exhibit comparable magnitudes of 1-day maximum rainfall events. However, over the entire Sina Basin, the maximum rainfall events are observed to be the lowest. This is perhaps due to the averaging effect on the 1-day heavy rainfall events for the comparatively larger basin. It is clear from the statistical analysis that the MRMI and the ratio of heaviest 1-day rainfall to the mean maximum rainfall are significantly associated (at 95% confidence level).

Similarly, annual variations in the peak monsoon runoff are reflected from the FFMI. The higher FFMI values over the Agrani, Yerala and Karha Basins (0.76, 0.70 and 0.57, respectively) indicate higher inter-annual variations in peak runoff volume. On the other hand, the larger basins (Sina and Man) depict comparatively lower values of FFMI. Here it is pertinent to mention that the obtained FFMI values for the study basins are higher than the world average (Kale 2002). It is evident that the FFMI values are positively associated with the ratio of 1-day maximum runoff over the gauge period to the mean peak runoff volume. This relationship is statistically significant at 95% confidence level.

Distribution and variability of seasonal groundwater levels

Figure 5a shows that the average post-monsoon GW level depths, which vary between 1 and 11 m from the ground surface, do not reveal any systematic spatial pattern over the study area. The eastern part of the Sina Basin, which receives the comparatively higher amount of rainfall, is characterized by GW closer to the ground surface (< 4 m bgl). It can be noticed that the same domain of the Sina Basin depicts higher annual variability of groundwater

depths during the post-monsoon season (Fig. 5b). Although the Karha and Yerala Basins receive a reasonably adequate amount of rainfall (> 490 and > 502 mm, respectively), they are characterized by higher post-monsoon groundwater level (between 5 and 11 m bgl). Another observation that could be made from Fig. 5b is that a very small area in the Yerala and Agrani Basins is characterized by highly variable GW resource during the post-monsoon season. There are some pockets in the Man, Yerala and Karha Basins that exhibit higher annual variations in the post-monsoon GW levels, which perhaps are linked to the meteorological chronic to severe drought-prone zone (where rainfall is < 510 mm and CV of rainfall varies between 35 and 50%). The pockets of distinctly high and low post-monsoon GW depths (with lower and higher variability, respectively) indicate a notable role of surface water management practices (Fig. 5a, b).

Figure 5c, d shows the distribution of average pre-monsoon GW depths and its annual variability in the study basins, respectively. The spatial variation in the pre-monsoon GW levels almost mimics the post-monsoon pattern of GW levels. Figure 5 shows that in the areas where the post-monsoon GW level is observed to be higher (4–11 m bgl), the same pockets are observed with higher groundwater depths (8–18 m bgl) during the pre-monsoon season. The Sina Basin receives comparatively higher rainfall (> 550 mm). This explains why about 70% area of this basin exhibits GW depth < 9 m even during the pre-monsoon season (Fig. 5c). In spite of receiving an adequate amount of rainfall in the upper Karha and western Yerala Basins, the pre-monsoon GW levels are observed to be far deeper (between 8 and 18 m bgl). Figure 5 shows that the pre-monsoon GW level depths have lower inter-annual variations (CV < 63%) as compared to the variations in post-monsoon GW level. Among all the study basins, the Upper Karha and eastern side of the Sina Basin exhibit the highest inter-annual variation (CV 34–76%) in the pre-monsoon GW level depths. Although a large part of the Man, Yerala and Agrani Basins fall in the low-rainfall zone, these basins reveal marginal annual variability (< 45%) in the pre-monsoon GW tables.

Seasonal fluctuations in the groundwater depths

In the study area, the groundwater recharge is primarily determined by the monsoon rainfall. Therefore, the results obtained from the Student's *t* test suggest about 80% of the selected dug wells reveal a statistically significant decrease in the depth of the water table (that is, a rise in the water table) at the end of the monsoon season.

Figure 6 exhibits the spatial variation in the average rise in water table after the monsoon season (at the end of October), which rises by about nine meters. Figure 6 shows that the pattern of groundwater table rise broadly follows the

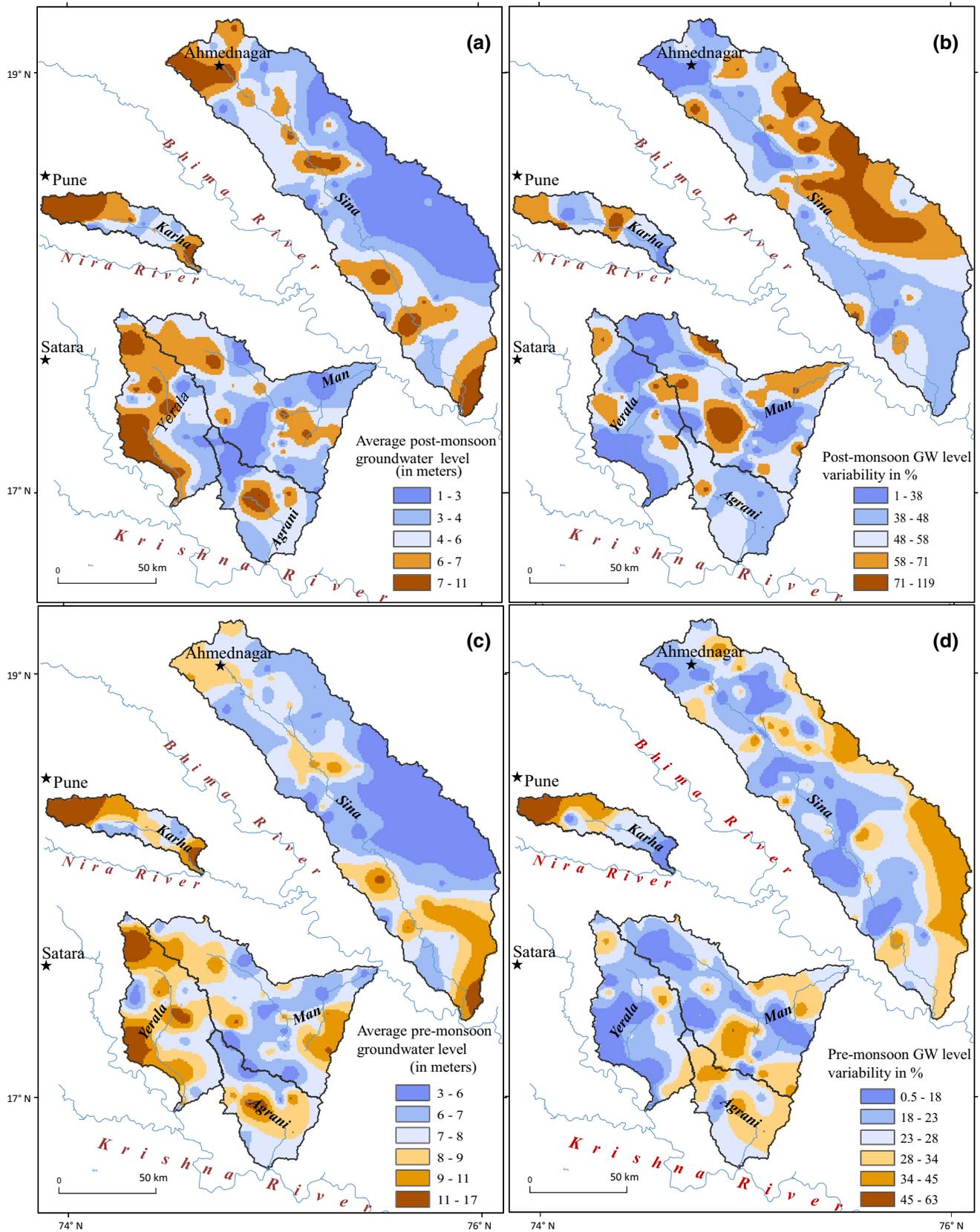
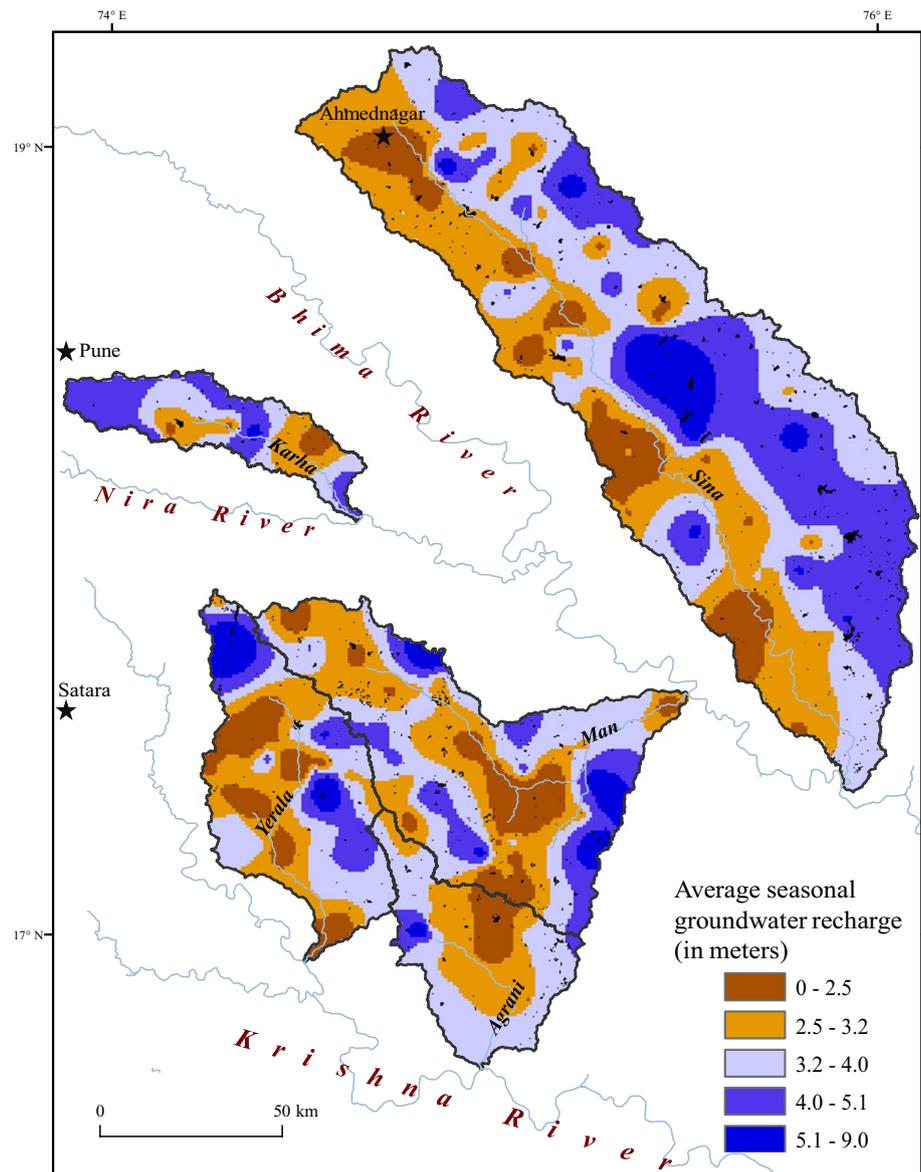


Fig. 5 Spatial distributions of **a** Average GW levels and **b** variability (CV) for the post-monsoon season. **c** Average GW levels and **d** variability for the pre-monsoon season. Classes based on natural breaks. Figures are based on the data for variable durations between 1975 and 2014

Fig. 6 Sitewise average difference in the GW table depths at the end of the monsoon season. Classes based on natural breaks. Results are based on the data of variable duration between 1975 and 2014. The pre-2000 water bodies derived from NDVI analysis of Landsat images are shown in black



spatial pattern of monsoon rainfall over the study basins (see Fig. 2). The middle domains of the Man, Yerala and Agrani Basins exhibit minimum rise in the GW table during the gauge period. These domains fall in meteorologically chronic to severe drought-prone zone. Similarly, almost half of the Sina Basins (western part) exhibits a marginal rise in the water table (by <math>< 4\text{ m}</math>), whereas the eastern part recorded average rise in water table by 4–9 m (Fig. 6). Apart from this, in the Karha, Yerala, Man and Agrani Basins, pockets are observed with the seasonal average rise of groundwater levels by 4–9 meters.

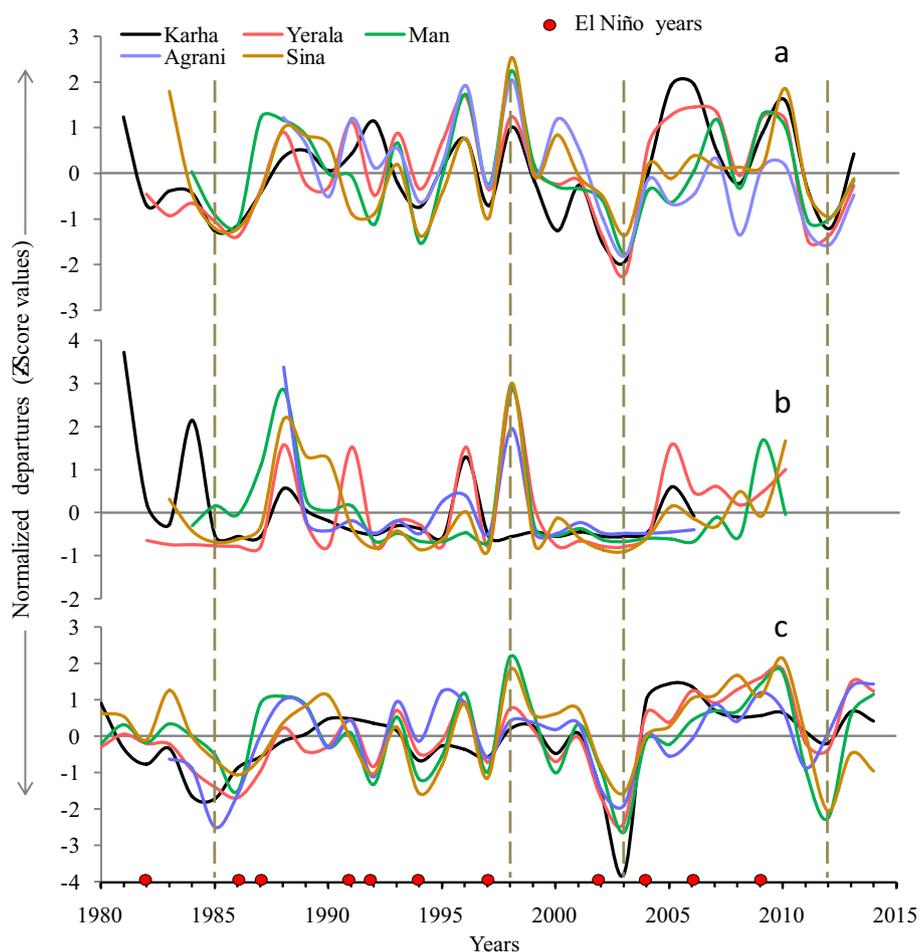
In order to ascertain the role of water impounding structures on the fluctuations in the water table, the water bodies present in the area were identified from the Landsat images (for the year 1998) via NDVI values. The identified water bodies are presented in Fig. 6. From the map, it can be noted

that where the rise in water table after monsoons is modest (<math>< 3.2\text{ m}</math>), there are fewer water bodies.

Temporal variations in rainfall, runoff and groundwater levels

Figure 7 shows that the synchronized temporal variations in the monsoon rainfall and post-monsoon groundwater level were observed over the study basins between 1980 and 2013. This can be observed from 1993 to 1999, from 2002 to 2004 and between 2008 and 2013. On the contrary, the temporal variations in monsoon runoff display consistency only during the extreme monsoon years, particularly in 1985, 1986, 1994, 1998 and 2003. Among these years, 1985–86, 2003 and 2012 were the widespread drought years when the notable decline in all the water resources in

Fig. 7 Basinwise temporal variations in **a** monsoon rainfall, **b** monsoon runoff and **c** post-monsoon groundwater table over the study area. Negative values of PMGW denote fall in the water table and vice versa. Dashed gray lines represent synchronized and region-wide hydrological events. Graphs are based on the basinwise available data between 1980 and 2014



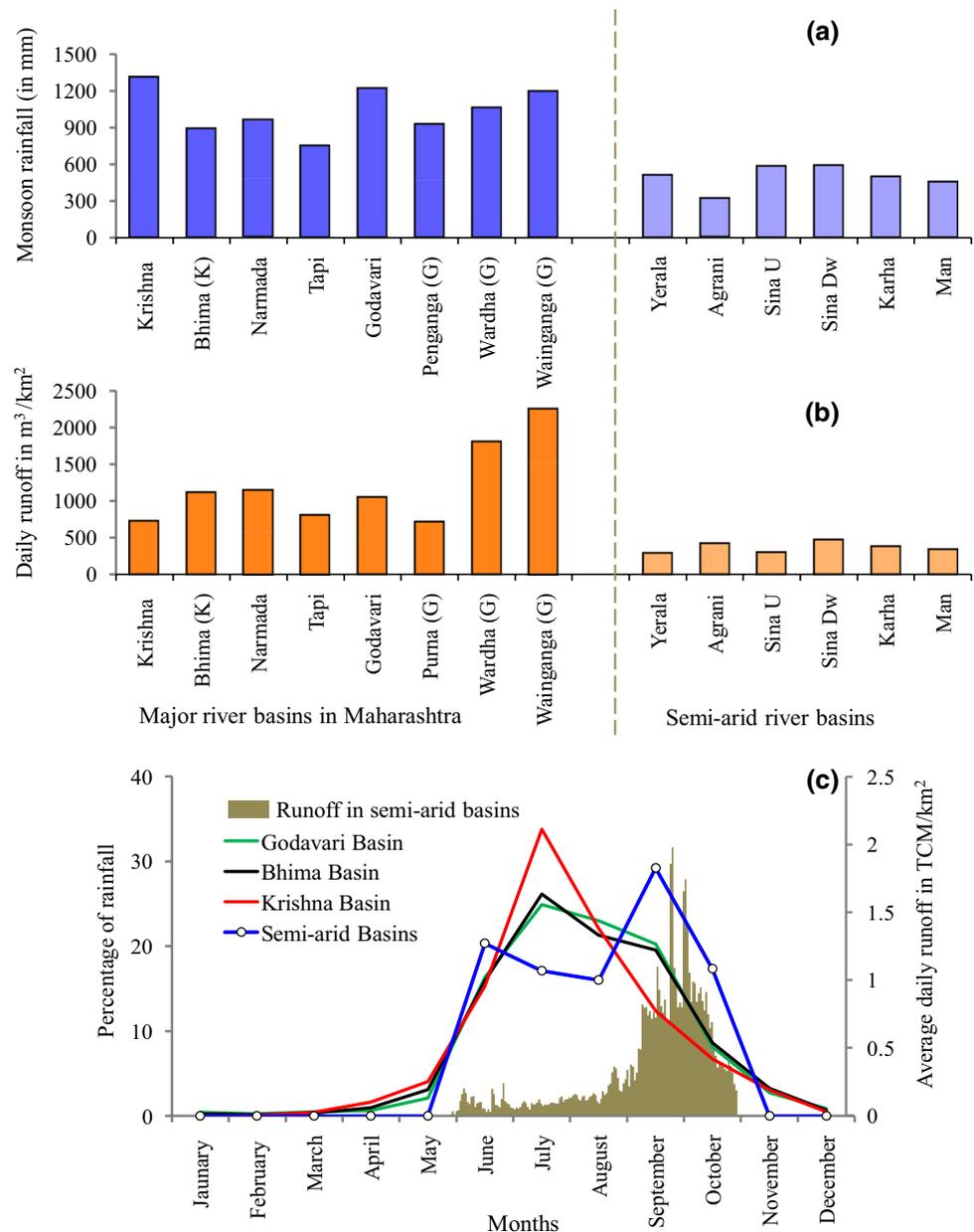
the study basins was experienced (Fig. 7). As 1998 was the wettest monsoon year, the monsoon runoff and groundwater table observed significantly higher. 1992, 1994, 1997, 2002, 2003 and 2012 years are observed with region-wide low monsoon runoff. Some of these are observed with no discharge in the rivers. Among all the study basins, the Karha Basin displays anomalous behavior of rainfall (CV 33%), runoff and groundwater level as well.

In the study area, only a few widespread low-rainfall years are associated with the El Niño events (Fig. 7). There are studies that indicate that the all-India monsoon rainfall is weakly associated with the El Niño events during the recent decades (Kumar et al. 1999; Shewale and Kumar, 2005). However, the study carried out by Todmal and Kale (2016) observed that about 65% of the negative departures in rainfall over the Karha Basin are observed in El Niño years. Similarly, the post-monsoon groundwater level in the study basins has a weak relationship with the El Niño events. In the case of monsoon runoff, about 33% occasion the study basins are dry during the El Niño years.

Discussion

The results given in the previous section highlight inherent nature of monsoon rainfall, runoff and groundwater levels in the semiarid region of Maharashtra. The selected river basins receive the lowest amount of rainfall in the state. The statistically significant result of the Student's *t* test supports this fact (Fig. 8a). Due to limited rainfall, these basins are characterized by higher annual water deficiency (800–1100 mm) (Dikshit 1983). As the Western Ghats is the predominant physical factor that determines the distribution of rainfall, the leeward side of Western Ghats exhibits a notable decline in the monsoon rainfall toward the east (Gadgil 2002). Therefore, the eastern parts of the Yerala and Karha Basins located in the vicinity of Western Ghats receive comparatively higher rainfall. The zone of minimum rainfall (< 510 mm) with maximum variability (CV > 35%) covers the Agrani Basin and parts of the Karha, Yerala and Man Basins. Therefore, among the study

Fig. 8 Comparison of **a** monsoon rainfall and **b** runoff over the semiarid river basins with the major river basins in Maharashtra. (K) and (G) denote tributary of the Krishna and Godavari Rivers, respectively. Areas of other than semiarid basins in figure **a** and **b** are not comparable. **c** Percentage share of monthly rainfall in the major river basins within Maharashtra vis-à-vis the semiarid basins under review. Monthly rainfall data for Bhima, Godavari and Krishna Basins from IMD (2005). Runoff data for major river basins in Maharashtra from Kale et al. (2014) and Subramanian (2004)



basins, the lowest monsoon rainfall (313 mm) with high variability is observed over the Agrani Basin, followed by the Man Basin (453 mm). In the previous studies (Chowdhury and Abhyankar 1984; GoM 1999; Deosthali 2002; Gadgil 2002), the same area was recognized as chronic to severe drought-prone zone where the drought probability is more than 30%. Based on the assured rainfall, in India, there are three pockets (observed in Gujarat–Punjab, Maharashtra and Karnataka States) with the lowest crop potential (Biswas and Nayar 1984). The observed chronic to severe water scarcity zone in the present study is one of them. Furthermore, the monsoon rainfall over the semiarid basins has another characteristic that the average rainfall has a significant inverse relation with CV (Fig. 9a). It

means that the places with lower rainfall are associated with the higher CV and vice versa. Under the scenario of climate change, it has been reported that the water scarcity in such semiarid regions is very likely to aggravate (IPCC 2013; TERI 2014). Therefore, the present study area seeks special attention to water resources management.

As the monsoon runoff mimics the spatial pattern of monsoon rainfall, the highest monsoon runoff variability (207%) is observed over the Agrani Basin, followed by the Man Basin. It is important to mention here that these basins fall in the high-rainfall-variability (>35%) zone. The annual runoff in the Sina Basin is observed with the least variability (Table 2). These differences in CV could be ascribed to the basin rainfall characteristics and the basin area. The

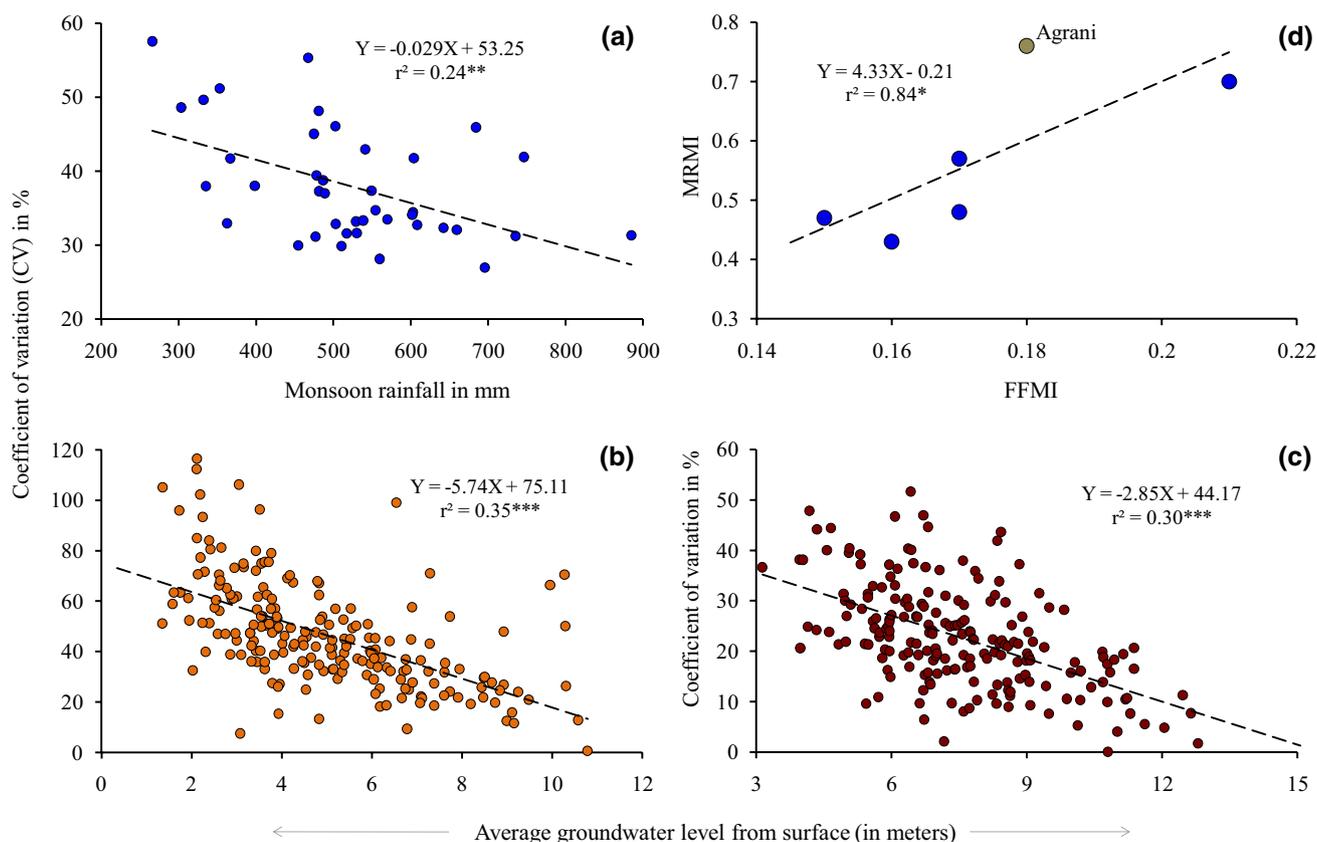


Fig. 9 Linear relationships of **a** monsoon rainfall **b** post-monsoon GW level **c** pre-monsoon GW level with their respective coefficient of variation and **d** relationship between MRMI and FFMI. Relation-

ship between MRMI and FFMI is significant by excluding the Agrani Basin. *, ** and *** represent significant relationship at 95%, 99% and 99.9% confidence levels, respectively

same fact is observed in the case of dams located within (Mhaswad and Yeralwadi) and outside (Sina Nimgaon and Nazare) of the low-rainfall (and high-variability) zone. The Karha and Agrani Basins are the part of chronic to severe drought-prone zone; therefore, these rivers were observed dry for about > 25% of gauge period (Fig. 3). Similarly, among the selected dams, the catchment area of Mhaswad Dam completely falls in the low-rainfall zone. Therefore, the lowest amount of average monsoon runoff (52 TCM/km²) is observed at this dam site. The rainfall distribution has pronounced effect on the surface water availability in Maharashtra. The result obtained from Student's t test suggests that the surface runoff (daily runoff TMC/km²) generated in the semiarid basins is remarkably low as compared to the other river basins in Maharashtra (Fig. 8b). In a nutshell, the ripple effect of rainfall on the surface water hydrology over the semiarid region is noticeable in a pattern of spatial distribution and character as well. The significant relationship between extreme (1-day maximum) rainfall (MRMI) and peak runoff events (FFMI) also supports this fact (Fig. 9d). The MRMI and FFMI for the major Indian rivers range from 0.05 to 0.11 and from 0.05 to 0.3, respectively (Kale

2003, 2012). The higher range of MRMI and FFMI values (0.15–0.21 and 0.4–0.8, respectively) for the study basins indicates the inherent character of higher inter-annual variations in extreme rainfall and flood discharges as compared to the Indian rivers from different climatic zones. It suggests need to adopt the measures which can restrain the flashy character of surface water and increase its availability.

As the monsoon rainfall over the study area is primary source to recharge the groundwater table, rise in GW level during post-monsoon season mimics the spatial pattern of rainfall. Broadly, the areas with deeper post-monsoon groundwater level are experiencing the deepest groundwater table during pre-monsoon season of the next calendar year. The parts of study basins fall in the rainfall deficiency zone display marginal rise (< 4 m) in GW table after monsoon season (Fig. 6). The pockets of distinctively higher groundwater table are very likely associated with surface water management practices. The areas with dense surface storage structures have a considerably higher rise (> 4 m) in water table (Fig. 6). It clears that the intensification of surface harvesting is the most appropriate strategy to improve the groundwater resource. Figure 8b and 8c shows statistically

significant negative relationships between sitewise groundwater depths (pre- and post-monsoon, respectively) and values of coefficient of variation (CV). This implies that places with lower GW level depths are associated with higher annual variability and vice versa. It appears that such places are influenced by either the inherent variability in the monsoon rainfall or the watershed management practices and/or local hydrogeology. Another fact that could be noted here is that while obtaining the GW level records when the dug well is completely dry, the Groundwater Survey and Development Agency (GSDA) has considered the pre-monsoon reading same as the depth of dug well (Gokhale and Sohani 2015). This could be another possible reason behind the lower inter-annual variations (CV) for the higher GW levels, especially during the pre-monsoon season. It is obvious to reflect the character of rainfall variability in the subsequent water resource. However, it is feasible to minimize the intra-seasonal and inter-annual variability in surface and subsurface water resources.

Based on the assured rainfall, the basins under investigation have lower crop potential (Biswas and Nayar 1984). It is, therefore, important to understand the intra-seasonal pattern of monsoon rainfall to plan the rainfed agricultural crops. The monsoon rainfall over the study basins shows a bimodal distribution. The first and second peaks in monsoon rainfall are observed in the month of June and September, respectively. In between two rainfall peaks, there is a lean period (July to August). This is a peculiar characteristic of the semiarid basins, whereas the major river basins in Maharashtra (Godavari, Krishna, Tapi and Narmada) exhibit a unimodal distribution of rainfall (IMD 2005) during the monsoon period (Fig. 8c). This inherent nature of semiarid region needs to be considered separately in the planning of available water resources. Although most of the rainfall characteristics in the study area are reflected in the surface and subsurface water resources, the pattern of runoff regime reveals an exceptional behavior. The monsoon runoff has unimodal distribution. The only peak in runoff is observed in the month of September and October. The early monsoon months (June to August) contribute a marginal amount of runoff (<20%). It is probably due to the marginal rainfall

(<40% of total monsoon rainfall), higher hydrological losses and storage during the early monsoon season. Secondly, to facilitate irrigation during the post-monsoon season, many medium-scale dams (command area between 2000 and 10,000 hectares) have been constructed in the study region, particularly after 1980 (Biggs et al. 2007; Samal and Gedam 2012; GoM 2014). The runoff during early monsoon days is stored in such structures. Therefore, most of the runoff during June to August cannot reach the downstream reaches of the rivers. The effect of surface storages is remarkable in the small river basin such as Agrani Basin, where about 5% of runoff is generated during the early monsoon months. Furthermore, the abrupt rise in daily runoff (during the first decadal of September) over the Karha and Agrani Basins indicates control of surface storages on the monsoon runoff formation during the early monsoon period, particularly in the smaller catchments (Fig. 4b). It is pertinent to mention here that the previous studies (Biggs et al. 2007; Todmal et al. 2018) have observed significantly declining monsoon runoff within the Krishna Basin due to the increase in surface storage structures. Apart from this, the gradual fall in the peak runoff is an indication of base flow contribution (Deodhar 2000), and the sharp decline in the falling limb of the peak runoff phase in the Karha and Agrani Basins suggests that there is very little contribution of base flow to the daily discharges (Fig. 4b). Needless to say, such moderate-sized rivers become dry immediately after the termination of the monsoon season.

The spatial and temporal consistency in rainfall, runoff and post-monsoon GW level can be noticed over the study basins (Fig. 7), particularly during the dry (1985–86, 2003 and 2012) and wet (1998) years. In order to understand the broad picture about interrelationships in hydrometeorological variables over the semiarid region of Maharashtra, the average data series of rainfall (in mm), runoff (in TCM/day/km²) and post-monsoon GW level (in meters) were derived for the entire study area (average for study basins) for the period between 1980 and 2014. Table 3 shows that the monsoon rainfall is the primary source of water in the semiarid hydrology, as it significantly determines the surface and groundwater resources.

Table 3 Correlation (*r*) between hydrometeorological variables over the study area

Variables	Monsoon rainfall	Monsoon runoff	Post-monsoon GW level (bgl)	SOI	SOI (monsoon)
Monsoon rainfall	1	0.71 ^a	−0.47 ^a	0.24	0.53 ^a
Monsoon runoff	0.71 ^a	1	−0.31	0.42 ^a	0.64 ^a
Post-monsoon GW level	−0.47 ^a	−0.31	1	−0.15	−0.18
SOI (annual)	0.24	0.42 ^a	−0.15	1	0.80 ^a
SOI (monsoon months)	0.53 ^a	0.64 ^a	−0.18	0.80 ^a	1

^aDenote statistically significant relationship at 95% confidence level. *bgl* below ground level. Decrease in groundwater level imply rise in water table and vice versa. Negative values in SOI data series used represent weak to strong El Niño conditions

Relationship between ENSO and water resources

The relationship between El Niño events and monsoon rainfall at all-India level has become weaker in recent decades (Kumar et al. 1999; Shewale and Kumar 2005); almost the same fact is observed for the present study area (Table 3). However, the monsoon rainfall has a significant connection with the average SOI for monsoon period. The study conducted by Todmal and Kale (2016) corroborates the same results. It suggests that the El Niño events (or negative SOI during June to October) cause retard of monsoon over the semiarid region of Maharashtra. The insignificant relationship between monsoon runoff and post-monsoon groundwater level indicates that the groundwater resource in selected basins depends on the monsoon rainfall and surface storages rather than river discharges. Therefore, there is a significant correlation between monsoon rainfall and post-monsoon GW level. Moreover, the surface water storages and antecedent moisture availability also affect the GW level in the post-monsoon season. Very likely due to this reason, in spite of strong association with monsoon rainfall, the PMGW level does not show a considerable connection with SOI (El Niño). The hydrology of Indian rivers is linked with the SOI and El Niño/La Niño events (Kale 2005). At the regional level also, the monsoon runoff in the semiarid region has good agreement with SOI, which is statistically significant. As the monsoon runoff directly depends on the monsoon rainfall, it is obvious to observe no or low discharges in the years with negative SOI (El Niño years).

Implications for water resource management

The present study basins in the semiarid region of Maharashtra are characterized by low water availability with higher variability. On account of this characteristic of the study basins, higher frequency and intensity of droughts is observed during the last century (Gore and Sinha Ray 2002). In order to cope with regional water scarcity and reduce the agrarian distress, findings from the present investigation may improve the water management strategies as follows:

- The areas with low rainfall (< 500) and high variability (> 35%) (Agrani Basin and parts of the Karha, Yerala and Man Basins) are the most sensitive in Maharashtra, as the estimated rise in temperature (by 1.02 °C) very likely to aggravate the water deficiency and augment the agricultural water demand (Todmal et al. 2018). Therefore, the state government should give priority to such domains in water and irrigation management, particularly to improve the crop potential.
- In order to minimize the effect of intra-seasonal variability in rainfall and surface water resource, a sufficient number of surface impoundment structures should be

constructed which can trap the rainwater, particularly during the second peak of monsoon. Thus, the collected water during the early monsoon period can be used to irrigate the monsoon crops (*Kharif*) and such reservoirs can be refilled during the late monsoon peak which may facilitate irrigation to the post-monsoon crops (*Rabbi*).

- The knowledge of the bimodal and unimodal distribution of monsoon rainfall and runoff, respectively, emerged from this investigation can be useful for agricultural planning and scheduling the regional crop calendar.
- Every year about 7687 Million Cubic Meters (MCM) (from Fig. 8a) water drains out from the study area, mostly during the flashy rainfall and discharge events. It suggests a wide scope for surface water impoundment which has the potential to irrigate > 100,000 ha of agricultural land.
- The marginal base flow contribution to the river discharges indicates limited recharge of the groundwater resource. As it is clear from the present study that the areas with dense surface water storages show a higher rise in the water table, the extension of surface water harvesting practice will definitely maintain the groundwater table throughout the year and helps to diminish the intra-seasonal and inter-annual variability.
- As the ENSO considerably determines the availability of water resources in the study region, this finding can be implicated in forecasting of water resources and irrigation planning.
- The government should compel and heavily subsidize the water-saving technologies for agriculture and industries in the study area.

Conclusion

The present investigation confirmed that the semiarid region of Maharashtra is characterized by deficient water resources with higher variability. The variability in rainfall resulted due to natural and climatic change has created a great challenge for water resource managers. Although the observed inherent nature of rainfall cannot be modified, it is necessary to alter the agricultural cropping pattern and irrigation schedule to optimize the available rainwater. Although the study highlights the effect on surface and groundwater resources, a substantial volume of water which drains out (~7687 MCM/per year) from the study area exhibits hydrological potential and seeks judicial water management. With the consideration of runoff and groundwater characteristics, the availability of water resources can be enhanced with the extension of surface and subsurface water harvesting which may also diminish the inter-annual variability in available water and obviously improves the regional crop potential. The government and non-government agencies should

priorities such sensitive but potent regions while defining the strategies for water management and implementation of water-saving technologies. Apart from this, for the precise forecast of regional water availability in the semiarid region of Maharashtra it is indispensable to consider the ENSO predictions.

Acknowledgements The author would like to thank all the government agencies for supplying the required data for this study. The author also expresses his sincere gratitude to Dr. Vishwas S. Kale for his valuable guidance to carry out the present work. The author is also thankful to the anonymous reviewers for their comments and suggestions which helped in improving this manuscript.

Compliance with ethical standards

Conflict of interest The author states that there is no conflict of interest.

References

- Biggs TW, Gaur A, Scott CA, Thenkabail P, Parthasaradhi GR, Gumma MK, Acharya S, Turrall H (2007) Closing of the Krishna Basin: irrigation, streamflow depletion and macroscale hydrology, Colombo, Sri Lanka. International Water Management Institute (IWMI) Research Report 111, <http://re.indiaenvironmentport.al.org.in/files/RR111.pdf>. Cited 10 July 2018
- Biswas BC, Nayar PS (1984) Quantification of drought and crop potential. *Mausam* 35:281–286
- Chowdhury A, Abhyankar VP (1984) A Markov chain model for the probability of drought incidence in India. *Mausam* 35:403–405
- Deodhar L (2000) Flood hydrology and geomorphic effectiveness in the Western Upland Maharashtra. Unpublished Ph. D thesis submitted to the University of Pune
- Deosthali V (2002) Dry farming in Maharashtra. In: Diddee J, Jog SR, Kale VS, Datye VS (eds) *Geography of Maharashtra*. Rawat Publications, Jaipur, pp 180–198
- Dikshit J (1983) Identification of drought prone areas and prediction of crop expectation in Maharashtra. *Geogr. Rev. India* 42:48–81
- Gadgil AS (2002) Rainfall characteristics of Maharashtra. In: Diddee J, Jog SR, Kale VS, Datye VS (eds) *Geography of Maharashtra*. Rawat Publications, Jaipur, pp 89–102
- Gokhale R, Sohani M (2015) Data driven behavioral characterization of dry-season groundwater-level variation in Maharashtra, India. *J Earth Syst Sci* 124:767–781
- GoM (1999) Maharashtra Water and Irrigation Commission Report. Government of Maharashtra, Water and Land Management Institute (WALMI), Aurangabad, p 1
- GoM (2014) Irrigation (Chapter IV). In: Maharashtra State Development Report. http://planningcommission.nic.in/plans/stateplan/sdr_maha/ch-4-14-02-05.pdf. Cited 12 July 2018
- Gore PG, Sinha Ray KC (2002) Variability of drought incidence over districts of Maharashtra. *Mausam* 53:533–538
- Guhathakurta P, Saji E (2013) Detecting changes in rainfall pattern and seasonality index vis-à-vis increasing water scarcity in Maharashtra. *J Earth Syst Sci* 122:639–649. <https://doi.org/10.1007/s12040-013-0294-y>
- IMD (2005) Climate of Maharashtra. India Meteorological Department, Pune
- IPCC (2013) Climate Change 2013: the physical science basis. In: Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
- Kalamkar SS (2011) Agricultural growth and productivity in Maharashtra: trends and determinants. Allied Publishers Pvt. Ltd., New Delhi, pp 1–218
- Kale VS (2002) Fluvial geomorphology of Indian rivers—an overview. *Prog Phys Geogr* 26:400–433
- Kale VS (2003) Geomorphic effects of monsoon floods on Indian Rivers. *Nat Hazards* 28:65–84
- Kale VS (2005) Fluvial hydrology and geomorphology of monsoon-dominated Indian rivers. *Rev Bras Geomorfol* 6:63–73
- Kale VS (2012) On the link between extreme floods and excess monsoon epochs in South Asia. *Clim Dyn* 39:1107–1122
- Kale VS, Todmal RS, Kulkarni P (2014) The monsoon-fed rivers of Maharashtra: their hydro-geomorphic characteristics and management. *Spec Publ J Geol Soc India* 3:26–34
- Khot V, Sabanna T (2018) Water resources and cropping pattern. Lulu Publication, Morrisville, pp 1–227
- Kumar KK, Rajagopalan B, Cane MA (1999) On the weakening relationship between the Indian monsoon and ENSO. *Science* 284:2156–2159
- Mehta AK, Satpathy T (2011) Escaping poverty: the Ralegan Siddhi Case. Indian Institute of Public Administration, New Delhi, CPRC-IIPA Working Paper No.38. p. 53
- Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ (2008) Stationarity is dead: whither water management? *Science* 319:573–574
- Mishra S (2006) Farmers' suicides in Maharashtra. *Econ Polit Wkly* 41:1538–1545
- Muthreja KN (1986) Applied hydrology. Tata McGraw Hill, New Delhi, pp 148–163
- NOAA (National Oceanic and Atmospheric Administration). website: http://archive.stats.govt.nz/browse_for_stats/environmental-reporting-series/environmental-indicators/Home/Atmosphere-and-climate/el-nino-oscillations/el-nino-oscillations-archived-19-10-2017.aspx. Accessed 11 May 2018
- Pahl-Wostl C (2007) Transitions towards adaptive management of water facing climate and global change. *Water Resour Manag* 21:49–62
- Pathak MD, Gadkari AD, Ghate SD (1999) Groundwater development in Maharashtra State, India. In: 25th WEDC conference on “Integrated development for water supply and sanitation”, Addis Ababa Ethiopia in 1999
- Pavelic P, Patankar U, Acharya S, Jella K, Gumma MK (2012) Role of groundwater in buffering irrigation production against climate variability at the basin scale in southwest India. *Agr Water Manag* 103:78–87
- Purandare P (2013) Water governance and droughts in Marathwada. *Econ Polit Wkly* 48:18–21
- Samal DR, Gedam SS (2012) Impact of land use dynamics on streamflow: a case study in upper Bhima basin, Maharashtra, India. In: 2012 IEEE International on geoscience and remote sensing symposium (IGARSS). 7165–7168. <https://doi.org/10.1109/igars.2012.6352010>
- Sanyal J, Lu XX (2004) Application of remote sensing in flood management with special reference to monsoon Asia: a review. *Nat Hazards* 33:283–301
- Sethi LN, Panda SN, Nayak MK (2006) Optimal crop planning and water resources allocation in a coastal groundwater basin, Orissa, India. *Agr Water Manag* 83:209–220
- Shewale MP, Kumar S (2005) Climatological features of drought incidences in India. Meteorological Monograph (Climatology 21/2005), National Climate Centre, India Meteorological Department. http://www.imdpune.gov.in/Clim_Pred_LRF_New/Reports/NCCResearchReports/

- Subramanian V (2004) Water quality in south Asia. *Asian J Water Environ Pollut* 1:41–54
- Suresh D (2015) Farmers Suicides in Maharashtra (October 7, 2015). Available at SSRN: <https://ssrn.com/abstract=2670554> or <http://dx.doi.org/10.2139/ssrn.2670554>
- Surinaidu L, Bacon CGD, Pavelic P (2013) Agricultural groundwater management in the Upper Bhima Basin, India: current status and future scenarios. *Hydrol Earth Syst Sci* 17:507–517
- TERI (2012) Climate change in Maharashtra: A pioneering adaptation strategy. The Energy and Resource Institute, New Delhi. Available at: http://www.metoffice.gov.uk/media/pdf/c/a/GOM_brochure_for_web.pdf Accessed 2 Sept 2014
- TERI (2014) Assessing Climate Change Vulnerability and Adaptation Strategies for Maharashtra: Maharashtra State Adaptation Action Plan on Climate Change (MSAAPC). The Energy and Resources Institute, New Delhi, 302. <http://www.moef.gov.in/sites/default/files/Maharashtra%20Climate%20Change%20Final%20Report.pdf>
- Todmal RS (2019) Droughts and agriculture in the semi-arid region of Maharashtra, western India. *Weather Clim Soc*. <https://doi.org/10.1175/WCAS-D-18-0131.1>
- Todmal RS, Kale VS (2016) Monsoon rainfall variability and rain-fed agriculture in the water-scarce Karha Basin, western India. *Mausam* 67:927–938
- Todmal RS, Korade MS, Dhorde AG, Zolekar RB (2018) Hydro-meteorological and agricultural trends in water-scarce Karha Basin, western India: current and future scenario. *Arab J Geosci*. <https://doi.org/10.1007/s12517-018-3594-3>
- Udmale PD, Ichikawa Y, Kiem AS, Panda SN (2014) Drought impacts and adaptation strategies for agriculture and rural livelihood in Maharashtra State of India. *Open Agric J* 8:41–47
- Wang LZ, Fang L, Hipel KW (2003) Water resources allocation: a cooperative game theoretic approach. *J Environ Inform* 2:11–22
- World Bank (2008) Climate change impacts in drought and flood affected areas: case studies in India. Available at: http://www.preventionweb.net/files/12563_WBIndia.pdf. Accessed 1 May 2014



Mechanism of biochar soil pore–gas–water interaction: gas properties of biochar-amended sandy soil at different degrees of compaction using KNN modeling

Ankit Garg¹ · He Huang¹ · Vinod Kushvaha² · Priyanka Madhushri³ · Viroon Kamchoom⁴ · Insha Wani² · Nevin Koshy² · Hong-Hu Zhu⁵

Received: 29 September 2019 / Accepted: 18 November 2019 / Published online: 25 November 2019

© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

Soil compaction has contrasting effect on soil strength (i.e., positive) and vegetation growth (i.e., negative), respectively. Biochar has been utilized mostly in combination with soils in both agricultural fields (i.e., loose soils) and geo-structures (i.e., dense soil slopes, landfill cover) for improving water retention due to its microporous structure. Biochar is also found to be useful to reduce gas permeability in compacted soil recently. However, the efficiency of biochar in reducing gas permeability in loose and dense soils is rarely understood. The objective of this study is to analyze effects of compaction on gas permeability in soil at different degrees of compaction (i.e., 65%, 80% and 95%) and also different biochar amendment contents (0%, 5% and 10%). Another aim is to identify relative significance of parameters (soil suction, water content, biochar content and compaction) in affecting gas permeability. Experiments were conducted before applying *k*-nearest neighbor (KNN) modeling technique for identifying relative significance of parameters. Biochar was synthesized from a coastal invasive species (water hyacinth), which has relatively no influence on food chain (as unlike in biochar produced from biomass such as rice husk, straw, peanut shell). Based on measurements and KNN modeling, it was found that gas permeability of biochar-amended soil is relatively lower than that of soil without amendment. It was found from KNN model that for denser soils, higher amount of soil suction is mobilized for a significant increase in gas permeability as compared to loose soils. Among all parameters, soil suction is found to be most influential in affecting gas permeability followed by water content and compaction.

Keywords KNN modeling · Biochar-amended soil · Coastal species · Compaction · Suction

Introduction

Biochar, a carbon rich material, which is usually generated from pyrolysis of biomass has been utilized for both agricultural soil (i.e., loose soil for crop growth; Kavitha et al. 2018) and dense soil (i.e., in landfill covers; Bordoloi et al. 2018) applications. Addition of biochar helps to improve soil water retention which is important for crop growth fertility (Lehmann and Joseph 2015) and also reduces infiltration (Gopal et al. 2019) in slopes. However, the interaction of biochar, however, in both loose and dense soils has not been systematically compared and analyzed. Soils in agriculture are generally compacted at 65% degree of compaction (Garg and Ng 2015), whereas in geo-environmental infrastructure, it may vary between 80 and 95% degree of compaction (Jien and Wang 2013).

Soils used in landfills are required to have low permeability toward water infiltration as well as gas permeation

✉ Vinod Kushvaha
vinod.kushvaha@iitjammu.ac.in

¹ Department of Civil and Environmental Engineering, Guangdong Engineering Center for Structure Safety and Health Monitoring, Shantou University, Shantou, China

² Department of Civil Engineering, IIT Jammu, Jammu, India

³ Staley Black and Decker, Atlanta, GA, USA

⁴ Department of Civil Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, Thailand

⁵ School of Earth Sciences and Engineering, Nanjing University, Nanjing, China

to minimize leakage of greenhouse gases (Mohareb et al. 2011; Mateus et al. 2012). Gas permeability depends on the void space or degree of compaction (Joseph et al. 2019). Generally, application of BC has been found to reduce gas permeability of dense soil (Garg et al. 2019). This is because of retention of higher amount of water in soil and biochar pores. However, the interaction of biochar in loose soil can be different from that of dense soil depending on the composition of biochar content. This is because there is porosity of biochar itself may compete with highly porous loose soil, which needs further investigation. Further, organic content in biochar can further enhance the bonding between large particles (Liu et al. 2012). As far as authors are aware, there is a lack of comprehensive study that investigates the effects of degree of compaction on gas permeability of biochar-amended soil. Further, there is a need to interpret the significance of soil suction and water content (due to complex soil and biochar porous structure) on gas permeability and possibly explores their relationship. Empirical modeling using *k*-nearest neighbor (KNN) has been adopted widely for establishing relationship. Analysis of predictability of soil load-bearing capacity can be done by machine learning by the use of data sets consisting of the data collected from the field measurements. Good predictions were recorded up to 200 m from the closest point with the known bearing capacity (Pohjankukka et al. 2016). The effects of the microbial activity on the permeability of the compacted clay soils were analyzed, and the data obtained were applied to the KNN for prediction of permeability of soils in the sites of landfill (Ozcoban et al. 2018). In another study, KNN has been used for derivation of pedotransfer functions (PTF) for the soil properties (Jagtap et al. 2004). KNN is used for the prediction of soil series and provides a suitable crop yield suggestion for a specific type of soil (Zaminur Rahman et al. 2018). For the estimation of the water content

present in the soil, KNN was used and the correlations were shown in terms of R^2 value (Nemes et al. 2006). A KNN algorithm, nonparametric approach, was developed for prediction of cation exchange capacity from the measured soil properties by considering two parameters which were optimized before implementation (Zolfaghari et al. 2016). Such model development would help to improve understanding of soil–BC–water interactions in loose and dense soils, which are commonly found in agriculture and engineering applications, respectively.

The major objective of this study is to develop a model for computing gas permeability with respect to soil suction, water content and temperature for different degrees of compaction of soil. Further, the relative significance of soil suction, water content and temperature has been analyzed. Three different degrees of compaction (i.e., 65%, 80% and 95%) have been considered. Based on measurements of gas permeability as well as other soil parameters, a new empirical model using KNN technique was developed to understand the mechanism of soil–BC interaction under varying degrees of compaction (i.e., void ratio). This study aims to improve understanding of soil–BC–water interaction under different degrees of compaction.

Materials and methods

Gas permeability and water retention property tests of biochar-amended soil

All the experimental tests were conducted in the greenhouse established at Shantou University, China. Soil selected for investigation was collected from the campus of Shantou University, China. The particles in soil corresponding to size range of 1.18–2.36 mm and 2.36–4.75 mm were found

Table 1 Properties of soil adopted for investigation

Properties	Standard	Soil	BC	BAS5%	BAS10%
Particle size distribution	ASTM D 422			–	–
2–4.75 mm		50.0		–	–
0.85–2 mm		29.7		–	–
0.425–0.8 mm		18.11		–	–
0.15–0.425 mm		1.16	100% > 0.425 mm	–	–
0.075–0.15 mm		0.81		–	–
0–0.075 mm		0.22		–	–
Atterberg limits	ASTM D 4318				
Liquid limit (LL/%)		29.2		–	–
Plastic limit (PL/%)		18.3		–	–
Plastic index (PI/%)		10.9		–	–
MDD (kN/m ³)	ASTM D 698	16.9		16.6	16.5
OMC (%)	ASTM D 698	18.8		19.5	19.8
Specific gravity	ASTM D 854	2.590	2.260	2.579	2.551

to be 29.7% and 50%, respectively. The properties of soil are summarized in Table 1. Based on unified soil classification system (USCS; ASTM D2487-17 2017), the soil can be categorized as SP (poorly graded sand). Three series of gas permeability tests were conducted including one series on bare soil and other two series on soil amended with 5% and 10% biochar content, respectively. These two different biochar contents were chosen owing to its improved soil properties as explored by Reddy et al. (2015). Soils were compacted at three different degrees of compaction (65%, 80% and 95%). All these tests were conducted in 9 soil columns with a diameter 300 mm and a height of 250 mm. These columns were manufactured (in-house) using PVC.

Biochar was also prepared in-house using water hyacinth collected locally. The details of the preparation of biochar are provided in Bordoloi et al. (2018). For understanding structure of biochar, FESEM and FTIR analyses were conducted as shown in Fig. 1. SEM images revealed high specific surface area and porosity, while FTIR analysis indicates

three major surface functional groups (i.e., $-\text{OH}$, $-\text{COOH}$ and $-\text{CO}$; shown in Fig. 1b). These functional groups indicate the biochar produced from locally collected water hyacinth is hydrophilic in nature.

For measuring gas permeability under different soil suction and water content, compacted soil columns were subjected to four 49-day monitoring cycle that for each cycle includes 42-day drying period and a 7-day continuous wetting period. The measurement system used for monitoring is shown in Fig. 2b, c. Soil suction was measured using MPS-6 sensor, which has a range from 10 to 100,000 kPa (Decagon Devices 2016), whereas volumetric water content was measured using soil water content probe EC-5 probe (Decagon Devices 2016). Both these sensors were inserted 100 mm into the soil from the top of the column. Average relative humidity of 69.9%, average temperature of 14.2° – 38.1° and evaporation rate of 0.28–11.07 mm/d were measured during the whole period. Gas apparatus as developed by Garg et al. (2019) is considered for investigation. CO_2 gas has been considered in this study for its non-reactivity with soil. Based on Darcy's law, the gas

Fig. 1 Structure of biochar produced in-house from water hyacinth using **a** FESEM, **b** FTIR images

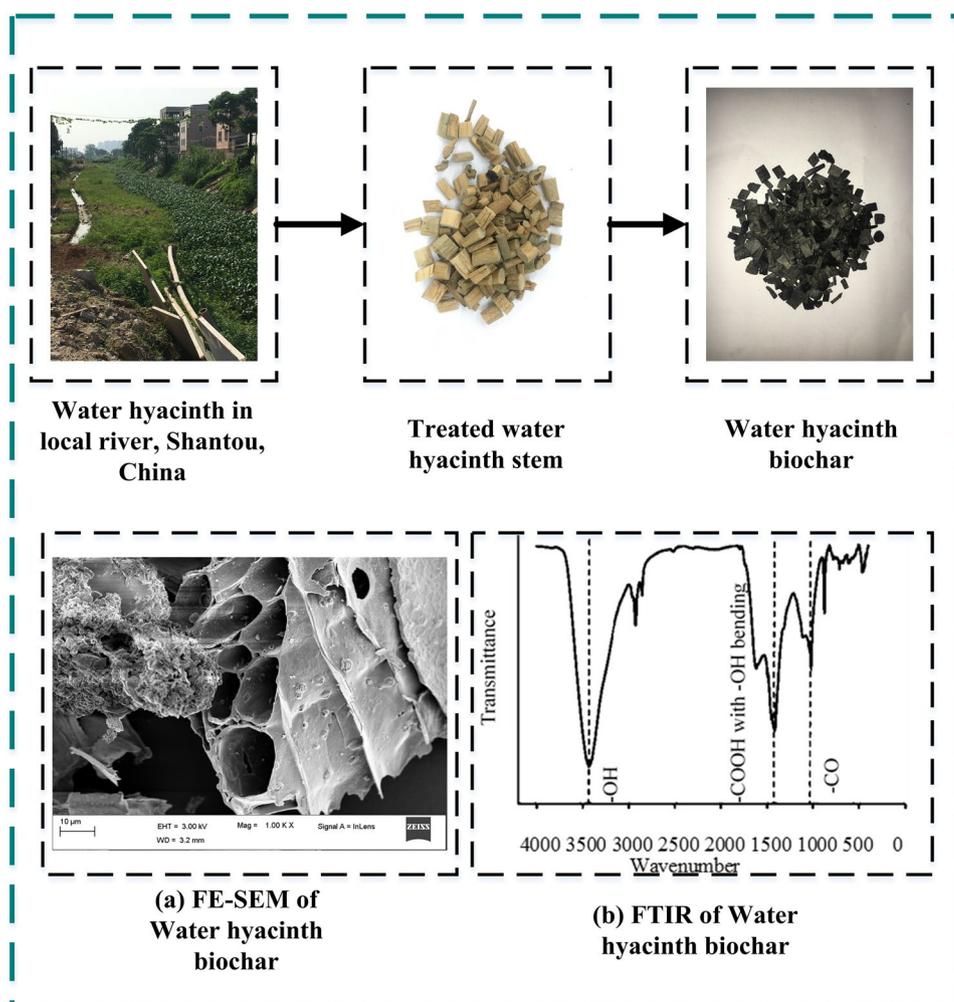
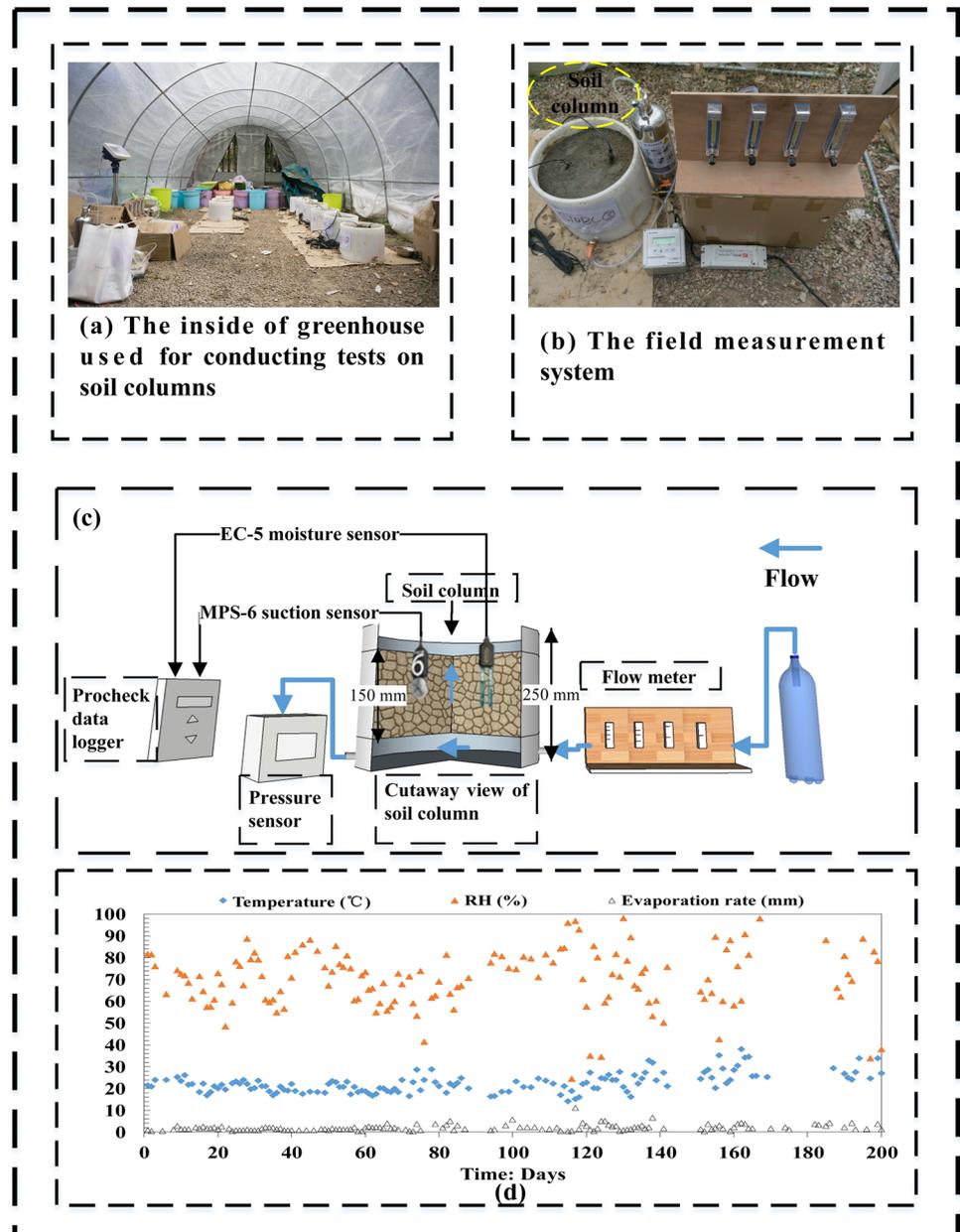


Fig. 2 Test column in greenhouse at Shantou University, China. **a** Greenhouse interior; **b** original diagram of soil column measurement system; **c** schematic diagram of soil column measurement system; **d** climate conditions



permeability of soil column can be determined by the following equation (Damkjær and Korsbech 1992):

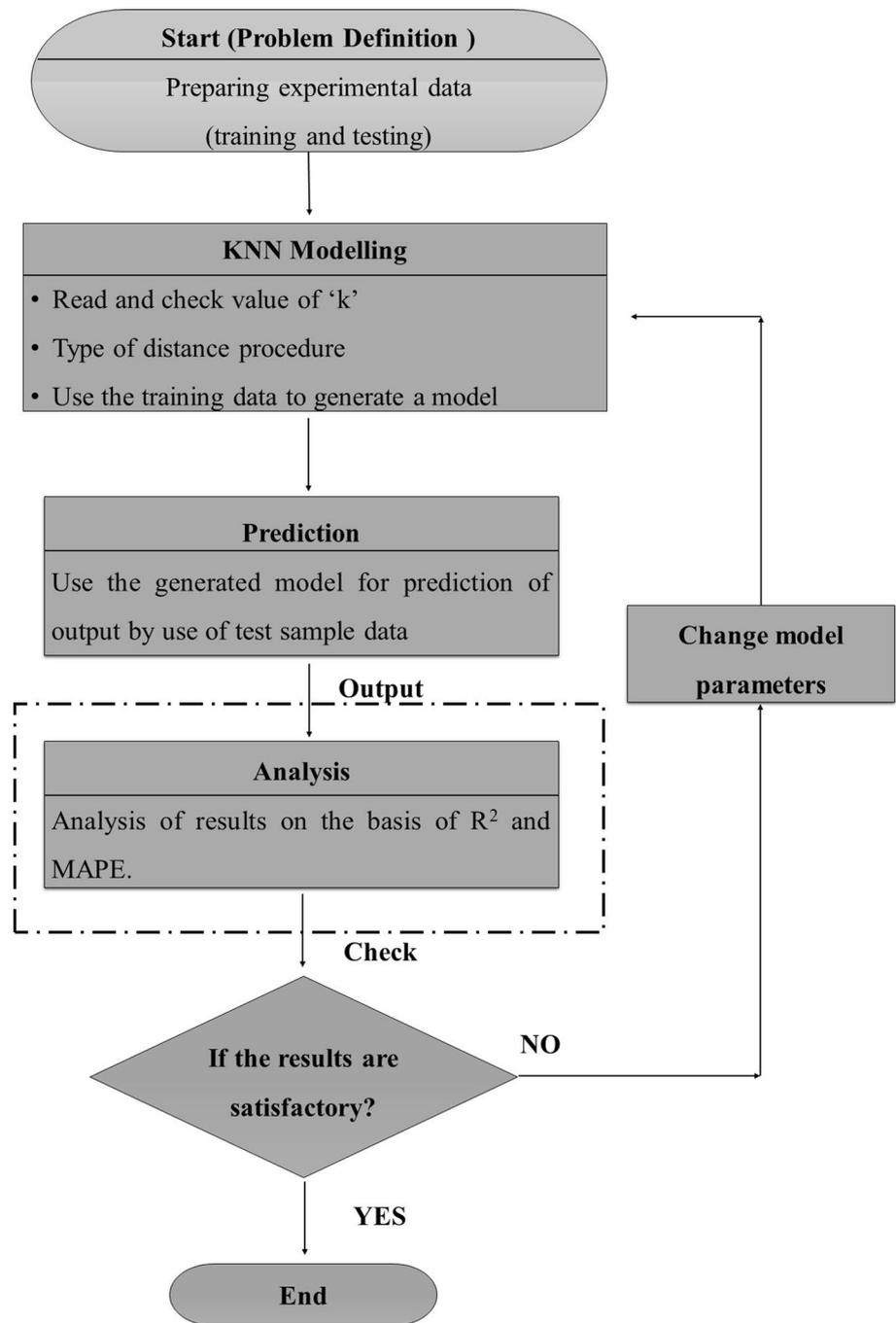
$$\text{Gas permeability} = \frac{\mu q L}{A \Delta P} \quad (1)$$

where ΔP is difference of gas pressure between the bottom and the top of soil column; q is the flow rate; A is the soil sample cross-sectional area; L is the length of the soil column; and μ is the absolute viscosity of the carbon dioxide gas flow ($14.8 \times 10^{-6} \text{ N s/m}^2$).

KNN modeling

In the present study, KNN algorithm was implemented in MATLAB (version 2019A). The KNN method does not use any predefined mathematical functions for the prediction of the target variable. In this methodology, based on k -nearest neighbor values of the input parameters, prediction of the target output is carried out. Most commonly, the nearest neighbor distances are calculated using the Euclidean distance algorithm.

The flowchart of the KNN modeling which includes the training and testing steps is shown in Fig. 3. The

Fig. 3 Flowchart of KNN modeling

first step is to prepare the experimental data and identify input and output parameters. The experimental data were divided into the training and testing data sets. Subsequently, to generate the model using training data set, the distances from the k -nearest neighbor were calculated using Euclidean distance function $\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$. In this distance function, the choice of k has a drastic impact on the prediction. In KNN modeling, the best selection of k depends on the data and is chosen among

the various weighting schemes of the samples selected. Therefore, an arbitrary value k was initialized and R^2 and mean absolute percentage error (MAPE) were calculated. The same procedure was repeated for different values of k until the maximum value of R^2 and the minimum value of MAPE were achieved. In this case, $k = 2$ provided the best prediction based on R^2 and MAPE. The model was used to predict the target output for test data set, and the predicted values were validated against the experimentally measured values. The measured and

predicted values were analyzed for its accuracy based on R^2 and MAPE.

The error in the KNN modeling was determined by comparing the predicted result with the experimentally measured values. The mean absolute percentage error MAPE and R -squared (R^2) were employed to assess how close prediction values were to the actual values, and the following equation was used for error analysis:

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \frac{|T_i - P_i|}{T_i} \quad (2)$$

where T_i is the target value (measured experimental value), P_i is the predicted value, and n is the number of data. The value of MAPE should be minimum to ensure the accuracy of the prediction model. R -squared (R^2) is a measure of closeness of the data to the fitted regression line. It is the percentage of the response of variable variation that is explained by the linear model. The higher the value of R -squared, the better the model fits the data. If the final data obtained are not satisfactory, the parameters of the model are changed to achieve the desired output; otherwise, the results are satisfactory.

Since KNN model is incapable of predicting the normalized importance, random forest method (another machine learning algorithm) was used to predict the normalized importance plots. The normalized importance plot provides the ranking of importance of each input parameter on the target output.

The suction, volumetric water content, soil compaction and biochar, at various levels, were employed for the KNN training and testing phases. For training the model, data corresponding to 0% and 10% biochar content for all degrees of compaction (65%, 80% and 95%) was selected. The testing

was done on data corresponding to 5% biochar content at all degrees of compaction. Similarly, in another case, data corresponding to 65% and 95% degrees of compaction (for all biochar contents) was used for training of model. The testing was conducted using 80% degree of compaction for all biochar contents. Figure 4 describes the network design used for the prediction of air permeability.

Results and discussion

Air permeability versus soil suction for soil–biochar composites at different degrees of compaction

Figure 5a–c shows the comparison between measured and computed air permeability versus suction curves for different degrees of soil compaction. The biochar content of 5% was kept as constant during prediction. It can be observed from these figures that the prediction using KNN technique seems reasonable considering the fluctuation in measured data of air permeability as a function of soil suction. For different degrees of compaction, air permeability generally rises with an increase in soil suction. This is primarily due to an increase in connectivity of pores (with air) with a lowering of volumetric water content. The observed trend is consistent with Garg et al. (2019) for 95% soil compaction. However, when we compare trends among different degrees of compaction, air permeability gradually increases with suction for 65% degree of compaction, while it increases significantly beyond a certain suction (~ 450 kPa) for 95% degree of compaction. This indicates that for higher degree of compaction, larger amount of soil suction is needed to enhance air

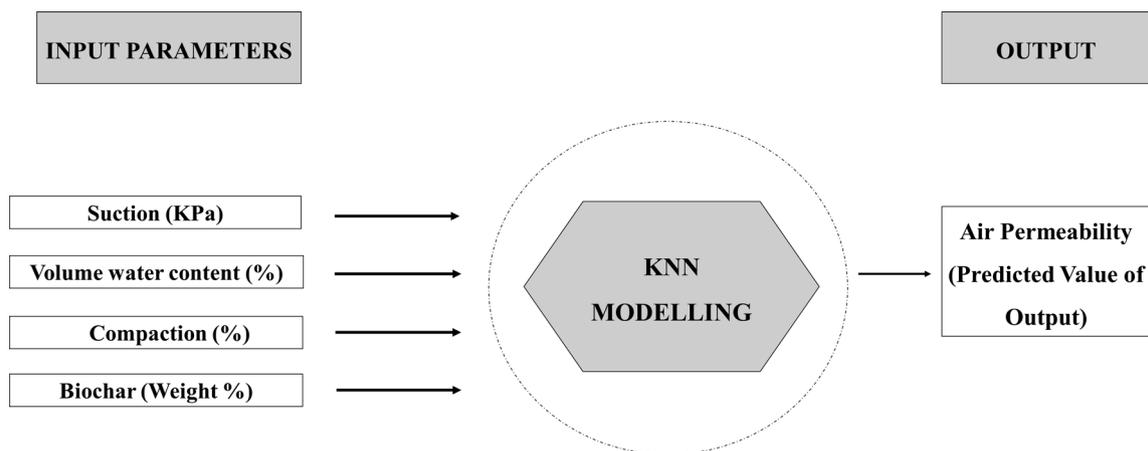


Fig. 4 KNN architecture applied to air permeability prediction

Fig. 5 KNN prediction of 5% biochar—air permeability versus suction plot on log scale **a** for 65% soil compaction, **b** 80% soil compaction, **c** 95% soil compaction, **d** normalized importance of input parameters

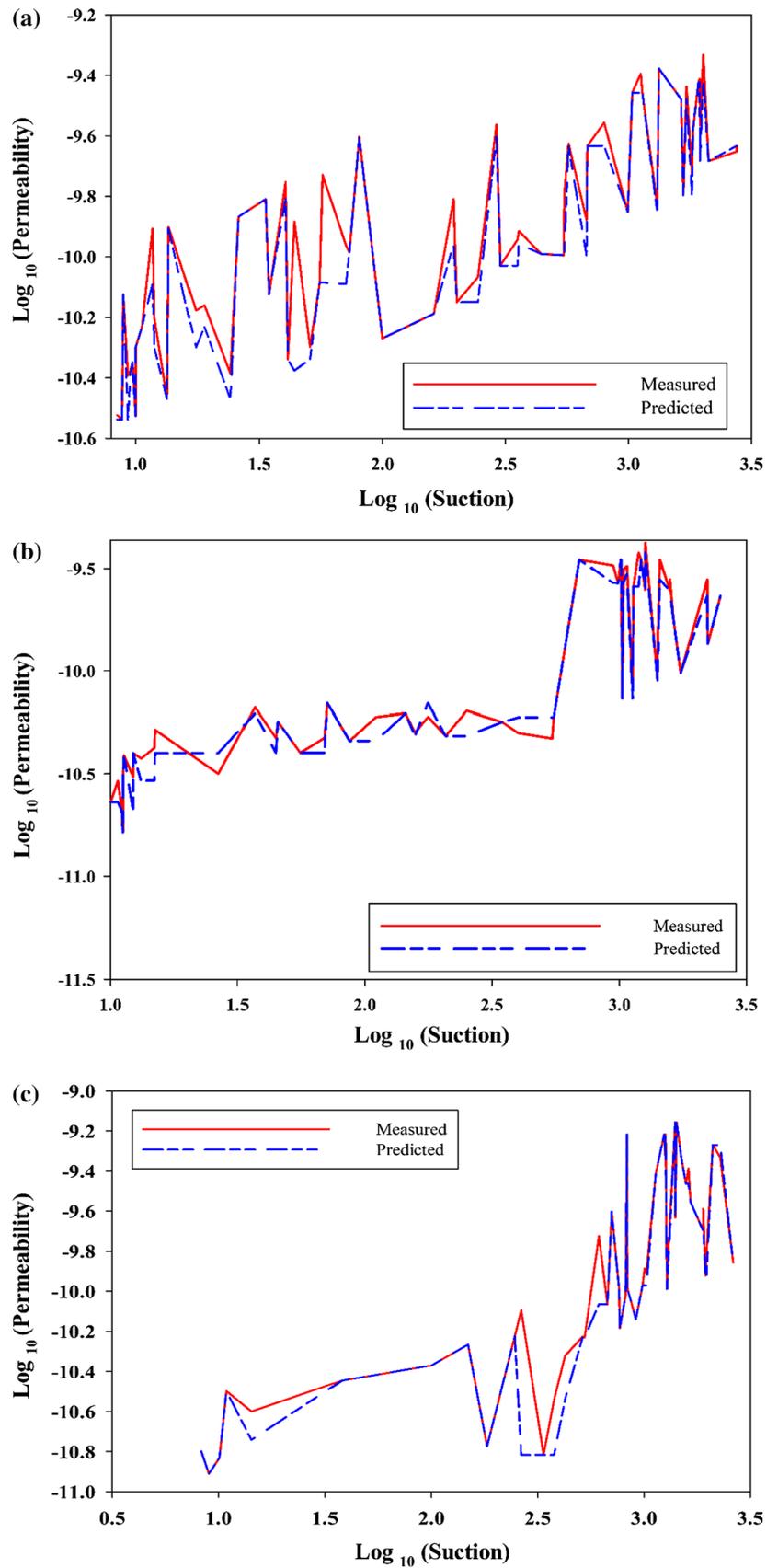
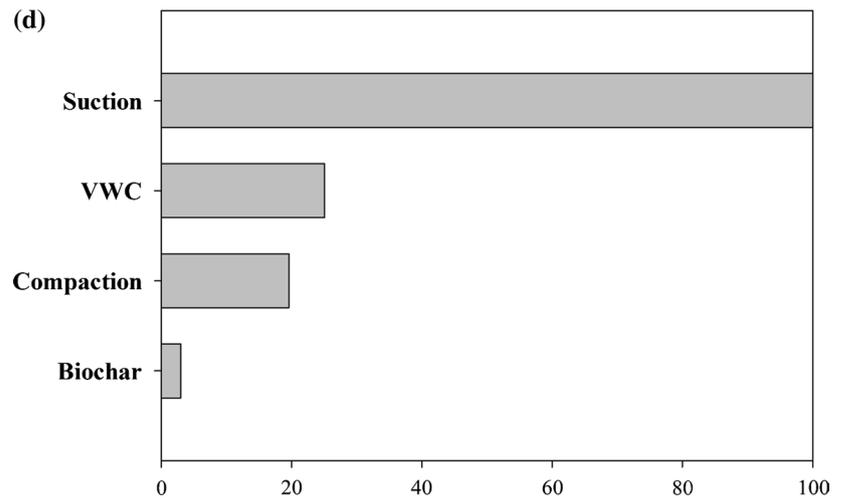


Fig. 5 (continued)



permeability. For application in landfill covers located in subtropical regions of heavy rainfall (suctions within lower range of up to 400 kPa), where minimum gas permeability is desirable, higher degree of compaction is more suitable. It is also interesting to note that air permeability of biochar-amended soil at lower suction range (between 10 and 350 kPa) is highest for loose soil (i.e., 65% compaction) followed by 80% and 95%. However, the difference in air permeability between loose and dense soils becomes lower at higher suction ranges. This seems to imply that compaction may not have a huge difference in air permeability of soil under very dry state. This is reasonable as under dry state, most of voids of soil are filled with air, thus providing improved connectivity for gas to flow.

In order to further interpret the soil–biochar–water–gas interaction, the relative importance of parameters (suction, volumetric water content (VWC), biochar content and compaction) affecting air permeability was investigated. Figure 5d shows the relative importance of parameters affecting air permeability. It was found that soil suction is the important parameter affecting air permeability of biochar-amended soil followed by volumetric water content, compaction (i.e., void ratio) and biochar content. The result seems to be consistent with the theory suggested by Fredlund and Morgenstern (1977) that soil suction is one of the important stress state parameters for understanding of unsaturated soil behavior. As per Fig. 5d, the second most important parameter is volumetric water content or compaction. This is expected that it itself has direct relationship with soil suction through soil–water characteristic curve (SWCC; Bordoloi et al. 2018). On the other hand, theoretical models proposed in the literature such as Carman–Kozeny equation (Carman 1956; Valdes-Parada et al. 2009) and Yang–Alpin equation (Yang and Aplin 2010) also suggest that dependence of gas permeability on porosity (or void ratio). Such models, however, have been proposed for granular material

Table 2 Coefficient of correlation and maximum percentage error for 5% biochar prediction using KNN

Prediction of air permeability		R^2	MAPE (%)
With biochar (%)	Soil compaction (%)		
5	65	0.92	8
	80	0.97	9
	95	0.98	8

and clay material, respectively. They further do not take into account directly any influence of soil suction, that itself is a measure of air–water interfaces in an unsaturated soil. Saiyouri et al. (2008) investigated gas permeability on cement grouted soils. They observed that gas permeability reduced significantly due to cementation of soil (i.e., reduction in void ratio). However, their study was performed on saturated soils without taking into account variation in soil moisture. Table 2 summarizes the R^2 values of predicted air permeability and suction curves for soil amended with biochar at different degrees of compaction. It can be observed that the R^2 values are greater than 0.9, suggesting reasonably good prediction capacity of KNN.

Air permeability versus soil suction for soil–biochar composites at different biochar contents

Figure 6a–c shows the plot of air permeability and suction for soil amended with different biochar contents such as 0, 5% and 10%. As per figures, it can be observed that generally with the presence of biochar, there is a reduction effect on gas permeability, especially for biochar content of 10%. The observation is consistent with that observed by Wong et al. (2016), where they found that with an increase in biochar content the gas permeability reduces. However, interestingly, their study found that at 80% degree of compaction,

Fig. 6 KNN prediction of 80% soil compaction—air permeability versus suction plot on log scale **a** for 0% biochar, **b** 5% biochar, **c** 10% biochar, **d** normalized importance of input parameters

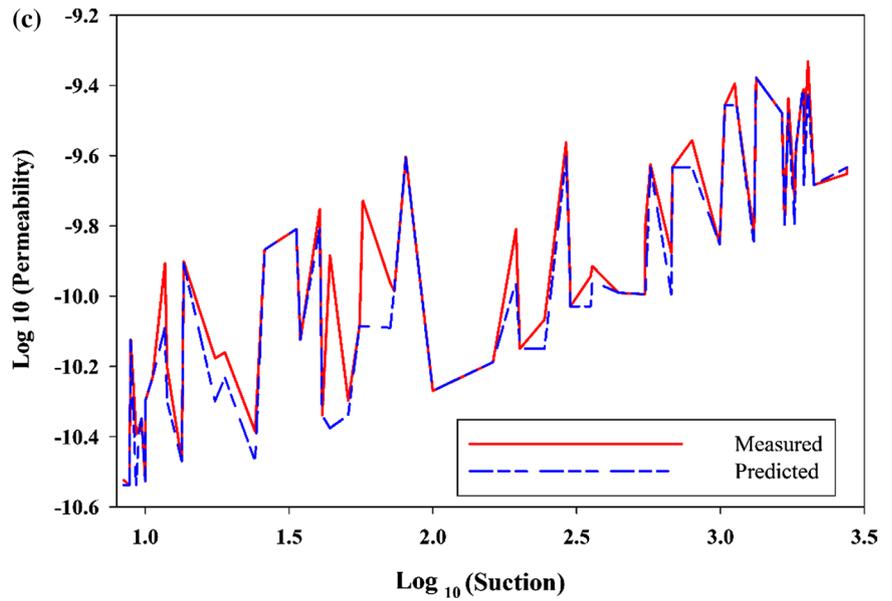
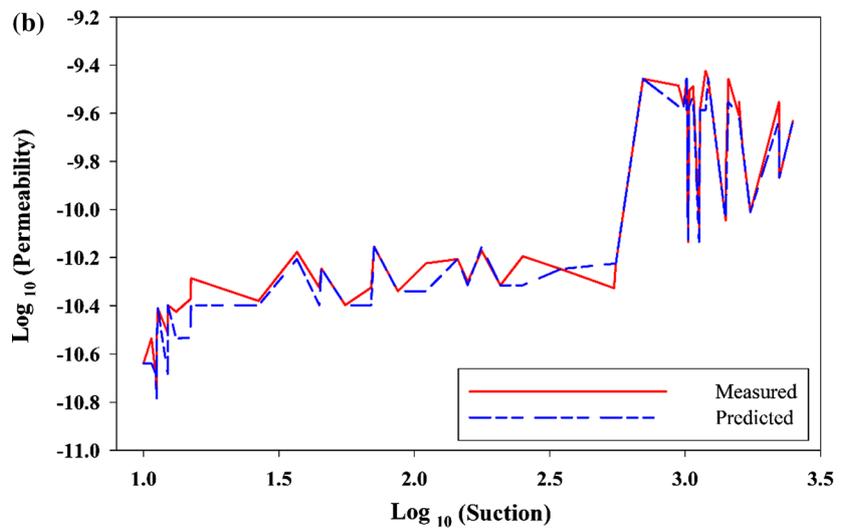
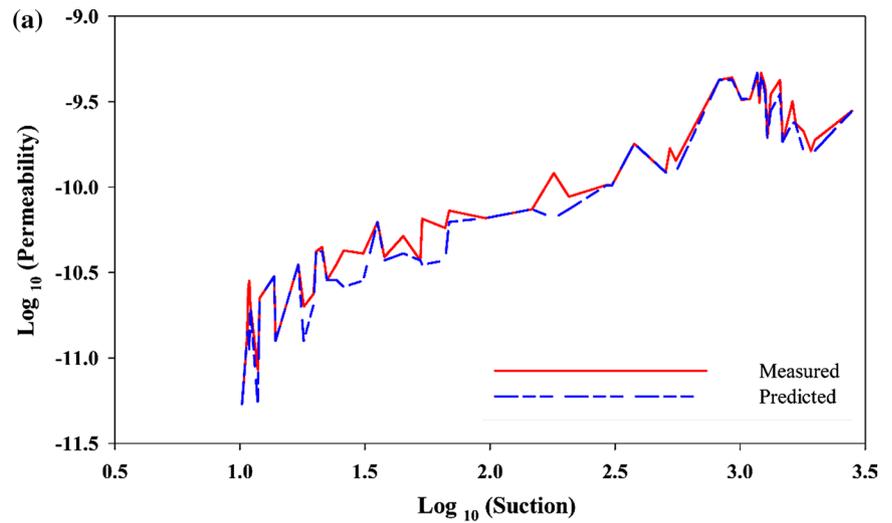


Fig. 6 (continued)

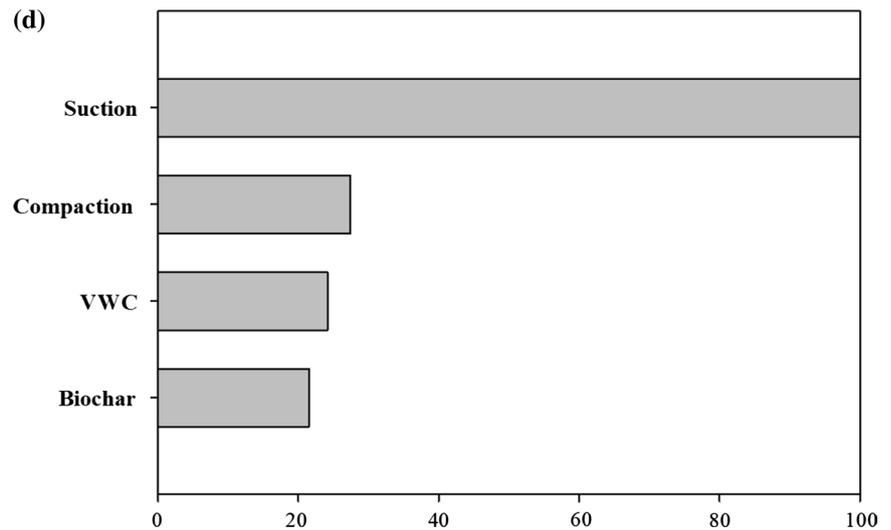


Table 3 Coefficient of correlation and maximum percentage error for 80% soil compaction using KNN

Prediction of air permeability		R^2	MAPE (%)
Soil compaction (%)	With biochar (%)		
80	0	0.97	12
	5	0.96	8
	10	0.96	9

the biochar effects are negligible. On the contrary, our study found that biochar content of 5% could be also impactful in reducing gas permeability at 65% degree of compaction. This variation can be due to difference in soil type (silty sand in our study as compared to clay in their study) as well as biochar type (produced from invasive weed as compared to commercial peanut shell in their study). Such differences can cause variation in packaging arrangement of soil–biochar particles and hence any observed gas permeability. Kumar et al. (2019) observed that feedstock type (animal or plant) from which biochar is produced can significantly influence erosion properties of soil. Kumar et al. (2019) observed that biochar produced from animal and plant waste can significantly alter its influence on soil erosion properties. The reason is due to different functional groups present in animal- and plant-based biochar that can cause its structure to be hydrophilic or hydrophobic in nature. Further systematic studies are needed to consider influence of soil type and biochar type (animal source or plant source). Similar to Fig. 5d, soil suction is found to be the most significant parameter in influencing gas permeability of soils amended at different biochar contents. Volumetric water content and void ratio have similar importance followed by biochar content to be the least. Based on the second KNN predictions, it can be

confirmed that soil suction is highly essential in governing gas permeability in unsaturated soils amended with biochar (Table 3).

Conclusions

The study involves the investigation of gas permeability of biochar-amended soils at different degrees of compaction. Further, KNN modeling has been applied to develop model and identify relative importance of parameters (soil suction, water content, biochar content and compaction) that influence gas permeability. As per study, it was found that biochar-amended soils have lower gas permeability as compared to unamended soils. Further, it was also found that higher degree of compaction requires larger suction (or drying) to enhance gas permeability. Gas permeability among different compacted soils (for 5% biochar content) was highest at lower suction range. Soil suction is found to be the most influential parameter for gas permeability in biochar-amended soils.

Acknowledgements The authors would like to acknowledge the National Natural Science Foundation (NSFC) Youth Project (Grant number 41907252) and also Shantou University Scientific Research Fund (NTF17007), China. The first and fourth authors would also like to acknowledge International Collaboration Fund from Academic Melting Pot Program (KREF206237) from King Mongkut's Institute of Technology Ladkrabang (KMITL).

Compliance with ethical standards

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

References

- ASTM D2487-17 (2017) Standard practice for classification of soils for engineering purposes (unified soil classification system). ASTM International, West Conshohocken. <https://doi.org/10.1520/D2487-17>
- Bordoloi S, Garg A, Sreedeeep S, Lin P, Mei G (2018) Investigation of cracking and water availability of soil–biochar composite synthesized from invasive weed water hyacinth. *Biores Technol* 263:665–677. <https://doi.org/10.1016/j.biortech.2018.05.011>
- Carman PC (1956) Flow of gases through porous media. Butterworth Scientific Publications, New York
- Damkjaer A, Korsbech U (1992) A small-diameter probe for in situ measurements of gas permeability of soils. *Radiat Prot Dosim* 45(1–4):85–89. <https://doi.org/10.1016/j.rsr.2016.06.003>
- DDI (Decagon Devices Inc) (2016) MPS-2 and MPS-6 dielectric water potential sensors operator’s manual. Decagon Devices, Pullman. http://manuals.decagon.com/Retired%20and%20Discontinued/Manuals/13755_MPS-2and6_Web.pdf
- Fredlund DG, Morgenstern NR (1977) Stress state variables for unsaturated soils. *J Geotech Geoenviron Eng (ASCE)* 129(19):103:447–466
- Garg A, Ng CWW (2015) Investigation of soil density effect on suction induced due to root water uptake by *Schefflera heptaphylla*. *J Plant Nutr Soil Sci* 178(4):586–591. <https://doi.org/10.1002/jpln.201400265>
- Garg A, Bordoloi S, Ni J, Cai W, Maddibiona PG, Mei G, Poulsent TG, Lin P (2019) Influence of biochar addition on gas permeability in unsaturated soil. *Géotech Lett*. <https://doi.org/10.1680/jgele.18.00190>
- Gopal P, Bordoloi S, Ratnam R, Lin P, Cai W, Buragohain P, Sreedeeep S (2019) Investigation of infiltration rate for soil–biochar composites of water hyacinth. *Acta Geophys* 67(1):231–246. <https://doi.org/10.1007/s11600-018-0237-8>
- Jagtap SS, Lall U, Jones JW, Gijnsman AJ, Ritchie JT (2004) Dynamic nearest-neighbor method for estimating soil water parameters. *Trans ASAE* 47:1437–1444
- Jien SH, Wang CS (2013) Effects of biochar on soil properties and erosion potential in a highly weathered soil. *CATENA* 110:225–233. <https://doi.org/10.1016/j.catena.2013.06.021>
- Joseph J, Kuntikana G, Singh DN (2019) Investigations on gas permeability in porous media. *J Nat Gas Sci Eng* 64:81–92. <https://doi.org/10.1016/j.jngse.2019.01.017>
- Kavitha B, Reddy PVL, Kim B, Lee SS, Pandey SK, Kim KH (2018) Benefits and limitations of biochar amendment in agricultural soils: a review. *J Environ Manag* 227:146–154. <https://doi.org/10.1016/j.jenvman.2018.08.082>
- Kumar H, Ganesan SP, Bordoloi S, Sreedeeep S, Lin P, Mei G, Sarmah AK (2019) Erodibility assessment of compacted biochar amended soil for geo-environmental applications. *Sci Total Environ* 672:698–707. <https://doi.org/10.1016/j.scitotenv.2019.03.417>
- Lehmann J, Joseph S (eds) (2015) *Biochar for environmental management: science, technology and implementation*. Routledge, Abingdon
- Liu X, Han F, Zhang X (2012) Effect of biochar on soil aggregates in the loess plateau: results from incubation experiments. *Int J Agric Biol* 14(6):975–979
- Mateus MDSCS, Machado SL, Barbosa MC (2012) An attempt to perform water balance in a brazilian municipal solid waste landfill. *Waste Manag* 32:471–481. <https://doi.org/10.1016/j.wasman.2011.11.009>
- Mohareb EA, MacLean HL, Kennedy CA (2011) Greenhouse gas emissions from waste management—assessment of quantification methods. *J Air Waste Manag Assoc* 61(5):480–493. <https://doi.org/10.3155/1047-3289.61.5.480>
- Nemes A, Rawls WJ, Pachepsky YA, Van Genuchten MT (2006) Sensitivity analysis of the nonparametric nearest neighbor technique to estimate soil water retention. *Vadose Zone J* 5(4):1222–1235. <https://doi.org/10.2136/vzj2006.0017>
- Ozcoban MS, Isenkul ME, Güneş-Durak S, Ormanci-Acar T, Övez S, Tüfekci N (2018) Predicting permeability of compacted clay filtrated with landfill leachate by *k*-nearest neighbors modelling method. *Water Sci Technol* 77(8):2155–2164. <https://doi.org/10.2166/wst.2018.139>
- Pohjankukka J, Riihimäki H, Nevalainen P, Pahikkala T, Ala-Ilomäki J, Hyvönen E, Heikkonen J (2016) Predictability of boreal forest soil bearing capacity by machine learning. *J Terramech* 68:1–8. <https://doi.org/10.1016/j.jterra.2016.09.001>
- Rahman SAZ, Mitra KC, Islam SM (2018) Soil classification using machine learning methods and crop suggestion based on soil series. In: 2018 21st international conference of computer and information technology (ICCIT). IEEE, New York, pp 1–4. <https://doi.org/10.1109/ICCITECHN.2018.8631943>
- Reddy KR, Yaghoubi P, Yukselen-Aksoy Y (2015) Effects of biochar amendment on geotechnical properties of landfill cover soil. *Waste Manag Res* 33(6):524–532. <https://doi.org/10.1177/0734242X15580192>
- Saiyouri N, Bouasker M, Khelidj A (2008) Gas permeability measurement on injected soils with cement grout. *Cem Concr Res* 38(1):95–103. <https://doi.org/10.1016/j.cemconres.2007.08.015>
- Valdes-Parada FJ, Ochoa-Tapia JA, Alvarez-Ramirez J (2009) Validity of the permeability Carman–Kozeny equation: a volume averaging approach. *Physica A* 388(6):789–798
- Wong JTF, Chen Z, Ng CWW, Wong MH (2016) Gas permeability of biochar-amended clay: potential alternative landfill final cover material. *Environ Sci Pollut Res* 23(8):7126–7131. <https://doi.org/10.1007/s11356-015-4871-2>
- Yang Y, Aplin AC (2010) A permeability–porosity relationship for mudstones. *Mar Pet Geol* 27(8):1692–1697. <https://doi.org/10.1016/j.marpetgeo.2009.07.001>
- Zolfaghari AA, Taghizadeh-Mehrjardi R, Moshki AR, Malone BP, Weldeyohannes AO, Sarmadian F, Yazdani MR (2016) Using the nonparametric *k*-nearest neighbor approach for predicting cation exchange capacity. *Geoderma* 265:111–119. <https://doi.org/10.1016/j.geoderma.2015.11.012>



Spatiotemporal variability in spate irrigation systems in Khirthar National Range, Sindh, Pakistan (case study)

Abdul Ghani Soomro^{1,2} · Muhammad Munir Babar¹ · Muhammad Arshad³ · Anila Memon^{1,2} · Babar Naeem¹ · Arshad Ashraf^{2,4}

Received: 24 May 2019 / Accepted: 2 December 2019 / Published online: 10 December 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

Satellite remote sensing and geographical information system (GIS) have been used successfully to monitor and assess the land use and land cover (LULC) dynamics and their impacts on people and the environment. LULC change detection is essential for studying spatiotemporal conditions and for proposing better future planning and development options. The current research analyzes the detection of spatiotemporal variability of spate irrigation systems using remote sensing and GIS in the Khirthar National Range, Sindh Province of Pakistan. We use Landsat images to study the dynamics of LULC using ArcGIS software and categorize five major LULC types. We obtain secondary data related to precipitation and crop yield from the provincial department of revenue. The maximum likelihood supervised classification (MLSC) procedure, augmented with secondary data, reveals a significant increase of 86.25% in settlements, 83.85% in spate irrigation systems, and 65% in vegetation, and a substantial negative trend of 39.50% in water bodies and 20% in barren land during the period from 2013 to 2018. Our study highlights an increase in settlements due to the inflow of local population for better means of living and an increase in spate irrigation systems, which indicates the water conservation practices for land cultivation and human purpose lead to the shrinkage of water bodies. The confusion matrix using Google Earth data to rectify modeled (classified) data, which showed an overall accuracy of 82.8%–92%, and the Kappa coefficient estimated at 0.80–0.90 shows the satisfactory results of the LULC classification. The study suggests the need to increase water storage potential with the appropriate water conservation techniques to enhance the spate irrigation system in the hilly tracts for sustainable developments, which mitigates drought impact and reduces migration rate by providing more opportunities through agricultural activities in the study area.

Keywords Spate irrigation · LULC · Remote sensing · GIS · Water bodies · Vegetation · Confusion matrix · Khirthar National Range

✉ Abdul Ghani Soomro
ag2005parc@gmail.com

Muhammad Munir Babar
mmunirbabar.uspcasw@faculty.muets.edu.pk

Muhammad Arshad
marshad205@gmail.com

Anila Memon
aneelahameem@gmail.com

Babar Naeem
babar.gis@gmail.com

Arshad Ashraf
mashr22@yahoo.com

¹ United States-Pakistan Center for Advanced Studies in Water (USPCAS-W), Mehran University of Engineering and Technology (MUET), Jamshoro, Pakistan

² Pakistan Agricultural Research Council (PARC), Islamabad, Pakistan

³ Department of Chemical Engineering, King Khalid University (KKU), Abha, Saudi Arabia

⁴ PARC/NARC, Islamabad, Pakistan

Introduction

Arid and semiarid regions have diverse, challenging issues in water management, mainly areas, which are entirely dependent on the rainfall pattern as a primary water source (Markhi et al. 2019). Land cover refers to the physical land type, such as the area covered by various modes of land cultivation and water preservation techniques. Land use refers to the existing activities for development, conservation, and other practices (NOAA 2015 as cited in Chowdhury et al. 2018). LULC change detection is an essential exploratory technique to study the temporal situations of land dynamics, and it is a useful tool for the planning and management of water resources.

There are broad collections of the literature that links LULC change detection with population dynamics (Autónoma et al. 2011). Recently, remote sensing has helped to access satellite imagery data (Adolfo and Rosalen 2007) with cost-effective and time-saving techniques.

Jain Figueroa (2019) focused the future food challenges and cited the review of Tilman et al. 2011; he projected the increasing food demand by about 50% more in 2050 globally. Jain 2019 discovered that the competition of food demand and the enhancement of more land under-cultivation might adversely affect the existing ecology.

According to Rawat and Kumar (2015), the variation of LULCs of a particular region reveals the natural and socioeconomic characteristics of that territory in terms of time and space. For sustainable management of natural resources, it is critical to understand the extent, management, and causes of LULC dynamics. The spatial–temporal features of the LULC provide an essential dataset for improved decision-making. LULC identifications are helpful in many ways, such as assessment of water bodies and evaluation of watersheds (Butt et al. 2015) and delineation of flood areas through LULC classification (Chignell et al. 2015; Fiorella and Julián 2018), and analyze the detection and assessment of climate impacts (Rokni et al. 2014). Anthropogenic activities are mainly responsible for LULC changes due to increasing urbanization, decreasing agriculture, and forestry, which lead to a negative impact on the existing environment and disturb the ecosystem (Girma and Hassan 2014). LULC assessments help in explaining the extent of human influence on the natural environment (Chowdhury et al. 2018). The social impact on the land cover includes land clearing for agriculture, deforestation, and urban expansion for socioeconomic benefit. However, LULC fluctuations are generating unintended multi-directional effects on the ecosystem (Birhanu 2018), such as increased surface runoff and reductions in the recharge of groundwater due to increases in impervious surfaces

(Pan et al. 2011). In terms of agricultural practices, there is a direct relationship between LULC and the amount of precipitation in the catchment areas. Satellite remote sensing data, using GIS techniques, have been successfully used to monitor and assess dynamics in LULC and their impact on people and the environment (Aydöner and Maklav 2009). Satya et al. (2020) analyzed the effect of temporary variations of LULC concerning the rainfall runoff during 1995 to compare with 2005 employing Landsat TM satellite image. The runoff model and precipitation have been used as the input data, and high-resolution imagery data have been employed. They observed an increasing pattern of runoff by peak discharge due to LULC for the studied period.

Spatiotemporal data have been successfully employed to detect LULC and deterioration using a dataset from 1956 in a river catchment arid region of Spain (Adolfo and Rosalen 2007). Dramatic LULC variations have been also observed in a peri-urban to urban areas in the case study of a semiarid region of Mexico, where the research conducted for water transfer (Autónoma et al. 2011). Dimitrios et al. (2012) conducted a survey to improve LULC classification results. They employed the TM/ETM+ images of wet and dry periods, rainfall, and ground spectra radiometric data and estimated the consequences using statistical analysis techniques, validated through various classification algorithms. Mohaideen and Varija (2018) experimented with detecting the possible and appropriate variables for the estimation of the infiltration rate. They employed the hydrological modeling approach and ran various hydrological components under two LULC classification scenarios for 2000 and 2010 in the upper Bhima Catchment of India. Vogels et al. (2019) researched to monitor the agricultural activities adopted by stakeholders of Africa based on spatial and temporal configurations using object-based image analysis and Sentinel-2 imagery in Africa. The results revealed that the farming field irrigated artificially within the controlled conditions, while the precipitation influences are uncontrolled. Besides, LULC thematic maps developed every month have been produced from September 2016 to August 2017 at 10×10 m resolution on the farm level. Three collections of thresholds are employed to define the vegetative enhancement of crop germination separately on the artificial application of irrigation practices and rainfall patterns. The land irrigated ranges from 27.96 to 37.13 million hectares (MH), which 2.8 to 3.7 times exceeds the current higher assessments. Developed, thematic maps globally at 1000×1000 m resolution, and 1.2 to 1.7 exceeding from the map of Asia (2000–2010) and Africa (2010) by inducing the irrigated and rainwater harvesting area. The exposed mapping has vast information for the future development of Africa.

It is most important to estimate the LULC precisely to understand the actual situation and to plan for future

management (Mahmood et al. 2010). The RS and GIS are possibly used to detect the climate change impacts on the LULC, as Mercy (2015) elaborated similar results in the current study. She revealed significant losses in forestry and agriculture and documented the impact of the increase in barren land and mountain cover (Mercy 2015). Nayak and Mandal (2012) experimented using Landsat imagery data over 37 years (1973–2009) of temperature. Their results disclose that temperature imposes an adverse effect on LULC of forestry with a significant reduction in agriculture lands. The dynamics of LULC impose substantial impacts on water resources.

Most of the researchers, including El Bastawesy (2014), El Bastawesy et al. (2013), and Diwan and Varija (2018), have strongly recommended the application of remote sensing (RS) for the successful detection of LULC. Betru et al. (2019) and Milanova and Telnova (2007) described the application of RS and GIS might be improved with ground validation and physical surveys, which play an essential role in identifying the dynamics in LULC. The current study focused on the spate irrigation system, which is the indigenous practice of agricultural activities in the study area and LULC detections concerning precipitation patterns.

The spate irrigation system is a floodwater harvesting and management system. It is the oldest system of its kind—about 70 centuries earlier—and relied on about 13 million people in 20 countries (Mehari et al. 2011). The primary objective of the study is “Spatiotemporal variability in Spate Irrigation Systems in Khirthar National Range, Sindh, Pakistan (Case Study).”

The study provides recommendations to stakeholders, mainly to use the best water resources applications for agricultural and livelihood practices and adopt appropriate indigenous water conservation techniques to mitigate the drought impacts. Our research focuses on the following areas of concern:

1. Types of land use and land cover and their covered areas
2. Influence of precipitation on land use and land cover change
3. Effect of settlement on water bodies

Study area

According to Pakistan’s Bureau of Census (2017), the study area—Tehsil Thana Bula Khan—is inhabited by 155,000 people and lies between longitude 67° 16′ 16.49″ to 67° 56′ 30.98″ east to east and latitudes 24° 58′ 7.7″ to 26° 6′ 59.36″ north to north. The area located in the southwest mountainous range of Khirthar National Park along the main pathway of the China–Pakistan Economic Corridor (CPEC). The project stretches over an area of 5273 km² in the district Jamshoro of Sindh Province in Pakistan, extending to the

coastal belt and metropolitan city of Karachi (Fig. 1). The climate of the study area falls in the arid to semiarid region, which depends on the precipitation for its agriculture and groundwater recharge.

The annual average precipitation in the region ranges from 170 to 190 mm, with an evaporation rate varying from 2000 to 2100 mm. The study area is comprised of many watersheds, which are divided into several sub-basins and streams. However, a canal irrigation system does not exist in the study area due to its high elevation. A torrential rainfall causes floods, especially during the monsoon (July to October) of each year. The flooding supplies water to sub-basins and streams. Local farmers use the traditional practice of diverting floodwater through earthen or solid structures for harvesting using a torrential network (Mehari et al. 2007). In the study area, the spate irrigation system is locally known as “Nai” (Ephemeral River, transit streams). The people are also engaged in rearing livestock to earn a livelihood. The temperature varies between 6 °C to almost 48 °C during the winter and summer seasons. From the observation of meteorological data, the wind direction is mostly southwards, and velocity varies between 8 and 12 km/h.

Materials and methods

The data used in this research are comprised of Landsat 5 TM and Landsat 08 OLI satellite images (L05_TM, 2010 and LC08_OLI, 2013, 2014, 2015, 2016, 2017, and 2018), at a resolution of 30 × 30 m, only for September of each year. The study area depends on the precipitation as a primary water source, so the images acquired are for the wet season immediately after the monsoon for the detection of LULC variability. The spatial–temporal images covering the study area, “Path 152–153 and Row 42–43,” are downloaded from the USGS Earth Explorer website (www.earthexplorer.usgs.gov). We use Environmental Systems Research Institute, Inc. (Esri), ArcGIS software for the processing of satellite images. Satellite imagery for the respective temporal period is executed to produce false color composite (FCC) images using an image analysis Maximum Likelihood tool in ArcGIS software. Ground truth data are collected using GARMIN Global Positioning System (GPS), Model MAP64 s during the monsoon seasons for the year 2010, and between the years 2015 and 2018 for the general assessment of LULC classification consequences. A severe flooding event is evident in the clear sky Landsat 5 Thematic Mapper (TM) satellite image captured on September 4, 2010. However, the mean precipitation in the study area is depicted in Table 1.

Similarly, the data that have validated, we scrutinized the classified data of ArcGIS for each year and verified by overlaying it on Google temporal earth images for the

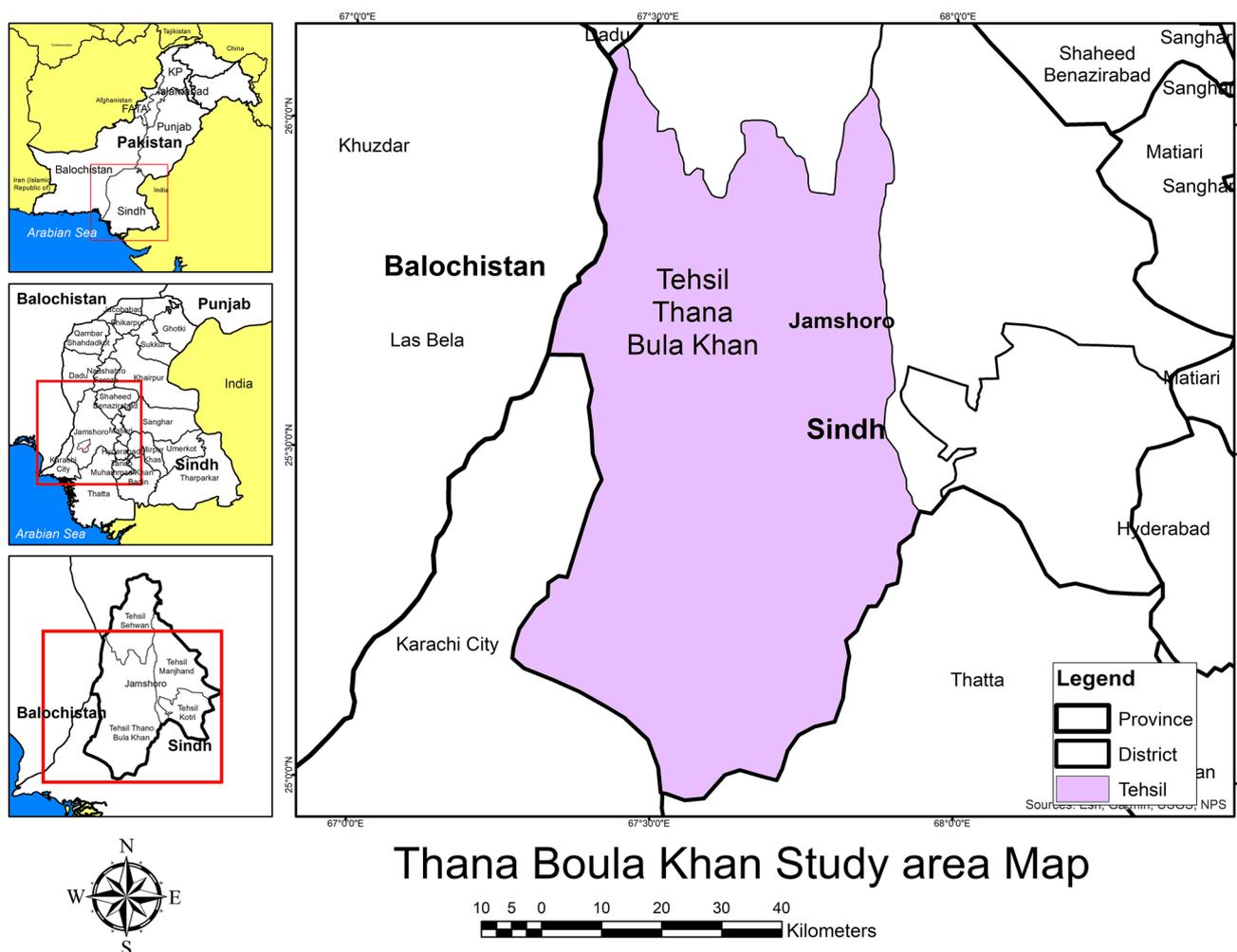


Fig. 1 Map of the study area

Table 1 Mean precipitation (mm) in the study area

Year	Precipitation (mm)	Remarks
2010	664.94	High flood
2013	111.59	Less precipitation
2014	115.02	Less precipitation
2015	130.09	Less precipitation
2016	155.85	Normal precipitation
2017	150.83	Normal precipitation
2018	167.16	Normal precipitation

same years. However, displaying the image in Google Earth Pro, we used a high-resolution temporal image for each year to extract the classified values of the random points, and we generated a ground truth populated table through ArcGIS. Finally, we classified images of the study area obtained using image change detection and classification tools. We reexamined the reliability of the spatial-temporal

dataset reexamined the confusion matrix tool to generate a table to assess the accuracy of the satellite-derived LULC classification.

Data analysis

For 2010 and between the years 2013 and 2018, totaling 7 years, we obtained change detection areas in percentages, whereas, and we retrieved feature classes for the studied area in square kilometers (km²). We used the maximum likelihood classification (MLC) technique, which uses the spectral signatures obtained from the training sample to classify images. According to Rawat and Kumar (2015), MLC is a standard supervised classification algorithm used to classify satellite images into thematic maps. To acquire thematic maps of the study area, from Landsat images, we applied the MLC tool available in the Spatial Analyst extension toolbar of ArcGIS 10.3.1 software, resulting in the segmentation of the study area into five classes. We used the maximum

likelihood classifier to calculate each category based on its highest probability of the cell belonging to that class, given its attribute values. The five types of LULC in the investigated region are barren land, settlements, vegetation, water bodies, and spate irrigation systems. Spate irrigation systems are indigenous (rainwater harvesting) practices that divert runoff to agricultural drylands. The vegetation class represents crop cultivation, while the water body class represents reservoirs and ponds. We used thematic change detection to analyze land use and land cover change, deforestation, urbanization, agricultural expansion, and water variability, among others. To assess the change detection in percentages for the study period of 2010 and from 2013 to 2018, we cross-tabulated comparative results from the regions obtained from the classified image pairs. The statistical approach is built in the ArcGIS to automate the spatial analysis to LULC as conducted experiments by Bonato et al. (2019), Olaoye et al. (2019) and Haack and Mahabir (2019).

Results and discussion

After processing and analyzing the Landsat images for the years 2010 and 2013–2018, we produced detailed maps of LULCC, as shown in Fig. 2 for the study area. We recognized five land use categories, delineating them as observed in Fig. 2 using a maximum likelihood classification technique in ArcGIS 10.3.1 software. The five types are barren land, settlements, vegetation, water bodies, and spate irrigation systems. The detailed analysis and interpretation of land use formation, as observed in Table 2, break land use formation into km^2 . Tables 3, 4 and 5 show LULCC during low- and average-precipitation events compared to a high-rainfall event. The analysis showed the land cover classification of the study area for the last wet year of 2010, observed as the massive flood year of the last 81 years. In 2010, heavy rainfall resulted in flooding, causing inundation of low-lying areas, sufficiently recharging the groundwater, filling surface water bodies for crop irrigation.

Furthermore, Table 3 represents changes in land cover classes for the years 2013–2015 due to below-average-precipitation events, which caused areas covered through surface water bodies, vegetation, and lands under spate

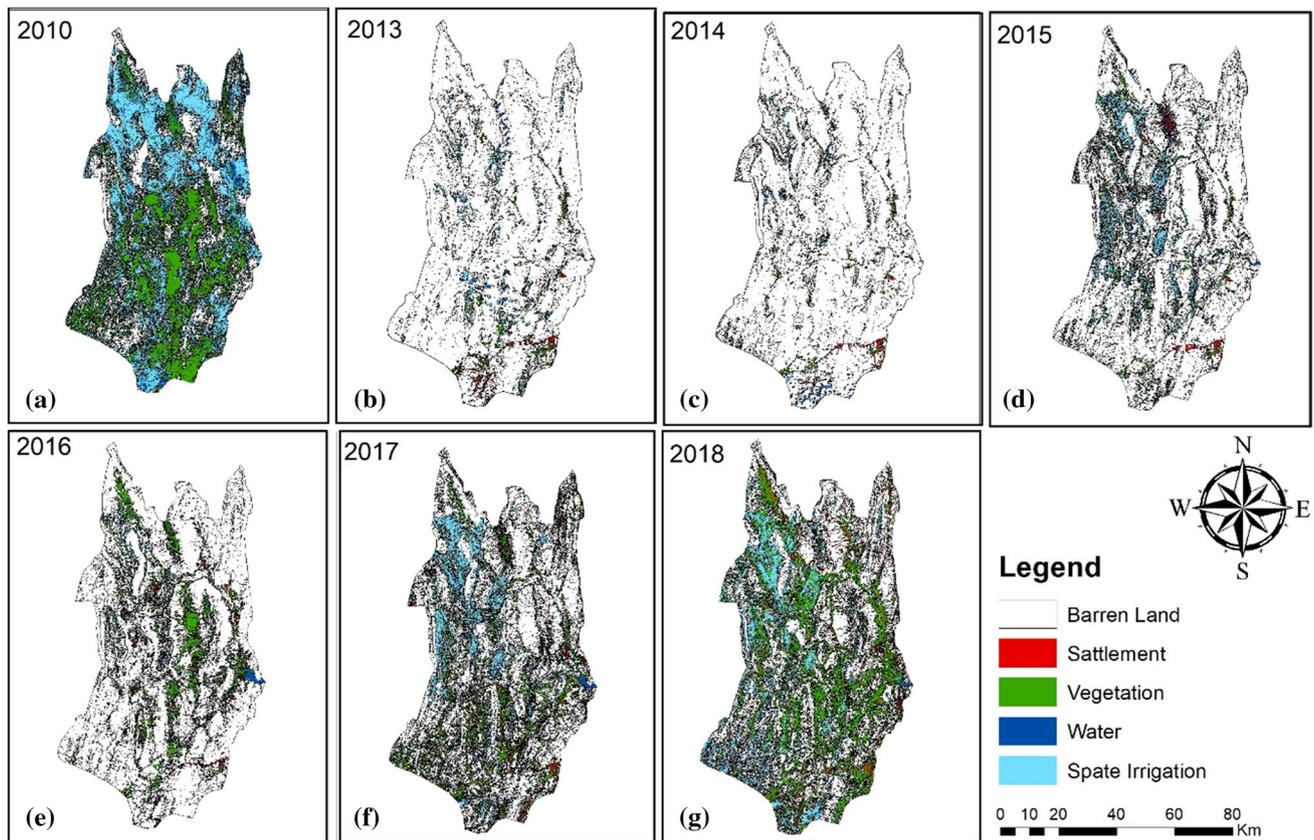


Fig. 2 LULC classification, 2010 and 2013–2018

Table 2 Overall land use LULCC during the study period

Type of land use/cover	Year							Observed change 2010–2018
	2010	2013	2014	2015	2016	2017	2018	
	(km ²)							
Barren land	2957.20	5041.85	5079.48	4782.59	4832.66	4427.36	3875.43	918
Settlement	16.17	34.68	39.00	49.25	53.90	111.92	117.80	102
Vegetation	867.38	76.03	58.88	110.47	283.63	314.80	807.41	−60
Water bodies	180.15	19.87	29.56	27.72	28.52	30.47	1.23	−169
Spate irrigation system	1252.35	100.81	66.32	303.20	74.54	388.70	461.37	−791

Table 3 LULC during periods of lower-than-average precipitation years

Type of class (%)	Years of less rain			Average	Years of high rain 2010	Difference	Remarks
	2013	2014	2015				
Barren land	95.61	96.33	90.70	94.21	56.08	38.13	Positive
Settlements	0.66	0.74	0.93	0.78	0.31	0.47	Positive
Vegetation	1.44	1.12	2.10	1.55	16.45	−15.00	Negative
Water bodies	0.38	0.56	0.53	0.50	3.42	−2.92	Negative
Spate irrigation system	1.91	1.26	5.76	2.98	23.75	−20.77	Negative

Table 4 LULCC during periods of normal and high precipitation

Type of class (%)	Years of normal to high precipitation			Average	Years of high rain 2010	Difference	Remarks
	2016	2017	2018				
Barren land	91.64	83.96	73.49	83.03	56.08	26.95	Positive
Settlements	1.02	2.12	2.23	1.79	0.31	1.48	Positive
Vegetation	5.38	5.97	15.31	8.89	16.45	−7.56	Negative
Water bodies	0.54	0.58	0.21	0.44	3.42	−2.98	Negative
Spate irrigation system	1.41	7.37	8.75	5.84	23.75	−17.91	Negative

Table 5 Dynamics of LULC during less and normal precipitation events 2013–2018

Type of land use/cover	Year							Observed change 2013–2018	Remarks
	2013	2014	2015	2016	2017	2018			
	(%)								
Barren land	95.61	96.33	90.70	91.64	83.96	73.63	−21.98	Negative	
Settlement	0.66	0.74	0.93	1.02	2.12	2.24	1.58	Positive	
Vegetation	1.44	1.12	2.09	5.38	5.97	15.34	13.90	Positive	
Water bodies	0.38	0.56	0.53	0.54	0.58	0.02	−0.35	Negative	
Spate irrigation system	1.91	1.26	5.75	1.41	7.37	8.77	6.85	Positive	

irrigation systems to shrink. On the contrary, barren lands increased due to less surface and subsurface water availability for agricultural practices. Table 3 shows precipitation pattern between the years 2016 and 2018; water scarcity reversed again due to average rainfall replenishing sources of surface and subsurface water bodies, increasing

lands under spate irrigation systems, and the ground covered under vegetation.

The detailed results of our study area presented in Tables 2, 3, and 4 elaborated that five land cover classes for 2010 explained the barren land covers 56.08%, i.e., an area of 2957 km², and settlements cover 0.31%, i.e.,

an area of 16 km². Vegetation covers 16.45%, equivalent to 867 km², water bodies cover 3.42%, equivalent to 180 km², and spate-irrigated lands cover 23.75%, amounting to 1250 km². Table 3 shows percentagewise land cover classes from 2013 to 2018. However, when we compare 2018 to 2010, we observe a remarkable decreasing trend overall in LULC changes in terms of water bodies (−2.98% or 169 km²), spate-irrigated lands (−17.91% or 791 km²), and vegetation cover (−7.56% or 60 km²). Conversely, we observe a substantial increasing in barren land (26.95% or 918 km²) and settlements (1.48% or 102 km²).

Tables 2, 3 and 4 indicate the positive and negative trends in certain land cover types by comparing years of high, less, and normal event precipitation rates. That showed a variable trend of change in barren land, with increases from 56.08% (2957.20 km²) in 2010 to 95.61% (5041.85 km²) in 2013 and 96.33% (5079.48 km²) in 2014, decreasing to 90.70% (4782.59 km²) in 2015 due to less precipitation. Barren land decreased from 91.64% (4832.66 km²) in 2016 to 83.96% (4427.36 km²) in 2017 to 73.49% (3875.43 km²) in 2018 due to normal precipitation. Our analysis of barren land cover change highlights the dependence of vegetation vigor and agricultural practices on the rate of precipitation received in the study area.

Likewise, a straight positive trend is shown in settlements during the study period, i.e., 0.31% (16.17 km²) in 2010, 0.66% (34.68 km²) in 2013, 0.74% (39.0 km²) in 2014, 0.93% (49.25 km²) in 2015, 1.02% (53.90 km²) in 2016, 2.12% (111.92 km²) in 2017, and 2.23% (117.80 km²) in 2018. The results reveal that the population (settlement) almost doubled in a single year (between 2016 and 2017), which is indicative of better means of livelihood in the study area.

In line with settlement land cover change, Tables 2, 3 and 4 depict dynamics in vegetative land cover during the study period. The highest vegetative land cover (16.45% or 867.38 km²) was observed due to historic heavy precipitation in 2010, which dropped to 1.44% (76.03 km²) in 2013 and to 1.12% (58.88 km²) in 2014 due to low precipitation and then steadily recovered from 2.10% (110.47 km²), 5.38% (283.63 km²), 5.97% (314.80 km²), and 15.31% (807.41 km²) in 2015, 2016, 2017, and 2018. Vegetation land cover jumped above 50%, i.e., 283.63 km² in 2016 to 314.80 km² in 2017. Nevertheless, vegetation land cover in 2018 was 807.41 km², which indicates the next wet year after 2010.

The analysis highlights the change in water bodies' land cover. Water bodies land cover was 3.42% (180.15 km²) in 2010, which suddenly dropped to 0.38% (19.87 km²) in 2013 and then slightly varied between 0.56% (29.56 km²), 0.53% (27.72 km²), 0.54% (28.52 km²), and 0.58% (30.47 km²) in 2014–2017 before dropping further to

0.21% (1.23 km²) in 2018. The decrease in water bodies is virtually 89% between 2010 and 2013.

Also, the tables show that cultivation practices under spate irrigation systems covered 23.75% (1252.35 km²) in 2010, withering to 1.91% (100.81 km²) in 2013—a decrease of 92%, equivalent to 1152 km²—and further dropping to 1.26% (66.32 km²) in 2014 before increasing to 8.75% (461.37 km²) in 2018.

Table 5 reflects the LULC scenario of the last 6 years (2013–2018) in the study area and elaborates the fact and figures of each class during less and normal precipitation events. The barren land occupied 95.61%, 96.33%, 90.70%, 91.64%, 83.96%, and 73.63 followed by the year 2013, 2014, 2015, 2016, 2017, and 2018. The increased values showed the scenario during the fewer precipitation events and the highest trend of barren land recorded during the normal precipitation events—a negative trend of barren land of −21.98% shown during 2018 in the above comparison.

Settlement in the study area has gradually increased: 0.66%, 0.74%, 0.93%, 1.02%, 2.12% and 2.24% in 2013, 2014, 2015, 2016, 2017, and 2018, respectively, shown in the figures. A positive trend of +1.58% is shown during the year of 2018 shown in Table 4. The increasing trend might be the result of the recent development in the study area because the study area is situated at the main super highway redeveloped and promoted as a smart motorway; this connected with the mega metropolitan city of Karachi is a major traveling source of China–Pakistan Economic Corridor.

The vegetation class showed a positive trend from 2013 to 2018, except for a slight decrease of 0.32% between the years 2013–2014. The vegetation in the study area revealed the trend of 1.44%, 1.12%, 2.09%, 5.38%, 5.97, and 15.37% in the year 2013, 2014, 2015, 2016, 2017, and 2018, respectively. An increase in the vegetation of +13.90% is shown in the year of 2018. The vegetation trend of the study area represents the indigenous practice of spate irrigation systems in the study area.

The classification of water bodies shows a fluctuated trend in the study area. A decreasing trend of 0.38% and 0.02% is shown during the year of 2013 and 2018. However, an increased trend has been revealed as 0.56%, 0.53%, 0.54%, and 0.58% in the year of 2014, 2015, 2016, and 2017, respectively. Actually, the LULC has developed through the Landsat images of September for each year. So it might be possible that the reduced trend happened due to percolation and contribution in the groundwater and utilized for spate irrigation system due to earlier events of precipitation in the month of July and August during the year of 2018. That is why the increasing trend of vegetation shows 13.90% in the study area. The overall scenario showed a negative value of −0.35% in the water bodies trend.

The classification of the spate irrigation system in the study area shows a positive trend of 6.85% as compared to the year

Table 6 Overall accuracy and Kappa coefficient of LULC classification

Accuracy assessment	2010	2013	2014	2015	2016	2017	2018
Overall accuracy (%)	82.80	87.20	92.00	85.20	86.40	88.80	91.60
Kappa coefficient	0.80	0.84	0.90	0.82	0.84	0.86	0.90

of 2013 and 2018. The trend has shown minimum 1.91%, 1.26%, and 1.41% practice of the spate irrigation system carried out during the year of 2013, 2014, and 2016 due to less precipitation in the year of 2013 and 2014, while there are mismanagement practices during the year of 2016. However, the trend of spate irrigation of 5.75%, 7.37%, and 8.77% during the year of 2015, 2017, and 2018 showed an increase in the spate irrigation system.

The confusion matrix shows 82.8%–92% overall accuracy, while the Kappa coefficient was 0.80–0.90 with acceptable results of LULC classifications, as detailed in Table 6.

In the current study, the significant influence of precipitation observed in LULC changes. As the discussions in the first section of this paper indicate, anthropogenic activities change the environment for centuries such as gradual increases in population, various livelihood activities, migration rates, and the atmospheric temperature impacts on the environment. Environmental changes have seen at the local, regional, national, and international levels in atmospheric temperature, precipitation, general LULC, and climate change impacts. Various researchers have elaborated on the significant implications (e.g., Chase et al. 2000; Kalnay and Cai 2003; Cai and Kalnay 2004; Trenberth 2004; Vose et al. 2004; Feddema et al. 2005; Christy et al. 2006; Mahmood et al. 2006; Ezber et al. 2007; Nuñez et al. 2008). The arid and semiarid regions are depending on precipitation as a primary source of water resources; climate change is the essential variable for such vulnerable areas to maintain the ecosystem. These regions are critical to monitoring the LULC and climate change impact on water resources. Yin et al. (2017) researched this issue in the Jinghe River Basin of about more than 45,000 km², located in the semiarid region of northwest China. Their study emphasized the impact of LULC and climate change on surface water resources. The investigation found multiple effects of LULC and climate change: runoff increased with the higher trend of the precipitation from 1970 to 1980, while the runoff reduced due to the increasing impact of LULC dynamics from 1980 to 1990, which contributed to 44% of runoff changes.

Locally, detailed analyses of spate irrigation systems reveal a scarcity of surface water bodies, which in turn severely affects the reserves of groundwater. Moreover, deficiency raises an alarm of drought conditions due to less occurrence of the precipitation in the study area. Djuma et al. (2017) highlighted the same issue in the context of the Peristerona Watershed—a semiarid region of Cyprus. Their study aimed to estimate the groundwater recharge by focusing on the assessment of recharge in spates (natural streams or ephemeral rivers) with

and without check dams. The current study disclosed that the hydraulic structures contributing to the potential water conservation practices in the study area, which mitigate the drought impacts by capturing the runoff.

The discussions revealed that the physiographic conditions of the study area are more appropriate to develop potential storage sites. Evapotranspiration has much importance to manage water resources in the study area and focused the return period for 10 to 1 years of rainfall, and the selection of cropping patterns would influence accordingly. The options of change in crops rendering the evapotranspiration lead to reduction in the climate change impact for sustainable development.

Conclusion

In this study, we estimate the dynamics of spate irrigation (indigenous rainwater harvesting) systems by focusing on five categories of LULC—barren land, settlements, spate irrigation, vegetation, and water bodies—and its drivers through the study of the Khirthar National Range, Sindh, Pakistan. The results show a strong relationship between precipitation and LULC in 2010 and from 2013 to 2018. During wet years, we observe an increase in land cover of water bodies and vegetation and a reduction in the barren land. Other direct observations reveal the enormous expansion of 86.25% in settlement land cover from 16.17 km² in 2010 to 117.80 km² in 2018, which almost doubled in a single year from 53.90 km² in 2016 to 111.92 km² in 2017. The overall accuracy and the Kappa coefficient show satisfactory results of the LULC classification. The spatiotemporal variability in spate irrigation systems showed discrepancy and dependence on precipitation patterns. The indigenous practice of harvesting through spate irrigation systems is under direct threat due to precipitation shortfalls and prevailing dry periods in the study area. Investigating the potential harvesting sites and arrangement of appropriate water conservation practices may reduce drought impacts from the region. LULC is a popular technique for sustainable development through future planning and management.

Acknowledgements The current research was conducted using the GIS and Computer Labs of the US-Pakistan Center for Advanced Studies in Water (USPCAS-W)/Mehran University of Engineering and Technology, Jamshoro. We also acknowledged the efforts, dedications, and knowledge of the local people who are engaged in the existing spate irrigation system; their dedicated efforts and expertise played a vital role in developing the current research. With great pleasure, the authors acknowledge the support provided by the FBLN/MetaMeta under “Africa to Asia and Back: Testing Adaptation in Flood Based Farming Systems” and the cooperation and supervision of the faculty

of the USPCAS-W. Last but not least, many thanks go to USAID for providing technical support for the establishment of the Center.

Funding No funding was received.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

References

- Adolfo C, Rosalen EA (2007) Land use change and land degradation in Southeastern Mediterranean Spain Elias Symeonakis *Æ*. *Environ Manag* 40:80–94. <https://doi.org/10.1007/s00267-004-0059-0>
- Autónoma CJ, Díaz-Caravantes RE, Sánchez-Flores E (2011) Water transfer effects on peri-urban land use/land cover: a case study in a semi-arid region of Mexico. *Appl Geogr* 31:413–425
- Aydöner C, Maktav D (2009) The role of the integration of remote sensing and GIS in land use/land cover analysis after an earthquake. *Int J Remote Sens*. <https://doi.org/10.1080/01431160802642289>
- Betru T, Motuma T, Kefyalew S, Habtemariam K (2019) Trends and drivers of land use/land cover change in Western Ethiopia. *Appl Geogr* 104:83–93
- Birhanu A (2018) Impacts of land use and land cover changes on hydrology of the Gumara catchment, Ethiopia. *Phys Chem Earth*. <https://doi.org/10.1016/j.pce.2019.01.006>
- Bonato M, Cian F, Giupponi C (2019) Combining LULC data and agricultural statistics for a better identification and mapping of High nature value farmland: a case study in the veneto Plain, Italy. *Land Use Policy* 83:488–504
- Butt A, Shabbir R, Saeed Ahmad S, Aziz N (2015) Land use change mapping and analysis using remote sensing and GIS: a case study of Simly Watershed, Islamabad, Pakistan. *Egypt J Remote Sens Space Sci*. <https://doi.org/10.1016/j.ejrs.2015.07.003>
- Cai M, Kalnay E (2004) Response to the comments by Vose et al. and Trenberth. *Impact of land-use change on climate*. *Nature* 427:214
- Chase TN, Pielke RA, Kittel TGF, Nemani RR, Running SW (2000) Simulated impacts of historical land cover changes on global climate in northern winter. *Climate Dyn* 16:93–105
- Chignell SM, Anderson RS, Evangelista PH, Laituri MJ, Merritt DM (2015) Multi-temporal independent component analysis and Landsat8 for delineating maximum extent of the 2013 Colorado front range flood. *Remote Sens* 2015(7):9822–9843
- Chowdhury M, Hasan ME, Abdullah-Al-Mamun MM (2018) Land use/land cover change assessment of Halda watershed using remote sensing and GIS. *Egypt J Remote Sens Space Sci*. <https://doi.org/10.1016/j.ejrs.2018.11.003>
- Christy JR, Norris WB, Redmond K, Gallo KP (2006) Methodology and results of calculating central California surface temperature trends: evidence of human-induced climate change? *J Climate* 19:548–563
- Dimitrios DA, Athos A, Diofantos GH, Adrianos R (2012) Optimizing statistical classification accuracy of satellite remotely sensed imagery for supporting fast flood hydrological analysis. *Acta Geophys* 60(3):959–984. <https://doi.org/10.2478/s11600-012-0025-9>
- Djuma H, Bruggeman A, Camera C, Eliades M, Kostarelos K (2017) The impact of a check dam on groundwater recharge and sedimentation in an ephemeral stream. *Water* 9(10):813. <https://doi.org/10.3390/w9100813>
- El Bastawesy M (2014) Hydrological scenarios of the renaissance Dam in Thiopia and its hydro-environmental impact on the Nile downstream. *J Hydrol Eng*. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001112](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001112)
- El Bastawesy M, Ramadan Ali R, Faid A, El Osta M (2013) Assessment of water logging in agricultural megaprojects in the closed drainage basins of the Western Desert of Egypt. *J Hydrol Earth Syst Sci* 17:1493–1501. <https://doi.org/10.5194/hess-17-1493-2013>
- Ezber Y, Sen OL, Kindap T, Karaca M (2007) Climatic effects of urbanization in Istanbul: a statistical and modeling analysis. *Int J Climatol* 27:667–679
- Feddema JJ, Oleson KW, Bonan GB, Mearns LO, Buja LE, Meehl GA, Washington WM (2005) The importance of land-cover change in simulating future climates. *Science* 310:1674–1678
- Fiorella OW, Julián BB (2018) Assessment of satellite-based precipitation estimates over Paraguay. *Acta Geophys* 2018(66):369–379. <https://doi.org/10.1007/s11600-018-0146-x>
- Girma H, Hassan R (2014) Drivers of land-use change in the southern nations, nationalities and people's region of Ethiopia. *Afr J Agric Resour Econ* 9(2):148–164
- Haack B, Mahabir R (2019) Optical and radar data analysis for land use land cover mapping in Peru. *Remote Sens Land* 3(1):15–27
- International Fund of Agricultural Development (IFAD) (2010) Spate Irrigation, Livelihood, Improvements, and Adaptation to Climate Change. MetaMeta & IFAD (2010), p 4
- Jain Figueroa A (2019) Sustainable agricultural management: a systems approach for examining food security tradeoffs (Doctoral dissertation, Massachusetts Institute of Technology)
- Kalnay E, Cai M (2003) Impact of urbanization and land use on climate change. *Nature* 423:528–531
- Mahmood R, Foster SA, Keeling T, Hubbard KG, Carlson C, Leeper R (2006) Impacts of irrigation on 20th century temperatures in the Northern Great Plains. *Glob Planet Change* 54:1–18
- Mahmood R, Pielke RA Sr, Hubbard KG, Niyogi D, Bonan G, Lawrence P et al (2010) Impacts of land use/land cover change on climate and future research priorities. *Bull Am Meteorol Soc* 91(1):37–46
- Markhi A, Laftouhi N, Grusson Y, Soulaïmani A (2019) Assessment of potential soil erosion and sediment yield in the semi-arid N'fs basin (High Atlas, Morocco) using the SWAT model. *Acta Geophys* 2019(67):263–272. <https://doi.org/10.1007/s11600-019-00251-z>
- Mehari A, van Steenberg F, Schultz B (2007) Water rights and rules, and management in Spate irrigation systems in Eritrea, Yemen, and Pakistan. *Community-based water law and water resource management reform in developing countries*, p 114
- Mehari A, Van Steenberg F, Schultz B (2011) Modernization of Spate irrigated agriculture: a new approach. *Irrig Drain* 60(2):163–173
- Mercy MW (2015) Assessment of the effects of climate change on land use and land cover using remote sensing: a case study from Kenya. Working paper series. Dresden nexus conference. Dnc2015/03
- Milanova E, Telnova N (2007) Land-use and land-cover change study in the transboundary zone of Russia-Norway. *Man in the landscape across frontiers: Landscape and land use change in central European border regions*. In: CD-ROM conference proceedings of the IGU/LUCC central Europe conference, pp 123–133
- Mohaideen MMD, Varija K (2018) Improved vegetation parameterization for hydrological model and assessment of land cover change impacts on the flow regime of the Upper Bhima basin, India. *Acta Geophys* 2018(66):697–715. <https://doi.org/10.1007/s11600-018-0161-y>
- Nayak S, Mandal M (2012) Impact of land-use and land-cover changes on temperature trends over Western India. *Curr Sci* 102(8):1166–1173
- Núñez MN, Ciapessoni HH, Rolla A, Kalnay E, Cai M (2008) Impact of land use and precipitation changes on surface temperature

- trends in Argentina. *J Geophys Res* 113:D06111. <https://doi.org/10.1029/2007JD008638>
- Olaoye IA, Ortiz J, Jefferson A, Shakoor A (2019) Landuse/landcover (LULC) change modeling of Old Woman Creek (OWC) Watershed using Remote Sensing and GIS. *Environmental Science & Design Research Initiative*. Paper 35. <https://digitalcommons.kent.edu/esdri2019/35>
- Pakistan's Bureau of Census (2017). http://www.pbs.gov.pk/sites/default/files/PAKISTAN%20TEHSIL%20WISE%20FOR%20WEB%20CENSUS_2017.pdf
- Pan Y, Gong H, Zhou D, Li X, Nakagoshi N (2011) Impact of land use change on groundwater recharge in Guishui River Basin, China. *Chin Geogr Sci* 21(6):734–743
- Rawat JS, Kumar M (2015) Monitoring land use/cover change using remote sensing and GIS techniques: a case study of Hawalbagh block, district Almora, Uttarkhand, India
- Rokni K, Ahmad A, Selamat A, Hazini S (2014) Water feature extraction and change detection using multitemporal landsat imagery. *Remote Sens* 2014(6):4173–4189
- Satya BA, Shashi M, Pratap D (2020) Effect of temporal-based land use–land cover change pattern on rainfall runoff. In: *Applications of geomatics in civil engineering*. Springer, Singapore, pp 175–182
- Tilman D, Balzer C, Hill J, Befort BL (2011) Global food demand and the sustainable intensification of agriculture. *Proc Natl Acad Sci* 108(50):20260–20264
- Trenberth KE (2004) Rural land-use change and climate. *Nature* 427:213
- Vogels MF, De Jong SM, Sterk G, Douma H, Addink EA (2019) Spatio-temporal patterns of smallholder irrigated agriculture in the horn of Africa using GEOBIA and Sentinel-2 imagery. *Remote Sens* 11(2):143
- Vose RS, Karl TR, Easterling DR, Williams CN, Menne MJ (2004) Impact of land-use change on climate. *Nature* 427:213–214
- Yin J, He F, Xiong YJ, Qiu GY (2017) Effects of land use/land cover and climate changes on surface runoff in a semi-humid and semi-arid transition zone in northwest China. *Hydrol Earth Syst Sci* 21(1):183–196. <https://doi.org/10.5194/hess-21-183-2017>



Influence of collars on reduction in scour depth at two piers in a tandem configuration

Sargol Memar^{1,2} · Mohammad Zounemat-Kermani¹ · Aliasghar Beheshti³ · Majid Rahimpour¹ · Giovanni De Cesare² · Anton J. Schleiss²

Received: 2 September 2019 / Accepted: 6 December 2019 / Published online: 11 December 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

Bridge failure, due to local scour at bridge pier foundations, has become a critical issue in river and bridge engineering, which might lead to transportation disruption, loss of lives and economic problems. A practical solution to prevent bridge collapses is the implementation of scour mitigation methods around bridge foundations. Based on an experimental perspective, this study is focused on the influence of the size and position of circular collars from the sediment bed on scour depth at two tandem piers. To meet this end, long-lasting experiments are performed under clear-water conditions using uniform sand for bed materials. Compared to the adjacent position of the collar on the bed, placing the collars below the bed would increase the delay time of scour at the piers up to four times. However, regardless of the delay time, the observations indicate that locating the collars on the initial bed surface results in maximum reduction in scour depths around the piers. It was found that diminishing the flow intensity has a dramatic impact on the scour reduction at the piers, so that maximum reduction in scour depths at piers increased on average from 20 to 70% with the reduction in the flow intensity from 0.95 to 0.9.

Keywords Bridge foundation · Collar · Countermeasure · Horseshoe vortex · Scour depth reduction

List of symbols

B	Width of the channel
D	Pier diameter
D^*	Equivalent pier width
D_{proj}	Sum of the non-overlapping projected width of the piers onto a plane normal to the flow direction
d_s	Depth of scour
d_{se}	Equilibrium depth of scour
$d_{\text{s(ext)}}$	Extrapolated depth of scour to an infinite time
d_{50}	Median sediment grain size
g	Acceleration of gravity
h	Upstream flow depth
H_c	Collar height from the sediment bed

K_m	Number of aligned rows factor
K_{sp}	Factor for the distance between the piers
R	Reduction in scour depth
R_u	Reduction in scour depth at the upstream pier
R_d	Reduction in scour depth at the downstream pier
$R_p = UD/\vartheta$	Pier Reynolds number
s	Center-to-center spacing of the piers
U	Mean upstream flow velocity
U_c	Mean threshold velocity
u_c^*	Shear critical velocity
T	Dimensionless time of scour
t	Time of scour
t_c	Thickness of the collar
t_d	Delay time of scour
t_e	Equilibrium time of scour
w_c	Width of the collar
Z	Dimensionless scour depth
$\Delta = \rho'_s/\rho$	Relative submerged sediment density
ϑ	Fluid kinematic viscosity
ρ	Density of fluid
ρ_s	Sediment grain density
ρ'_s	Submerged sediment density

✉ Mohammad Zounemat-Kermani
zounemat@uk.ac.ir

¹ Department of Water Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

² Laboratory of Hydraulic Constructions, Ecole Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland

³ Department of Water Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

σ_g	Geometric standard deviation of the sediment grain size distribution
φ	Function

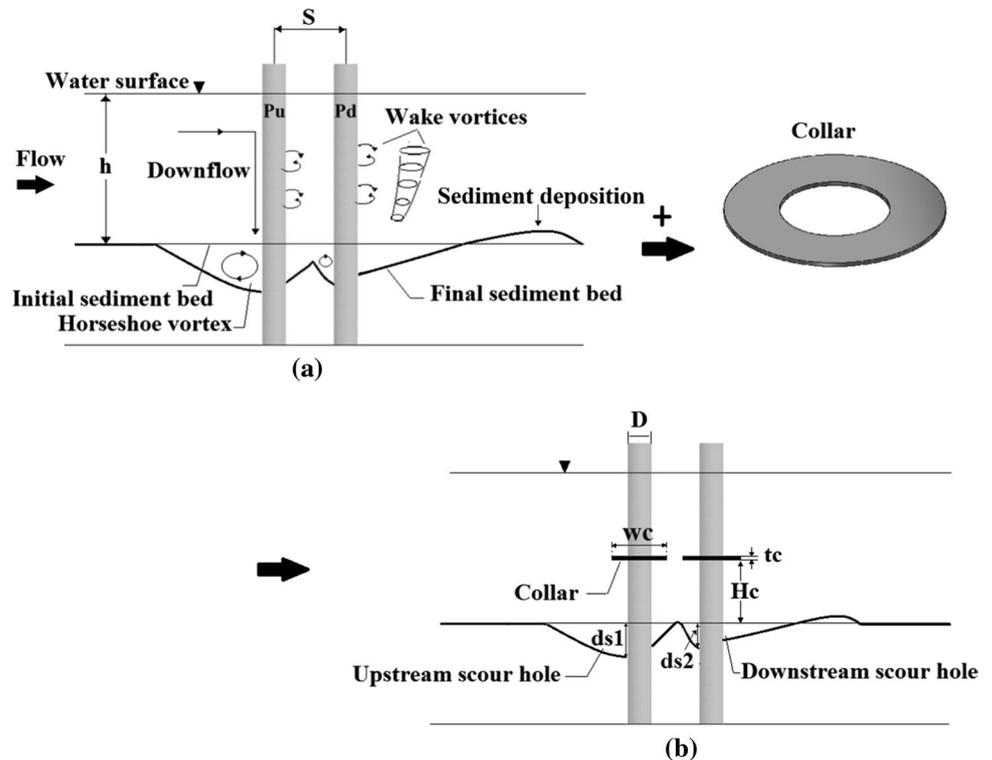
Introduction

Local scour at bridge foundations (piers and abutments) can be a significant hazard causing instability or collapse of such vital infrastructures. Pier scour is described as an erosion that undermines the bridge foundations. The erosive effect of flowing water forms bridge scour by the excavation and removal of sediment grains from around the abutments and piers of the bridges. Prediction of the depth of scour around bridge foundations is a fundamental part of designing safe bridges. Scour monitoring at bridges is performed regularly to maintain bridge safety against scour-induced failure (Ettema et al. 2006). In order to better understand the scour at the piers, it is of paramount importance to understand the flow pattern and performance of the large-scale coherent structures that are formed around the piers. In Fig. 1a, pier scour is depicted around two tandem piers. The approach flow velocity accelerates at the upstream of the piers and the flow is directed toward the sediment bed on the front face of the piers, which is called downflow. Where the downflow reaches the sediment bed, the horseshoe vortices are formed. Generally, scour begins at the flanks of the pier. Once the scour area extends as a hole around the pier, the

downflow and horseshoe vortices are strengthened, resulting in the development of the scour hole around the pier. By separating the approach flow from the sides of the piers, vortex shedding is generated in the piers' wake (Raudkivi 1986; Arneson et al. 2012 (HEC-18); Guo et al. 2012). In the case of two tandem piers, the existence of the upstream pier causes a reduction in the amount of the approach flow velocity toward the downstream pier. Thus, the rate of scour around the downstream pier is less than that around the upstream pier. When the scour holes around the piers overlap, the movement of sediment particles from upstream is facilitated, leading to increasing the depth of scour at the upstream pier (Hannah 1978; Lança et al. 2013). The variables P_u and P_d indicate the upstream and downstream piers, respectively.

Scour at bridge piers may occur under clear-water or live-bed scour conditions. According to Melville (1984) and Chiew (1984), the scour depth achieves two peaks depending on flow intensity. The first peak is called clear-water threshold peak and the second peak is termed as the live-bed peak. Under clear-water conditions and with uniform sediments, the scour depth increases approximately linearly with flow intensity and achieves its maximum value at the onset of sediment movement (at threshold velocity). The maximum scour depth is called threshold peak. Once the mean upstream flow velocity exceeds the threshold velocity, the scour depth reduces in the relatively small conversion of flow intensity and, then, increases again reaching the

Fig. 1 Schema of the flow field and scour holes around two tandem piers: **a** without collars; **b** with collars above the sediment bed, (d_s scour depth, h upstream flow depth, H_c collar height from the sediment bed, w_c width of the collar, t_c thickness of the collar, s center-to-center spacing of the piers, D diameter of the pier)



live-bed peak, leading to the formation of bedforms. Bedforms transport into the scour hole periodically, resulting in the periodic variation of maximum scour depth. In clear-water scour conditions, no upcoming sediment grain enters the scour hole (Shen et al. 1966; Ettema 1980; Melville and Chiew 1999; Sheppard and Miller (2006)).

Several methods have been examined for scour depth estimation at bridge piers by many investigators (Dey 1997a, b; Oliveto and Hager 2002, 2005; Zounemat-Kermani et al. 2009; Arneson et al. 2012; Sheppard et al. 2013; Pang et al. 2016). Many efforts have been made by researchers to reduce the dimensions of scour holes around the piers using several scour countermeasures. The choice of a proper pier countermeasure for a particular bridge is based on the flow conditions and features of the nearby channel (Johnson et al. 2002). There are two different types of scour countermeasures. Countermeasures such as bed sill, submerged vane and collar deflect the approach flow and minimize the power of the system of vortices around the pier, resulting in the reduction in scour depth. Other scour countermeasures are placed on the sediment bed to protect it from erosive and hydraulic forces consisting of riprap stones, cable-tied blocks and geobags (Tafarojnoruz et al. 2012; Khaple et al. 2017a).

Collars are thin horizontal disks that are attached around a pier (see Fig. 1). The collars protect the sediment bed from erosion caused by the downflow and horseshoe vortices. Thus, they reduce the scour potential around the pier. The effect of collars on scour depth at isolated bridge piers has been studied by several investigators (Kumar et al. 1999; Zarrati et al. 2004; Moncada-M et al. 2009; Masjedi et al. 2010; Tafarojnoruz et al. 2012;

Zokaei et al. 2013; Khodashenas et al. 2018; Karimaei Tabarestani and Zarrati, 2019). However, limited research has reported the influence of collars on scour reduction around the pier groups. Zarrati et al. (2006) examined the effect of the collars placed on the sediment bed on scour depths at two side-by-side and tandem piers. Heidarpour et al. (2010) studied the influence of collars on scour depth reduction at two and three tandem piers when the collars were set on the original bed. Figure 1b schematically illustrates scour holes around two tandem piers equipped with collars. Parameters d_{s1} and d_{s2} refer to scour depth at the upstream and downstream piers, respectively. Table 1 presents a summary of the essential experimental studies on the effect of collars on scour at the piers along with the obtained percentage reduction at scour depth in each study (the last column).

The effect of collars on scour depth at bridge piers is assessed based on the width of collar w_c and the collar height from sediment bed H_c . According to Table 1, the collar's effectiveness in reducing the scour depth around the pier increases with increasing the width of the collar. However, a collar that is wider than three times of the pier diameter is not applicable in practice (Zarrati et al. 2006). In addition, the optimal height of collars that results in maximum scour depth reduction around the piers depends on different factors of experimental condition, pier skewness and pier shape.

Furthermore, in most of the previous studies, the experiments have been performed for short durations of scour (e.g., 7 h), in which the equilibrium depth of scour has not been achieved. Given that the scour at bridge piers

Table 1 Experimental studies using collars against scouring at bridge piers (w_c width of the collar, D pier width or diameter, H_c collar height from the sediment bed, h upstream flow depth)

Authors	No. piers	Pier shape	Collar shape	w_c/D	H_c/h	Maximum reduction in scour depth (%)
Zarrati et al. (2004)	1	Rectangular rounded nose	Rectangular rounded nose	2, 3	From $0.4D$ to $-0.8D^a$	74 for: $w_c/D = 3$ (on the sediment bed)
Zarrati et al. (2006)	2	Circular	Circular independent and continues	3	0	30 (first pier) 75 (second pier)
Moncada-M et al. (2009)	1	Circular	Circular	2, 3	$0, \pm 3, \pm 6, \pm 9$	100 (for: $w_c/D = 3$, on the sediment bed) 96 (for: $w_c/D = 2$, on the sediment bed)
Masjedi et al. (2010)	1	Oblong	Oblong	1, 1.5, 2, 2.5, 3	$0, -0.1D, -0.5D, -1D$	92 (for: $w_c/D = 3$ and $H_c/D = -0.1$)
Heidarpour et al. (2010)	2, 3	Circular	Circular	2, 3	0	45 (third pier of three tandem piers) 44 (second pier of two tandem piers)
Tafarojnoruz et al. (2012)	1	Circular	Circular	3	0	28.7
Chen et al. (2018)	1	Circular	Hooked	1.25	0, 0.25	42 (on the sediment bed)

^aNote The negative values of H_c/h are corresponding to the collar height below the sediment bed

increases over time until the equilibrium condition is achieved, it is necessary to assess the effect of collars on scour reduction at the equilibrium time.

Zarrati et al. (2004) showed that the maximum scour depth reduction at a rectangular nose pier was obtained when the collar was placed on the sediment bed. In contrast Kumar et al. (1999) found that lowering the collar below the bed led to less scour at the pier, followed by increasing the dimensions of the scour hole compared to the case where the collar was positioned on the bed. Masjedi et al. (2010) indicated that, at an oblong pier situated in a 180° channel bend, locating the collar under the sediment bed at the height of 0.1 times of the pier width caused the maximum scour depth reduction around the pier. The results obtained by Moreno et al. (2015) revealed that the maximum reduction in scour depth at a complex bridge pier occurred when its pile cap was partly placed below the sediment bed. As the collar installation height increased above the sediment bed, more flow could pass under the collar. Therefore, the influence of the collar on reducing the scour around the pier was diminished (Tanaka and Yano 1967; Zarrati et al. 2004; Moncada-M et al. 2009).

Scour around two tandem piers has been investigated in several experimental and numerical studies (Hannah 1978; Selamoglu et al. 2014; Wang et al. 2016; Khaple et al. 2017b; Memar et al. 2018; Amini and Asadi Parto 2017; Keshavarzi et al. 2018). Ataie-Ashtiani and Beheshti (2006) demonstrated that scour depth around two tandem piers reached its maximum value at the center-to-center spacing of the piers of $s = 3D$, where D is the pier diameter. Numerical simulations were performed by Kim et al. (2014) to estimate the scour depth at two piers with the different center-to-center spacing of the piers. The results obtained from the above-cited studies indicated that the scour depth at two piers was highly influenced by the center-to-center spacing of the piers. Increasing the center-to-center spacing of the piers increased the scour depth, reaching the maximum values at $s = 3D - 3.5D$.

The experiments in this study were designed twofold. On the one hand, few studies have addressed the influence of collars on scour depth reduction at two tandem piers, for which the effect of collar height from the sediment bed has not been investigated (see Table 1). Thus, in the present study, this gap of knowledge was systematically studied by varying the width and height of collars from the sediment bed. Consequently, the best height of collars from the sediment bed, leading to the maximum scour depth reduction around the two tandem piers, was determined. On the other hand, in previous studies, the experiments have been conducted for a limited time (e.g., 7 h) without achieving the equilibrium scour depth. In this work, most of the experiments were performed for long durations (2–10 days),

reaching the equilibrium scour conditions. The effect of time on the performance of collars was examined.

Methods and materials

Dimensional analysis

The equilibrium depth of scour, d_{se} , at two cylindrical tandem piers protected by circular collars inserted in a wide rectangular channel with a movable bed consisting of uniform nonripple-forming quartz sand in clear-water conditions and uniform flow depends on the following: sediment features (submerged sediment density $\rho'_s = (\rho_s - \rho)$, sediment grain density ρ_s , median sediment grain size D_{50} , geometric standard deviation of the sediment grain size distribution σ_g), fluid features (density of fluid ρ , fluid kinematic viscosity ϑ), flow (mean upstream flow velocity U , upstream flow depth h), collar (countermeasure) features (width of collar w_c , collar height from sediment bed H_c , thickness of collar t_c), pier diameter D , width of channel B , mean threshold velocity U_c , which is defined as mean flow velocity for the onset of sediment movement (threshold condition), and the center-to-center spacing of piers s . Considering variables D , U and ρ as the repeaters and applying the Buckingham theorem lead to the following dimensionless relationship:

$$\frac{d_{se}}{D} = \varphi \left(\frac{U}{U_c}, \frac{\rho'_s}{\rho}, \frac{UD}{\vartheta}, \frac{s}{D}, \frac{h}{D}, \frac{B}{D}, \frac{w_c}{D}, \frac{H_c}{h}, \frac{t_c}{D}, \frac{D}{D_{50}}, \sigma_g \right) \quad (1)$$

where φ stands for the unknown function.

The effect of dimensionless parameters in Eq. (1) has been previously examined by researchers. To attain the equilibrium depth of scour in clear-water conditions, flow intensity (U/U_c) is adjusted near the threshold of sediment movement to a range of $0.9 \leq U/U_c \leq 1$. The uniform nonripple-forming quartz sand is characterized by $D_{50} \geq 0.6$ mm, $\rho_s \approx 2650$ (kg/m³) and $\sigma_g < 1.4$, as recommended by Dey et al. (1995). Moreover, the effect of sediment size (coarseness effect), flow depth (shallowness effect), viscosity and walls of the channel (wall effect) needs to be eliminated. In order to do so, the following conditions recommended by different researchers were applied to the experiments.

The effect of flow depth on local scouring was overlooked by using the criterion of Melville and Sutherland (1988), which is defined as $h/D \geq 2.6$. An upstream flow depth of $h = 0.165$ m was kept constant for all the experiments. The influence of sediment grain size on scour can be negligible if $D/D_{50} \approx 25$ –130 (Tafarojnoruz et al. 2010). In the case of wide channels, the wall effect on scour process associated with the existence of the piers in the channel was ignored considering $B/D \geq 10$ (Melville

Table 2 Conditions applied to the experiments and consequent interpretation

Dimensionless parameter	Theoretical range	Suggested by	Effect	Calculated values in this study	Resulted in the following conditions
D/D_{50}	$D/D_{50} \approx 25\text{--}130$	Tafarojnoruz et al. (2010)	Coarseness	35.4	The coarseness does not influence the scour depth
h/D	$h/D \geq 2.6$	Melville and Sutherland (1988)	Shallowness	2.6	Flow depth is not considered to be shallow (even though it is at the lowest margin for being shallow)
B/D	$B/D \geq 10$	Melville and Chiew (1999)	Wall	20.6	The scour depth is not influenced by the width of the channel
R_p	$R_p > 7000$	Monti (1994)	Viscosity	(22,680, 23,814)	Flow is fully turbulent
σ_g	$\sigma_g < 1.4$	Dey et al. (1995)	Uniform sediment	1.23	Sediment grains can be assumed uniform

and Chiew 1999). The pier Reynolds number $R_p = UD/\nu$ was greater than 7000 to avoid the impact of viscosity on the scour depth (Monti 1994). Table 2 also shows the conditions applied to the experiments.

Taking the experimental conditions of the study by Khaple et al. (2017b) into consideration, the center-to-center spacing of the piers was selected as three times of the pier diameter. The authors showed that the maximum depths of scour at two tandem piers could occur at the center-to-center spacing of the piers equal to three or two times of the pier diameter; for both spacings, the scour depth was approximately identical. The thickness of collar t_c was selected 3 mm to avoid influencing the scour development. Assuming the submerged sediment specific gravity (ρ'_s/ρ) almost invariant for sand and gravel (≈ 1.65) (Tafarojnoruz et al. 2012), Eq. (1) becomes:

$$\frac{d_{se}}{D} = \varphi \left(\frac{U}{U_c}, \frac{w_c}{D}, \frac{H_c}{h} \right) \tag{2}$$

In this study, the influences of collar height and the width of the collar on scour depth at two tandem piers with different flow intensities were examined. Table 3 presents the experimental conditions and dimensionless parameters.

Table 3 Experimental conditions and dimensionless parameters (Q discharge, D diameter of the pier, U mean upstream flow velocity, U/U_c flow intensity, U_c mean threshold velocity, h upstream flow depth, B width of the channel, D_{50} median sediment grain size, H_c

Experiment series	Q (l/s)	D (m)	U (m/s)	U/U_c	h (m)	B (m)	D_{50} (m)	H_c/h	w_c/D	s/D	R_p	F_r
I (reference)	77, 81	0.063	0.36	0.9,0.95	0.165	1.3	0.00178	–	–	3	22,680	0.283
											23,814	0.297
II	77	0.063	0.36	0.9	0.165	1.3	0.00178	0, –0.1	3	3	22,680	0.283
III	77	0.063	0.378	0.9	0.165	1.3	0.00178	0, –0.1	2	3	23,814	0.297
IV	81	0.063	0.378	0.95	0.165	1.3	0.00178	0, –0.1	2	3	23,814	0.297

Experimental installation

Laboratory experiments were performed in a straight part of a curved rectangular cross-section channel with 9 m length, 1.3 m width and 0.6 m depth at Laboratory of Hydraulic Constructions (LCH-EPFL), Lausanne, Switzerland. The channel included the experiment section (recess box) located 4 m downstream from the inlet of the channel with 0.25 m depth and 5 m length in the streamwise direction. The sediment utilized in this study was quartz sand of the median diameter $D_{50} = 1.78$ mm and the geometric standard deviation of the sediment grain size distribution $\sigma_g = 1.23$. This sediment is considered nonripple forming and non-cohesive. The model piers were provided using circular transparent PVC tubes with the diameter of 0.063 m and located at 7 m from the channel inlet, where a fully developed flow was established. The water was delivered to the channel by an automatically operated pump and the discharge was measured by an electromagnetic discharge meter. A gate that was positioned at the end of the channel regulated the water depth (h). The water and sediment bed levels were measured with a digital point gauge having the precision of 0.1 mm. To generate a uniform flow distribution, a metal net and a filter sponge were placed at the inlet of the channel. Periscopes were fixed inside the piers to read the time development of

collar height from the sediment bed, w_c width of the collar, s center-to-center spacing of the piers, R_p pier Reynolds number, F_r Froude number)

scour depth in front of the piers from the ruler papers glued to the piers. A Go Pro camera was installed on the wall of the channel under the water to take pictures and record videos of the scouring process around the piers. A 3D laser Baumer, OADM 1317480/S35A, mapped the topography of the scour holes around the piers at the end of the experiments.

The critical shear velocity u_c^* was initially determined through the method presented by Lança et al. (2015), which is $u_c^* = \sqrt{\Delta g D_{50} \tau_c^*} \approx 0.032$ and the critical shear stress is $\tau_c = \rho u_c^{*2}$. Parameter τ_c^* is the critical value of the shields parameter that is defined as: $\tau_c^* = 0.24/d^* + 0.055[1 - \exp(-d^{*1.05}/58)]$, where d^* is $\sqrt[3]{g\Delta D_{50}^3/\vartheta^2}$. The mean threshold velocity U_c was estimated using different relationships suggested by Shamov (1952) [cited in Dey (2014)], Melville and Sutherland (1988), Shepard et al. (2013) and Lança et al. (2015). Then, its value was determined by conducting some experiments without locating the piers and countermeasures in the channel. The results were in good agreement with the relationship recommended by Shamov (1952) [cited in Dey (2014)]. In the present work, most of the experiments were performed for a long duration (2–10 days). To stop the experiments, the criterion of Melville and Chiew (1999) was used. According to Melville and Chiew (1999), when the scour depth discrepancies during 24 h is not more than 5% of the pier diameter, the equilibrium phase is reached.

Configurations of experiments

For assessing the optimal height of collars from the sediment bed results in the minimum scour depth at the piers' front, the collars were positioned: (1) on the sediment bed $H_c = 0$ and (2) beneath the sediment bed at $H_c = -0.1h$ (in which h is upstream flow depth), under two different flow intensities of 0.9 (experiment series II and III) and 0.95 (experiment series IV) (see Table 3). It has to be clarified that, in the present work, the collars were not placed above the sediment bed. Given the outcomes presented by Tanaka and Yano (1967) and Zarrati et al. (2004), as the collars'

installation height above the sediment bed was increased, the flow could pass more easily under the collar. As a consequence, the protective effect of the collar on scour depth at the pier was reduced. Two different circular collars (made from PVC transparent sheets) of width $w_c = 2D, 3D$ (where D is the pier diameter) were employed. The schematic plan of two tandem piers equipped with the circular collars with different width is shown in Fig. 2. Figure 3 shows the configurations of two tandem piers without and with the collars.

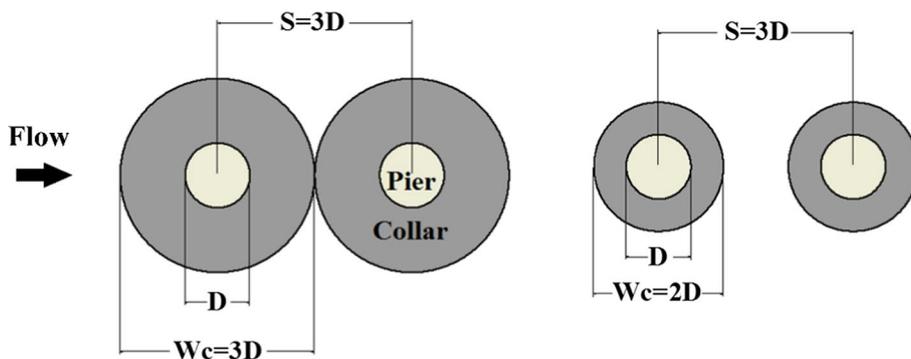
Results and discussion

Experimental observations

The results are shown in Table 4. Subscripts “ u ” and “ d ” refer to upstream and downstream piers. In the case of the experiments without collars (experiments A1 and A2), scour began at the front and sides of the piers, extended toward both downstream and upstream of them. By installing the collars around the piers, scour started from the downstream of the collars, then developed around the collars' edge and reached the piers' front after a time duration. Thus, the collars delayed the start of scour in front of the piers. The time from the beginning of the experiment to when the scour started at the piers' front is referred to as delay time and is shown by t_d in Table 4. Parameters t_{du} and t_{dd} indicate the delay time of scour at the upstream and downstream piers, respectively. In experiment C1, the rear of both collars was initially eroded. The scour began at the downstream pier's front after 1.5 h from the beginning of the experiment, while it reached the upstream pier's front later at $t_{du} = 2.5$ h. However, in the other cases (experiments C2, D1, D2), the delay time of scour at the upstream pier was greater than that at the downstream pier ($t_{du} > t_{dd}$) (see Table 4). This is due to the lower rate of scouring in experiment C1 than the other cases.

In the experiment series II, the broader collars $w_c = 3D$ caused maximum reduction in scour depth of about 100% at the collar height of $H_c = 0$ and the average of 82% at $H_c = -0.1h$ in front of the piers (at $t \approx 250$ min), where h stands for

Fig. 2 Schematic plan of two tandem piers equipped with collars (w_c width of the collar, D diameter of the pier, s center-to-center spacing of the piers)



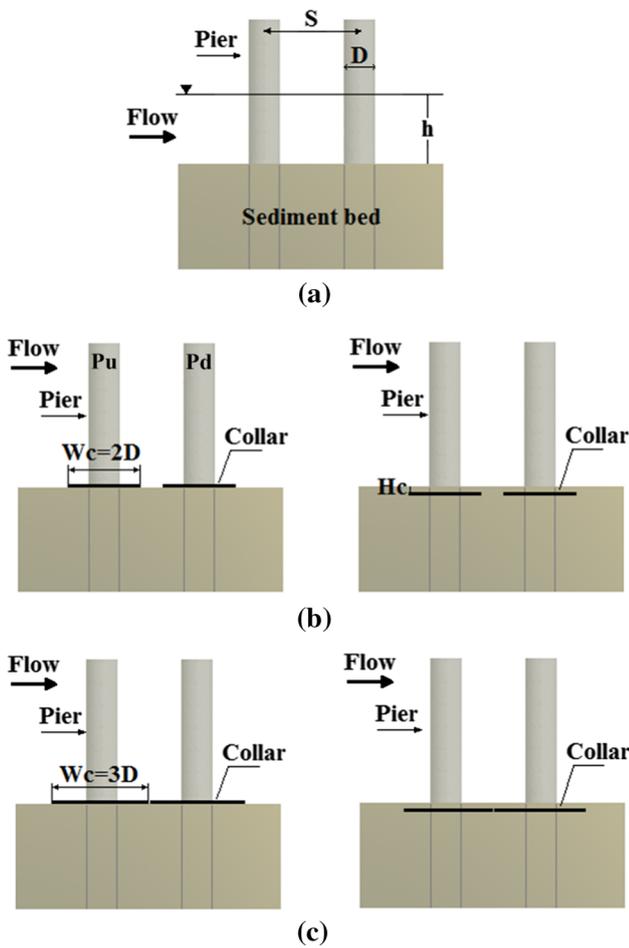


Fig. 3 Configurations of studied experiments: **a** two tandem piers, **b** two tandem piers equipped with collars of widths $w_c = 2D$, located on the sediment bed ($H_c = 0$) and beneath the sediment bed ($H_c = -0.1h$), **c** two tandem piers equipped with collars of widths $w_c = 3D$, located on the sediment bed ($H_c = 0$) and beneath the sediment bed ($H_c = -0.1h$) (s center-to-center spacing of the piers, D diameter of the pier, H_c collar height from the sediment bed, h upstream flow depth)

Table 4 Dimensionless parameters, experiment duration (time of scour) t , delay time t_d , scour depth at the front of the piers d_s , extrapolated scour depth at the front of the piers at infinite time $d_{s(exp)}$

Series	Experiment	U/U_c	w_c/D	H_c/h	t (h)	t_{du} (h)	t_{dd} (h)	d_{su} (m)	d_{sd} (m)	$d_{su(exp)}$ (m)	$d_{sd(exp)}$ (m)
I (reference)	A1 ^a	0.9	–	–	62.5	–	–	0.105	0.081	0.116	0.088
	A2 ^a	0.95	–	–	120	–	–	0.129	0.096	0.145	0.106
II	B1	0.9	3	0	250	–	–	0	0	–	–
	B2	0.9	3	–0.1	240	–	–	0.0165	0.0165	–	–
III	C1 ^a	0.9	2	0	72	2.5	1.5	0.024	0.0135	0.03	0.025
	C2	0.9	2	–0.1	41	11	13	0.055	0.0315	0.103	0.043
IV	D1 ^a	0.95	2	0	144	2	3	0.106	0.065	0.122	0.08
	D2	0.95	2	–0.1	44	9	11.5	0.09	0.045	0.124	0.084

Subscripts u and d refer to the first and second piers, respectively

^aEquilibrium condition

upstream flow depth. It was observed that approximately 5 min after the experiments began, sediments on the collars scoured. Collars covered the sediment bed around the piers and between them; therefore, scour occurred at the flank and downstream of collars with no upstream erosion (see Fig. 4).

Time development of scour depth in front of the two tandem piers (experiments A1 and A2) is represented in Fig. 5 on a logarithmic scale. The equilibrium depth of scour at upstream piers was on average 30% greater than their values for the downstream piers. The existence of the upstream pier led to flow velocity reduction toward the downstream pier. Thus, the power of downflow and horseshoe vortices around the downstream pier was weakened. In consequence, the rate of scouring at the upstream pier was higher than that at the downstream pier. Parameters P_u and P_d symbolize the upstream and downstream piers, respectively.

Influence of collar height on scour depth

Dimensionless time development of scour depth at two tandem piers equipped with the collars with width $w_c = 2D$ for different collar height can be compared, as shown in Fig. 6, with the coordinates of Oliveto and Hager (2002, 2005) and characterized as follows:

$$Z = \frac{d_s}{z_R}, \quad z_R = (hD^2)^{1/3} \tag{3}$$

$$T = \frac{t}{t_R}, \quad t_R = \frac{z_R}{\sigma^{1/3}(\Delta g D_{50})^{1/2}} \tag{4}$$

$$D^* = D_{proj} * K_{sp} * K_m \tag{5}$$

where $F_d = U/(\Delta g d_{50})^{1/2}$ is densimetric particle Froude number, D_{50} is median sediment size, $\Delta g = [(\rho_s - \rho)/\rho]g$, ρ is fluid density, ρ_s is sediment grain density, g is acceleration

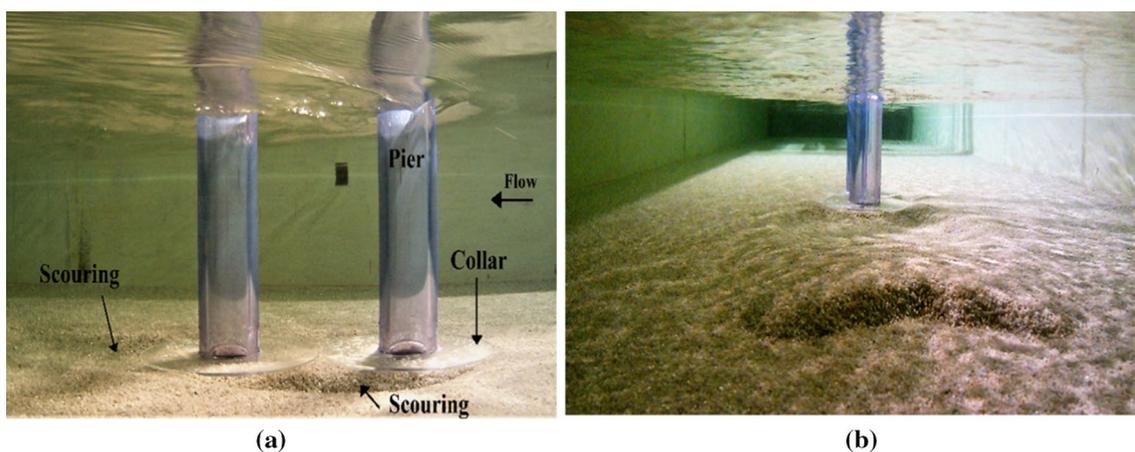


Fig. 4 The scour around two tandem piers equipped with collars of width $w_c = 3D$ located on the sediment bed: **a** side view and **b** view from downstream in the upstream direction ($t > 200$ h, under the water)

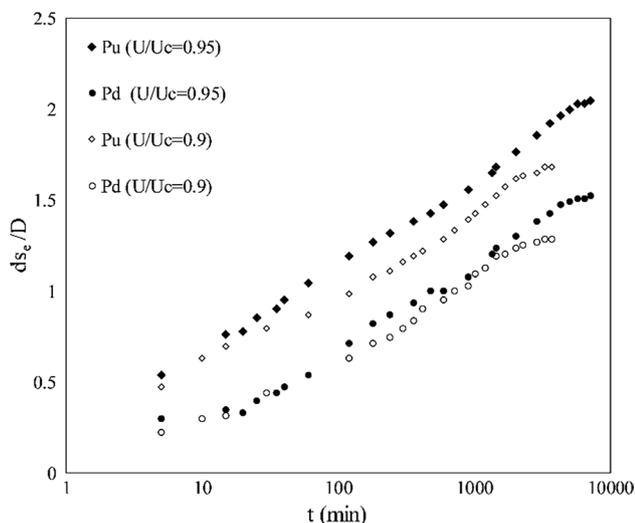


Fig. 5 The scour depth development at the front of two tandem piers at different flow intensities, in logarithmic scale (d_{se} equilibrium depth of scour, D diameter of the pier, t time of scour)

of gravity, Z is dimensionless scour depth, T is dimensionless time of scour, d_s is the maximum depth of scour, t is time of scour, σ_g is geometric standard deviation of the sediment grain size distribution and h is upstream flow depth. To consider the piers interaction effect on scour depth, the effective width of an equivalent full depth pier (D^*), recommended by FHWA (HEC-18), was replaced with the pier diameter. Thus, D_{proj} is the sum of the non-overlapping projected widths of the piers, K_{sp} is center-to-center spacing factor and K_m is the number of aligned rows factor (see Arneson et al. 2012).

Results indicated that locating the collars beneath the sediment bed increased delay time t_d compared to the case that the collars were set on the bed. However, greater scour depth

occurred at the front of the piers (see also Table 4). In the case of $H_c = -0.1h$, the obtained delay times were on average 4 and 5.5 times larger than the case of $H_c = 0$ at the upstream and downstream piers, respectively. After time t_d , the collars on the sediment bed can more efficiently slow down the scouring rates than the collars under the bed. This is proved by the slopes of the scour development curves in Fig. 6. Therefore, the optimal height of collars from the sediment bed, for which the maximum reduction of scour depth around the piers occurred, was $H_c = 0$ (on the original sediment bed).

Effect of time of scour on the collar’s effectiveness

As already mentioned, the dimensions of the scour hole around a bridge pier are increasing until the equilibrium scour condition is achieved. Conversely, some authors claim that the scour process may carry on even if the scour topography seems to attain an equilibrium state (Oliveto and Hager 2002, 2005). Given these points, it is essential to assess the effect of collars on scour depth around the piers in the equilibrium condition. Additionally, assuming that the equilibrium phase is attained asymptotically (Chabert and Engeldinger 1956; Etema 1980), the measured depth of scour at the end of the experiments was extrapolated to the infinite time ($t = \infty$) using the equation recommended by Lança et al. (2010). The extrapolated depth of scour is presented by $d_{s(exp)}$ in Table 4. The values of scour depth reduction R are calculated using Eq. (6):

$$R(\%) = \frac{d_{s0} - d_s}{d_{s0}} * 100 \tag{6}$$

where subscript 0 indicates the experiments without collars. In Table 5, the scour depth reduction values for different times of the experiments and for the infinite time (for the

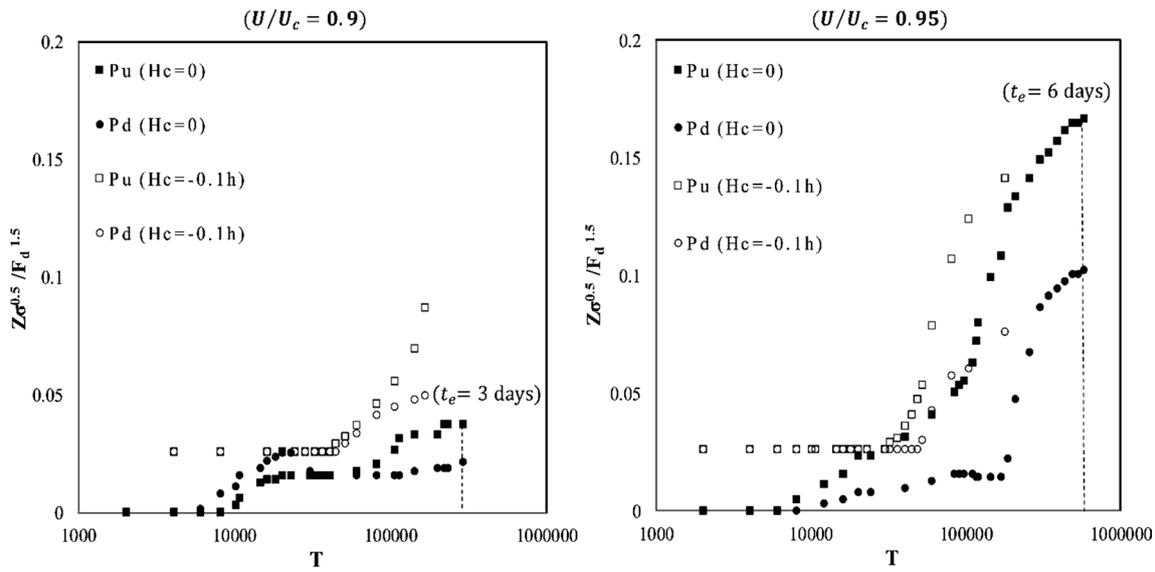


Fig. 6 Dimensionless development of the scour depths at the front of two tandem piers protected by collars for collar heights of $H_c = 0$ and $H_c = -0.1h$, at different flow intensities, with the coordinates of Oliveto and Hager (2002, 2005), in logarithmic scale (Z dimension-

less scour depth, T dimensionless time of scour, σ_g geometric standard deviation of the sediment grain size distribution, F_d densimetric particle Froude number, h upstream flow depth (also see Fig. 3b)

Table 5 Percent reduction in scour depth at different times t , at the front of the first pier R_u and the second pier R_d (H_c/h collar height from sediment bed to the upstream flow depth, U/U_c flow intensity)

Series	Experiment	H_c/h	U/U_c	R_u (%)			R_d (%)		
				$t = 24$ h	$t = 40$ h	$t = \infty^a$	$t = 24$ h	$t = 40$ h	$t = \infty$
II	C1	0	0.9	83.3	82	74.13	86.6	86.5	71.5
	C2	-0.1	0.9	64.5	52	11.2	63.3	63	51.13
IV	D1	0	0.95	67	34	16	87	88	24.5
	D2	-0.1	0.95	28.3	15.1	14.4	40	43	20.1

^aInfinite time

extrapolated scour depths) are given. Parameters R_u and R_d represent the scour depth reduction in front of the upstream and downstream piers. The results showed that the effectiveness of collars was a time-dependent parameter. Comparing the scour depth reduction at both collar heights revealed that, when the collars were placed on the bed, the maximum efficiency for them was obtained.

As can be observed in Fig. 7, when the collars were installed on the bed (experiments C1 and D1), the effect of the collar was decreased over time at the upstream pier. At the downstream pier, the scouring process was influenced by the upstream sediment transport. The moving sediment grains from upstream were continuously entering and leaving the scour hole around the downstream pier until the region between the piers was flattened. The minimum reduction in scour depth was related to the final state (extrapolated depth of scour). For $H_c = -0.1h$, the scour depth reduction was depicted after the delay time t_d .

Additionally, in experiments C1 and D1, as the flow intensity enhanced from 0.9 to 0.95, the maximum

reduction in scour depth was decreased from the average of 70–20% at the piers. The equilibrium time of scour under the flow intensity of 0.95 (experiment D1) was nearly two times of its value under the flow intensity of 0.9 (experiment C1). As mentioned before, after the scour reached the piers’ front, the rate of scour around the piers under the flow intensity of 0.95 was higher than under the flow intensity of 0.9, as evident by the slopes of scouring curves in Fig. 6. According to Melville and Chiew (1999) under clear-water conditions, scour depth enhances with flow intensity (which is determined as the relative mean flow velocity (U) to the mean threshold velocity (U_c)), reaching a maximum amount at the flow intensity equal to 1. Generally, increasing the flow intensity increases the erosive effect of the flowing water, as the bed-shear stress enhances. As the consequence, the scouring rate around the piers enhances. The process of sediment grains erosion is consistent with prevailing bed-shear stress, which can be defined from the measured velocity profile (Houwing and Van Rijn 1998).

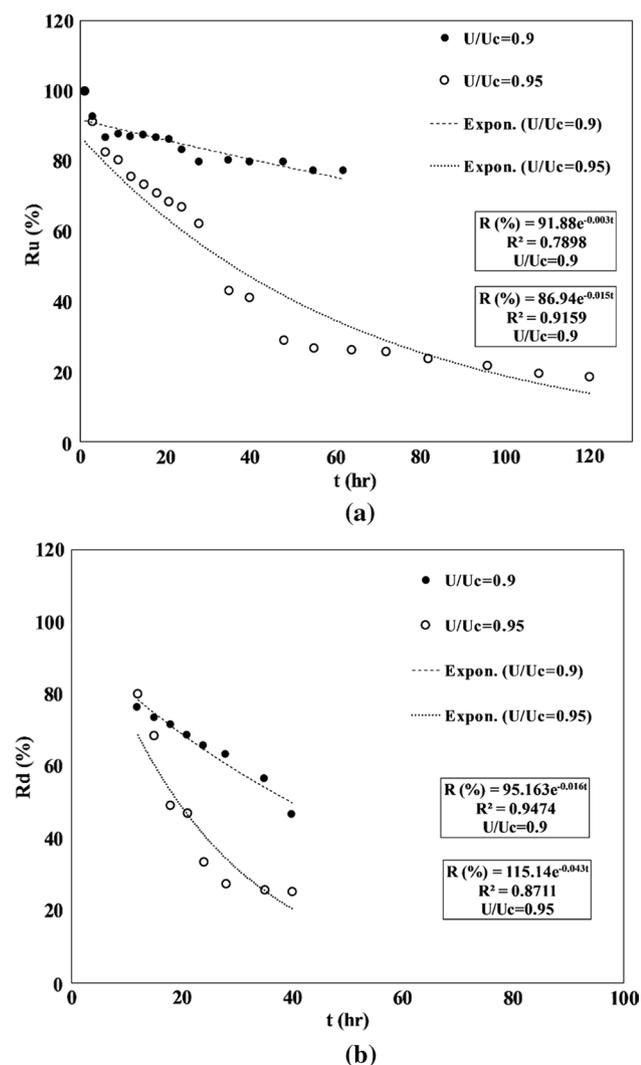


Fig. 7 Reduction in scour depth over time t (experiments C1 and D1): **a** at the front of the first pier R_u , **b** at the front of the second pier R_d , at different flow intensities U/U_c

Given the results explained above, at the piers protected by collars, the flow intensity makes a significant contribution to scour phenomena at the pier's site. The results are in line with those found by Karimaei Tabarestani and Zarrati (2019). The authors indicated that the flow intensity was a

dominant variable affecting the effectiveness of the collar in scour depth reduction around the pier. In Fig. 6, the equilibrium times are shown by the dotted lines.

Figure 8 shows the column diagram of the scour depth reduction at different times of the experiments. The eroded sediment beds at the end of experiments C1 and D1 are demonstrated in Fig. 9. The equilibrium depth of scour in front of two tandem piers protected by the collars is demonstrated with contour lines in Fig. 10. Recall that experiments C1 and D1 were performed until the equilibrium scour condition was achieved.

Conclusions

In this experimental work, the effect of the collar width and collar height from the sediment bed on the reduction in scour depths at two tandem piers was investigated. The following conclusions were made:

1. The optimal height of the collars from the bed, for which the maximum scour reductions around the piers was obtained, was the height of the initial sediment bed (collars situated on the sediment bed).
2. The flow intensity significantly influenced the scour depth at the piers equipped with collars. Generally, under clear-water conditions, by increasing the flow intensity, the scour depth at piers enhances. Herein, reducing the flow intensity from 0.95 to 0.9 increased the maximum reduction in scour depth from the average of 20–70% (for the collars placed on the bed).
3. For the collars set underneath the sediment bed, scour reached the piers' front at the much longer time than the condition, in which the collars were on the sediment bed. However, greater scour depth occurred in front of the piers. To put it another way, the delay time (the time from starting the experiment to when the scour started at the front of the piers) was increased for the average of 4 and 5.5 times at the upstream and downstream piers, correspondingly.
4. The collar's performance depended on time and decreased over time.

Fig. 8 Column diagram of the scour depth reduction for different time durations of the experiments at different flow intensities U/U_c (R_u reduction in scour depth for the first pier, R_d reduction in scour depth for the second pier, t time of scour)

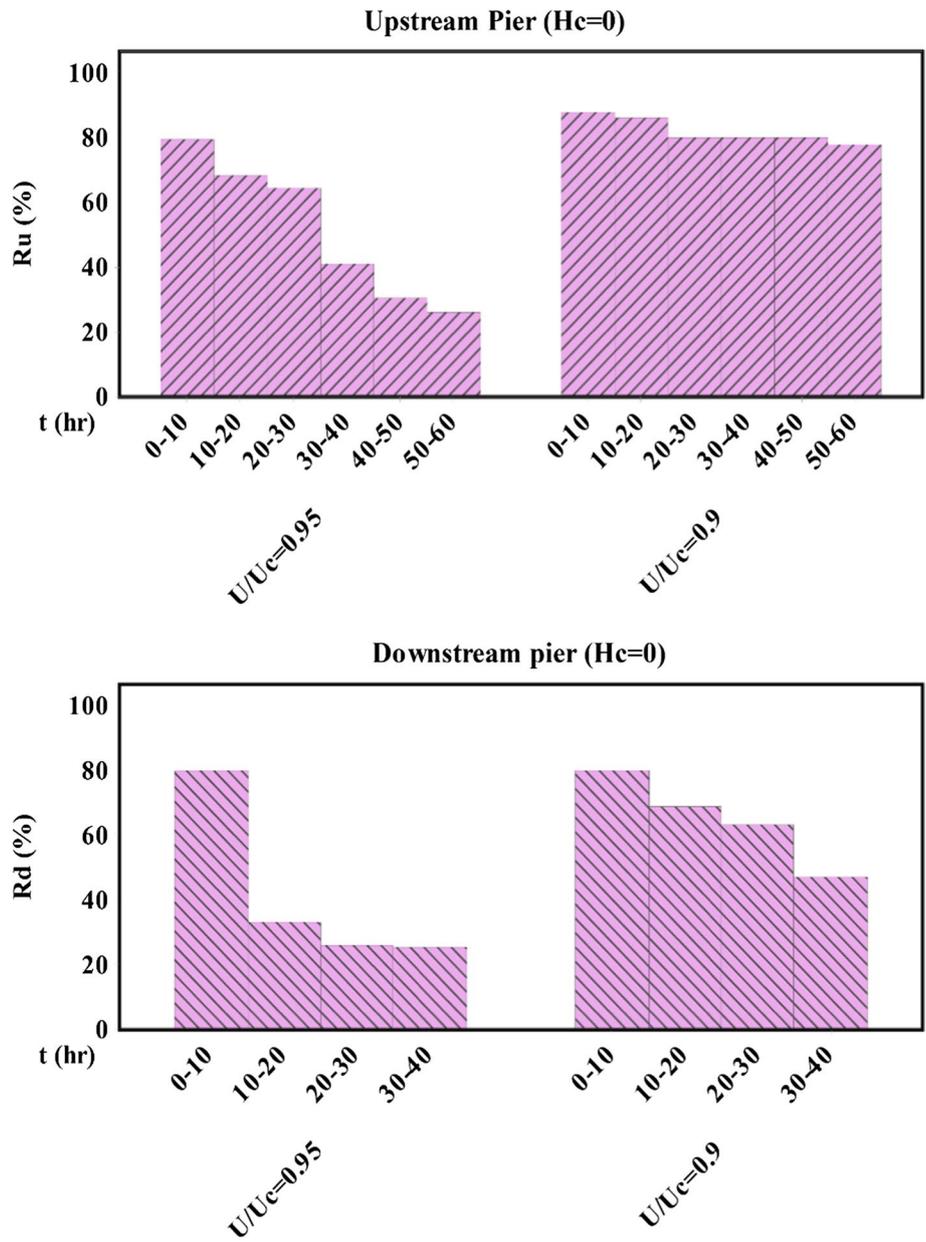


Fig. 9 Eroded sediment bed at the end of experiments C1 and D1 at different flow intensities (for collar width $w_c = 2D$, and collar height $H_c = 0$)

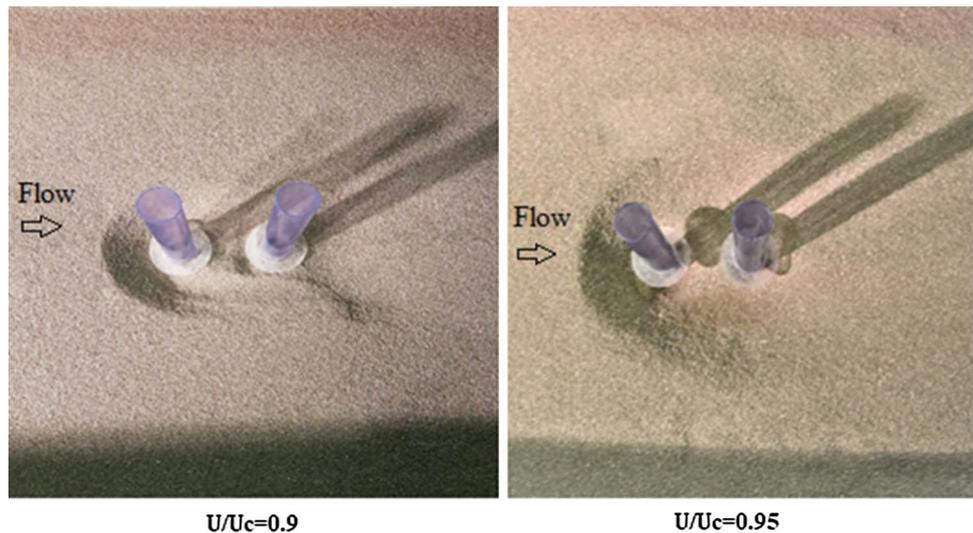
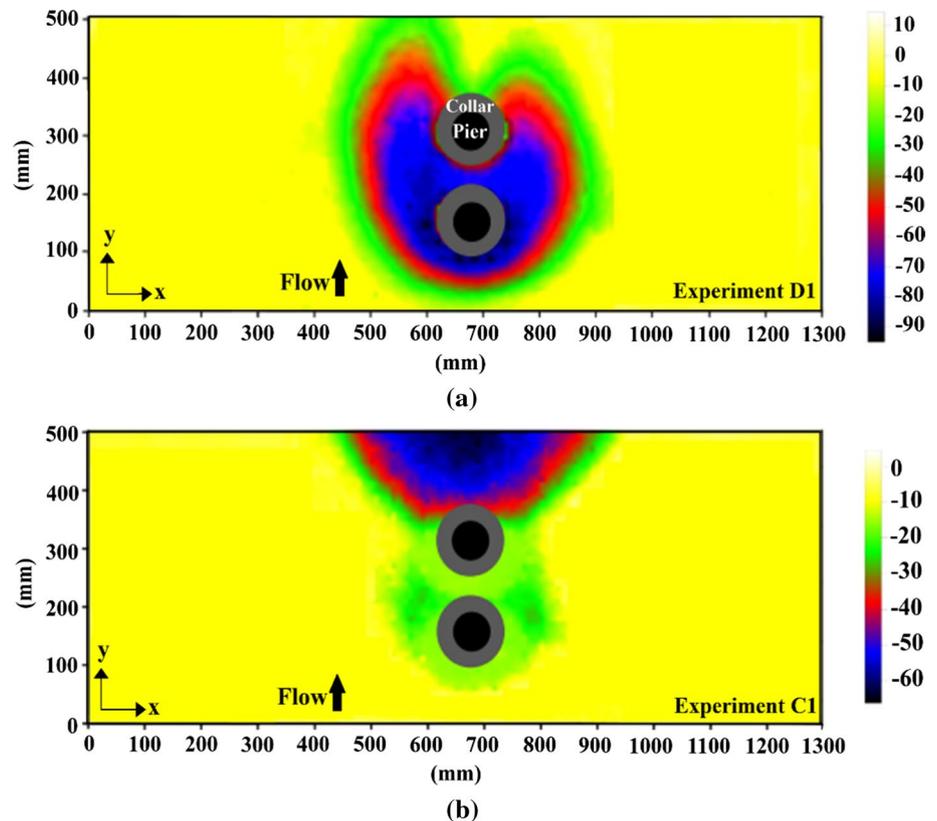


Fig. 10 Contour lines of scour at tandem piers equipped with collars on the sediment bed ($H_c = 0$) at different flow intensities, at the equilibrium time: **a** experiment D1, **b** experiment C1



Acknowledgements This experimental research was conducted at Laboratory of Hydraulic Constructions (LCH-EPFL), Lausanne, Switzerland, which supported the project financially as well.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Amini A, Asadi Parto A (2017) 3D numerical simulation of flow field around twin piles. *Acta Geophys* 65(6):1243–1251. <https://doi.org/10.1007/s11600-017-0094-x>
- Arneson PF, Zevenbergen LA, Lagasse LW (2012) Evaluating scour at bridges. Hydraulic engineering circular no. 18 (HEC-18) (report no. FHWA NHI 01-001). Federal Highway Administration, Washington, DC

- Ataie-Ashtiani B, Beheshti AA (2006) Experimental investigation of clear-water local scour at pile groups. *J Hydraul Eng* 132(10):1100–1104. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2006\)132:10\(1100\)](https://doi.org/10.1061/(ASCE)0733-9429(2006)132:10(1100))
- Chabert J, Engeldinger P (1956) Study of scour around bridge piers. Report prepared for the Laboratoire National d'Hydraulique
- Chen SC, Tfwala S, Wu TY, Chan HC, Chou HT (2018) A hooked-collar for bridge piers protection: flow fields and scour. *Water* 10(9):1251. <https://doi.org/10.3390/w10091251>
- Chiew YM (1984) Local scour at bridge piers. Doctoral dissertation, University of Auckland
- Dey S (1997a) Local scour at piers, part I: a review of developments of research. *Int J Sediment Res* 12(2):23–46
- Dey S (1997b) Local scour at piers, part II: bibliography. *Int J Sediment Res* 12(2):47–57
- Dey S (2014) *Fluvial hydrodynamics*. Springer, Berlin
- Dey S, Bose SK, Sastry GLN (1995) Clear water scour at circular piers: a model. *J Hydraul Eng* 121:869–876. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1995\)121:12\(869\)](https://doi.org/10.1061/(ASCE)0733-9429(1995)121:12(869))
- Ettema R (1980) Scour at bridge piers. Report no. 216. University of Auckland, Auckland, New Zealand
- Ettema R, Nakato T, Muste MVI (2006) An illustrated guide for monitoring and protecting bridge waterways against scour (No. Project TR-515). IIHR-Hydroscience & Engineering, University of Iowa
- Guo J, Suaznabar O, Shan H, Shen J (2012) Pier scour in clear-water conditions with non-uniform bed materials (No. FHWA-HRT-12-022). Turner-Fairbank Highway Research Center, McLean
- Hannah CR (1978) Scour at pile groups. Report no. 78-3. Canterbury University, Canterbury, New Zealand
- Heidarpour M, Afzalimehr H, Izadinia E (2010) Reduction of local scour around bridge pier groups using collars. *Int J Sediment Res* 25(4):411–422. [https://doi.org/10.1016/S1001-6279\(11\)60008-5](https://doi.org/10.1016/S1001-6279(11)60008-5)
- Houwing EJ, Van Rijn LC (1998) In situ erosion flume (ISEF): determination of bed-shear stress and erosion of a kaolinite bed. *J Sea Res* 39(3–4):243–253
- Johnson PA, Hey RD, Brown ER, Rosgen DL (2002) Stream restoration in the vicinity of bridges. *J Am Water Resour Assoc* 38(1):55–67. <https://doi.org/10.1111/j.1752-1688.2002.tb01534.x>
- Karimaei Tabarestani M, Zarrati AR (2019) Local scour depth at a bridge pier protected by a collar in steady and unsteady flow. In: *Proceedings of the Institution of Civil Engineers—water management*. Thomas Telford Ltd, London, pp 1–11. <https://doi.org/10.1680/jwama.18.00061>
- Keshavarzi A, Shrestha CK, Melville BW, Khabbaz H, Ranjbar-Zahedani M, Ball J (2018) Estimation of maximum scour depths at upstream of front and rear piers for two tandem circular columns. *Environ Fluid Mech* 18(2):537–550. <https://doi.org/10.1007/s10652-017-9572-6>
- Khaple S, Hanmaiahgari PR, Gaudio R, Dey S (2017a) Splitter plate as a flow-altering pier scour countermeasure. *Acta Geophys* 65(5):957–975. <https://doi.org/10.1007/s11600-017-0084-z>
- Khaple S, Hanmaiahgari PR, Gaudio R, Dey S (2017b) Interference of an upstream pier on local scour at downstream piers. *Acta Geophys* 65(1):29–46. <https://doi.org/10.1007/s11600-017-0004-2>
- Khodashenas SR, Shariati H, Esmaeli K (2018) Comparison between the circular and square collar in reduction of local scouring around bridge piers. In: *E3S web of conferences*. EDP sciences, vol 40, p 03002
- Kim HS, Nabi M, Kimura I, Shimizu Y (2014) Numerical investigation of local scour at two adjacent cylinders. *Adv Water Resour* 70:131–147. <https://doi.org/10.1016/j.advwatres.2014.04.018>
- Kumar V, Raju KGR, Vittal N (1999) Reduction of local scour around bridge piers using slots and collars. *J Hydraul Eng* 125(12):1302–1305. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1999\)125:12\(1302\)](https://doi.org/10.1061/(ASCE)0733-9429(1999)125:12(1302))
- Lança R, Fael C, Cardoso A (2010) Assessing equilibrium clear water scour around single cylindrical piers. In: *Proceedings of the international conference on fluvial hydraulic (river flow)*, Braunschweig, Germany, September 8–10
- Lança R, Fael C, Maia R, Pêgo JP, Cardoso AH (2013) Clear-water scour at pile groups. *J Hydraul Eng* 139(10):1089–1098. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0000770](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000770)
- Lança RMM, Simarro G, Fael CMS, Cardoso AH (2015) Effect of viscosity on the equilibrium scour depth at single cylindrical piers. *J Hydraul Eng* 142(3):06015022. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0001102](https://doi.org/10.1061/(ASCE)HY.1943-7900.0001102)
- Masjedi A, Bejestan MS, Esfandi A (2010) Experimental study on local scour around single oblong pier fitted with a collar in a 180 degree channel bend. *Int J Sediment Res* 25(3):304–312. [https://doi.org/10.1016/S1001-6279\(10\)60047-9](https://doi.org/10.1016/S1001-6279(10)60047-9)
- Melville BW (1984) Live-bed scour at bridge piers. *J Hydraul Eng* 110(9):1234–1247. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1984\)110:9\(1234\)](https://doi.org/10.1061/(ASCE)0733-9429(1984)110:9(1234))
- Melville BW, Chiew YM (1999) Time scale for local scour at bridge piers. *J Hydraul Eng* 114(10):1210–1226. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1999\)125:1\(59\)](https://doi.org/10.1061/(ASCE)0733-9429(1999)125:1(59))
- Melville BW, Sutherland AJ (1988) Design method for local scour at bridge piers. *J Hydraul Eng* 114(10):1210–1226. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1988\)114:10\(1210\)](https://doi.org/10.1061/(ASCE)0733-9429(1988)114:10(1210))
- Memar S, Zounemat-Kermani M, Beheshti A, De Cesare G, Schleiss AJ (2018) Investigation of local scour around tandem piers for different skew-angles. In: *International conference on fluvial hydraulics (river flow)*, Lyon-Villerurbanne, France, September 5–8
- Moncada-M AT, Aguirre-Pe J, Bolivar JC, Flores EJ (2009) Scour protection of circular bridge piers with collars and slots. *J Hydraul Res* 47(1):119–126. <https://doi.org/10.3826/jhr.2009.3244>
- Monti R (1994) Indagine sperimentale delle caratteristiche fluidodinamiche del campo di moto intorno ad una pila circolare. Tesi di Dottorato di Ricerca, Politecnico di Milano, Milan, Italy (**in Italian**)
- Moreno M, Maia R, Couto L (2015) Effects of relative column width and pile-cap elevation on local scour depth around complex piers. *J Hydraul Eng* 142(2):04015051. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0001080](https://doi.org/10.1061/(ASCE)HY.1943-7900.0001080)
- Oliveto G, Hager WH (2002) temporal evolution of clear-water pier and abutment scour. *J Hydraul Eng* 128(9):811–820. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2002\)128:9\(811\)](https://doi.org/10.1061/(ASCE)0733-9429(2002)128:9(811))
- Oliveto G, Hager WH (2005) Further results to time-dependent local scour at bridge elements. *J Hydraul Eng* 131(2):97–105. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2005\)131:2\(97\)](https://doi.org/10.1061/(ASCE)0733-9429(2005)131:2(97))
- Pang ALJ, Skote M, Lim SY, Gullman-Strand J, Morgan N (2016) A numerical approach for determining equilibrium scour depth around a mono-pile due to steady currents. *Appl Ocean Res* 57:114–124. <https://doi.org/10.1016/j.apor.2016.02.010>
- Raudkivi AJ (1986) Functional trends of scour at bridge piers. *J Hydraul Eng* 112(1):1–13. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1986\)112:1\(1\)](https://doi.org/10.1061/(ASCE)0733-9429(1986)112:1(1))
- Selamoglu M, Yanmaz AM, Koken M (2014) Temporal variation of scouring topography around dual bridge piers. In: *Proceedings of the seventh international conference on scour and erosion*, Perth, Western Australia
- Shen HW, Schneider VR, Karaki SS (1966) *Mechanics of local scour*. Report no. CER66HWS10. Colorado State University, Fort Collins, CO
- Sheppard DM, Miller W Jr (2006) Live-bed local pier scour experiments. *J Hydraul Eng* 132(7):635–642
- Sheppard DM, Melville B, Demir H (2013) Evaluation of existing equations for local scour at bridge piers. *J Hydraul Eng* 140(1):14–23. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0000800](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000800)

- Tafarojnoruz A, Gaudio R, Grimaldi C, Calomino F (2010) Required conditions to achieve the maximum local scour depth at a circular pier. In: Proceeding XXXII Convegno Nazionale di Idraulica e Costruzioni Idrauliche, 14–17 September, Palermo, Italy, Farina, Palermo
- Tafarojnoruz A, Gaudio R, Calomino F (2012) Evaluation of flow-altering countermeasures against bridge pier scour. *J Hydraul Eng* 138(3):297–305. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0000512](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000512)
- Tanaka S, Yano M (1967) Local scour around a circular cylinder. In: Proceeding of 12th IAHR congress. International Association for Hydraulic Research, Delft, Netherlands
- Wang H, Tang H, Liu Q, Wang Y (2016) Local scouring around twin bridge piers in open-channel flows. *J Hydraul Eng* 142(9):06016008. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0001154](https://doi.org/10.1061/(ASCE)HY.1943-7900.0001154)
- Zarrati AR, Gholami H, Mashahir MB (2004) Application of collar to control scouring around rectangular bridge piers. *J Hydraul Res* 42(1):97–103. <https://doi.org/10.1080/00221686.2004.9641188>
- Zarrati AR, Nazariha M, Mashahir MB (2006) Reduction of local scour in the vicinity of bridge pier groups using collars and riprap. *J Hydraul Eng* 132(2):154–162. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2006\)132:2\(154\)](https://doi.org/10.1061/(ASCE)0733-9429(2006)132:2(154))
- Zokaei M, Zarrati AR, Salamatian SA, Tabarestani MK (2013) Study on scouring around bridge piers protected by collar using low density sediment. *Int J Civ Eng* 11(3A):199–205
- Zounemat-Kermani M, Beheshti AA, Ataie-Ashtiani B, Sabbagh-Yazdi SR (2009) Estimation of current-induced scour depth around pile groups using neural network and adaptive neuro-fuzzy inference system. *Appl Soft Comput* 9(2):746–755. <https://doi.org/10.1016/j.asoc.2008.09.006>



Riverbed armoring and sediment exchange process in a sand–gravel bed reach after the Three Gorges Project operation

Linlin Li¹ · Junqiang Xia¹ · Meirong Zhou¹ · Shanshan Deng¹

Received: 27 September 2019 / Accepted: 2 December 2019 / Published online: 1 January 2020
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2020

Abstract

Upstream damming greatly altered the flow and sediment regime entering downstream reaches in the Middle Yangtze River, and the bed material in a sand–gravel bed reach coarsened continuously, which had a significant influence on the sediment transport and bed evolution. In order to study the riverbed armoring, the sediment exchange process (SEP) among bed material, bed load and suspended load in a sand–gravel bed river is firstly clarified, and then, the three-state transition probability model (Markov chain) is proposed in this study, with the hiding-exposure effect of non-uniform sediment being considered. Finally, the equilibrium equation of sediment quantity in an active layer is presented to calculate the grain size distribution of bed material. In this model, the influences of flow and sediment conditions, riverbed erosion and deposition on the SEP are discussed. The results show that the composition of surface bed material at the Zhicheng station became obviously coarse, and the median grain size (d_{50}) of surface bed material increased from 0.230 to 0.424 mm in 2003–2017, with an upward increasing trend. The proposed probabilistic model was validated against field measurements of bed material, and calculated results show reasonable agreement with the measured data at Zhicheng. Accordingly, the probabilistic model can be used to predict the riverbed armoring and to investigate the non-equilibrium transport of non-uniform sediment in a sand–gravel bed river.

Keywords Riverbed armoring · Sediment exchange process · Markov chain · Sand–gravel bed · Three Gorges Project

Introduction

The flow and sediment regime entering downstream reaches was greatly altered due to the operation of the Three Gorges Project (TGP), which led to significant channel degradation and corresponding riverbed coarsening. Especially in sand–gravel bed reaches, channel adjustments were more complicated owing to a remarkable bed material coarsening process (Xia et al. 2017; Zhou et al. 2018). In addition, it also caused a more complicated sediment exchange process (SEP) and a more random distribution among grains. Therefore, it is necessary to study the internal mechanism of riverbed armoring and sediment exchange processes in a sand–gravel bed river.

Recently, studies in the relevant literature have been classified into two major groups: riverbed armoring process, interactions among bed material, bed load and suspended load (SEP). Studies on riverbed armoring are based on the exchange process between bed material and bed load, considering the influence of riverbed erosion or deposition. On the basis of previous models (Gessler 1965; Little and Mayer 1972; Shen and Lu 1983), Ettema (1984) defined the rough layer thickness as $R = Ad_{95}$ (A is a constant; d_{95} is a characteristic grain size of surface bed material, with 95% of the sample being finer than this value) and proposed the probabilistic volume method, which was verified against flume experimental measurements and field observations. Other studies (Xu et al. 1999; Sun and Sun 2000) proposed the sediment motion probabilities considering the hiding-exposure effect and then established a multi-step model for calculating the grain size distribution (GSD) of the armor layer, which can also be used to predict the bed scour depth simultaneously. However, studies of Xu et al. (1999) showed that the active layer thickness E varied with the exchange of bed material during the coarsening process. For

✉ Junqiang Xia
xiajq@whu.edu.cn

¹ State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

the sand–gravel bed, the active layer thickness E is equal to d_{95} , and E increases continuously with the bed material coarsening. The multi-step model developed by Sun and Sun (2000) did not refer to the effects of sediment suspension and deposition on the SEP, assuming that the active layer thickness E is a constant during the process of bed coarsening, and only changed the GSD of the surface bed material and the corresponding riverbed elevation. In addition, a three-state continuous-time Markov chain model for the mixed-size sediment transport was investigated using different methods (Christina and Yang 2013; Christina and Lai 2014). For the evolution of bed material in natural rivers, Clayton and Pitlick (2008) presented a clear field evidence for the existence of coarse surface layer in a gravel bed river; several researchers (Lisle 1995; Powell et al. 2001; Singh et al. 2012; Guerit et al. 2018; Sziło and Bialik 2018; Peirce et al. 2019) observed the evolution of bed material mobility and the GSD of bed load under a range of discharges in a gravel bed river.

In terms of the SEP, earlier studies were primarily based on the two-state or three-state transition probability matrix (interactions among bed material, bed load or suspended load layer). However, the applications of three-state transition probability in sediment transport were seldom investigated in recent studies. Since the threshold probability of a single particle is the component of two-state or three-state transition probability, the rolling and suspension probabilities of uniform sediment were developed from different aspects based on forces analysis which worked out by many scholars (Cheng and Chiew 1999; Wu and Chou 2003; Elhaheem et al. 2017). Other scholars further considered the influence of hiding exposure on sediment motion and established the rolling and lifting probabilities of mixed-size sediment particles (Wu and Yang 2004; Bose and Dey 2013), and the results were verified against flume experimental data. Subsequently, Sun and Donahue (2000) proposed the two-state transition probability matrix and state occupancy probability of non-uniform sediment based on the Komogorov differential equation and the Markov theorem, and then, the fractional sediment transport formula was derived. Yang et al. (2010) developed a probabilistic model

for riverbed armoring depending on the birth–death and immigration–emigration Markov process and then obtained the piecewise function between bed load transport and scouring duration. The transition probability and exchange intensity among the four sediment motion states (static, rolling, lifting and suspension) were systematically summarized by Han (2018), and then, six probabilities of sediment exchange on the bed surface were presented; Furthermore, sediment transport formulas were established based on 12 exchange intensities of sediment motion, which are different from two common kinds of formulas (bed load transport and suspended sediment transport).

According to the analysis above, the probabilistic model is rarely used to study the SEP in a sand–gravel bed river. In this study, based on the Markov chain and inter-granular effect, the probabilistic model of SEP and riverbed armoring is developed by the approaches of mechanical analysis and probability statistics. In addition, the influences of flow and sediment regime, bed erosion and deposition on SEP have been discussed. The inter-annual bed coarsening and annual SEP are compared with the measured data from 2003 to 2009 at the Zhicheng station.

Study area

As shown in Fig. 1a, the Jingjiang Reach (JR) is located at about 102 km downstream of the TGP in the Middle Yangtze River, China, with a length of 347 km (Yu and Lu 2008; Xia et al. 2017). With the boundary at Ouchikou, the total JR is usually divided into the Upper JR, with a length of 172 km, and the Lower JR, with a length of 175 km (Xia et al. 2016). The Upper JR consists of six bends, such as Zhicheng, Jiangkou and Shashi. The river upstream of Jiangkou is composed of sand–gravel bed, with the surface bed being composed mainly of sand and gravel. In this study, the measured data are collected from the Zhicheng station, including hydrographs of discharge and suspended sediment concentration, transport rate of bed load, grain size distribution (GSD) of bed material (CWRC 2018), and the composition of bed



Fig. 1 Sketch of the Jingjiang Reach in the Middle Yangtze River

material at this station is mainly characterized by sand and gravel.

In order to investigate the process of bed erosion and deposition in the downstream reaches below the Three Gorges Dam (TGD), the natural flow and sediment regime entering the Middle Yangtze River should be analyzed in detail. The temporal variations in water volume and sediment load during flood seasons and hydrological years at Zhicheng in 1994–2017 are presented in Fig. 2a, b (Zhou et al. 2018; CWRC 2018). In terms of flood seasons, the average annual water volume (after the TGP operation) was $3030 \times 10^8 \text{ m}^3/\text{a}$, with a reduction of 10.2%, as compared with the average value of $3375 \times 10^8 \text{ m}^3/\text{a}$ during the period 1994–2002. However, the average sediment load had a remarkable reduction of 88.2%, as compared with the value before the TGP operation. In addition, the average sediment load also had a remarkable reduction of 88.4% after the TGP operation during hydrological years.

Figure 3a, b shows the evolution of bed elevation at two sections of a local reach of Zhicheng during the period 2002–2017. As shown in Fig. 1b, Jing3 and Jing4 sections in the local reach of Zhicheng are located 101.8 km and 103.6 km downstream of the TGD, respectively. Figure 3 indicates that bed scour occurred because of the remarkable reduction in sediment load entering the Middle Yangtze River after the TGP operation, and the composition of bed material was continuously coarsening, especially in the local reach of Zhicheng (Zhang et al. 2017). The measured GSD

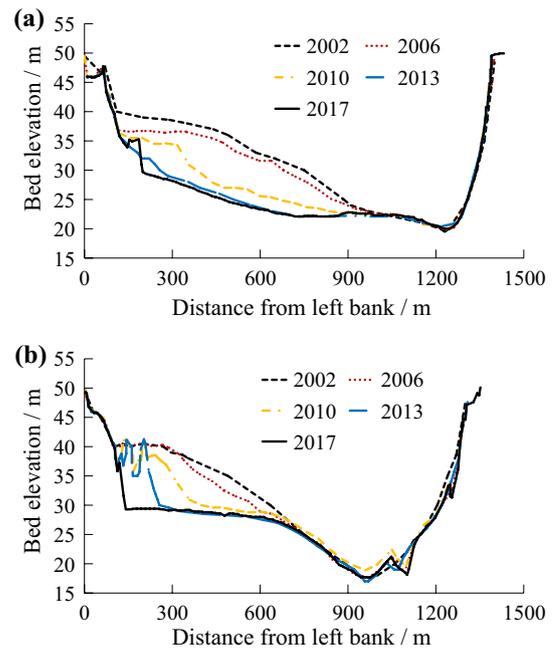


Fig. 3 Evolution of the bed elevation from 2002 to 2017 after the TGP operation at sections of **a** Jing3; **b** Jing4

and median diameter (d_{50}) for the bed material at Zhicheng from 2003 to 2017 are shown in Fig. 4a, b. In general, the median grain size at Zhicheng increased gradually, and the range in grain size became wider, but the mutation occurred

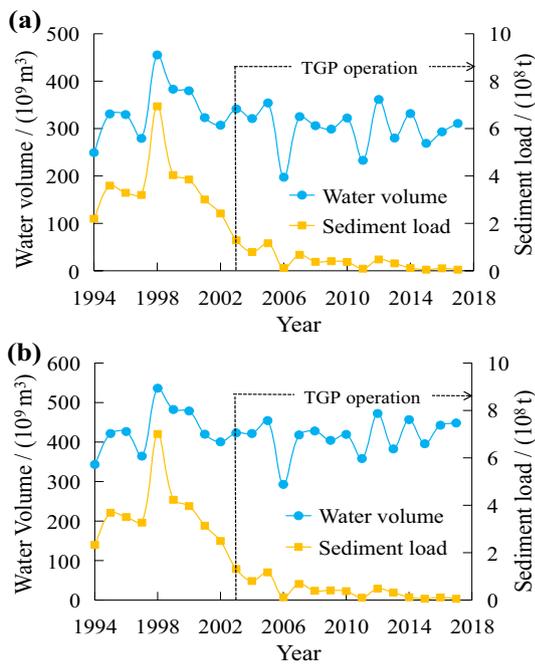


Fig. 2 Temporal variations in water volume and sediment load from 1994 to 2017 at Zhicheng during **a** flood season; **b** hydrological year

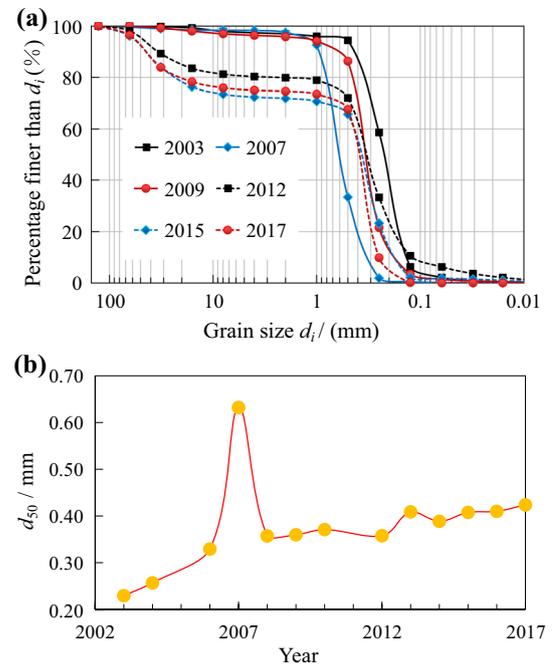


Fig. 4 Temporal variations in bed material from 2003 to 2017 at Zhicheng: **a** grain size distribution; **b** median particle size d_{50}

in 2007 due to a remarkable variation in water volume and sediment load.

As can be seen from Fig. 4a, the original grain size range of surface bed material at Zhicheng was 0.002–41 mm in 2003, with $d_{50} = 0.230$ mm; however, the particle size range of bed material was 0.031–86 mm in 2017, with $d_{50} = 0.424$ mm. Figure 4 confirms that the median diameter of bed material generally showed an increasing trend during the period 2003–2017, but it fluctuated slightly after 2010. Meanwhile, the coarsening rate of bed material in 2010–2017 was slower than that in 2003–2009.

Probabilistic model for SEP

Transition probabilities for the three-state model

As shown in Fig. 5a, b, three states (1—bed material; —bed load; and 3—suspended load) of sediment motion can be exchanged from each other under steady or unsteady flow and sediment parameters, and it is usually called the Markov stochastic process or the Markov chain.

In the literature, only the interaction between bed material and bed load or suspended load layer was generally considered, namely the two-state model (Lisle 1995; Cheng and Chiew

1999; Wu and Chou 2003; Wu and Yang 2004; Bose and Dey 2013; Li et al. 2018). Nevertheless, the two-state model is only suitable for flume experiments, uniform sediment and equilibrium sediment transport, and it has limitations when applied to natural rivers. Therefore, a three-state model needs to be developed, which includes the transition probabilities among bed material, bed load and suspended load comprehensively. It is assumed that both saltation and rolling of sediment motion are regarded as the same state 3 (Han and He 1999), and then, the three-state transition probability matrix of non-uniform sediment can be expressed as follows:

$$\epsilon_i^n = \begin{bmatrix} \epsilon_{11i}^n & \epsilon_{12i}^n & \epsilon_{13i}^n \\ \epsilon_{21i}^n & \epsilon_{22i}^n & \epsilon_{23i}^n \\ \epsilon_{31i}^n & \epsilon_{32i}^n & \epsilon_{33i}^n \end{bmatrix} \tag{1}$$

in which the subscript “i” refers to the size fraction of non-uniform sediment; ϵ_i^n = the transition matrix of different sediment motion states for the n th step; ϵ_{xyi}^n = the transition probability, which represents a single-step probability for a particle to move from the state x to the state y . Among them, $x = 1, 2$ or 3 and $y = 1, 2$ or 3 . “1,” “2” and “3,” respectively, represents the bed material, bed load and suspended load. For example, $\epsilon_{1yi}^n = \epsilon_{11i}^n + \epsilon_{12i}^n + \epsilon_{13i}^n$.

Kuai and Tsai (2016) assumed that the transition probability ϵ_{23i}^n for a particle to move from the state 2 to the state 3 is approximately equal to the suspension probability P_{Si}^n (Cheng and Chiew 1999), and the transition probability ϵ_{31i}^n for a particle to move from the state 3 to the state 1 is equal to 0. Hence, the three-state transition probability matrix ϵ_i^n for the i th size fraction in Eq. (1) is given as:

$$\epsilon_i^n = \begin{bmatrix} 1 - P_{Ti}^n & P_{Ti}^n - \epsilon_{13i}^n & \epsilon_{13i}^n \\ 1 - (P_{Ti}^n - \epsilon_{13i}^n) & (P_{Ti}^n - \epsilon_{13i}^n) - P_{Si}^n & P_{Si}^n \\ 0 & \beta \{1 - (P_{Ti}^n - \epsilon_{13i}^n)\} & 1 - \beta \{1 - (P_{Ti}^n - \epsilon_{13i}^n)\} \end{bmatrix} \tag{2}$$

where P_{Ti}^n = the total probability of sediment motion, and $P_{Ti}^n = P_{Ri}^n + P_{Li}^n$; P_{Li}^n = lifting probability; and $\beta = \epsilon_{32i}^n / \epsilon_{21i}^n$.

However, Kuai and Tsai (2016) believed that the probability for a particle to move from the state 3 to the state 1 might not be zero at any time, and Kuai’s model in Eq. (2) was not verified against field measurements. Accordingly, the modified model based on the inter-granular effect of non-uniform sediment can be written as follows:

$$\epsilon_i^n = \begin{bmatrix} 1 - P_{Ti}^n & P_{Ti}^n - P_{Si}^n & P_{Si}^n \\ 1 - P_{2i}^n & P_{Ri}^n P_{2i}^n & P_{2i}^n (1 - P_{Ri}^n) \\ 1 - P_{3i}^n & P_{3i}^n (1 - P_{Si}^n) & P_{Si}^n P_{3i}^n \end{bmatrix} \tag{3}$$

where P_{Ri}^n = the rolling probability of sediment motion; P_{Si}^n = suspension probability, and according to the assumption above, $P_{Si}^n = P_{Li}^n$; $1 - P_{2i}^n$ = the probability to stop moving from rolling state; $1 - P_{3i}^n$ = the probability to stop suspension. In Eq. (3), the diagonal values ($\epsilon_{11i}^n, \epsilon_{22i}^n, \epsilon_{33i}^n$) are

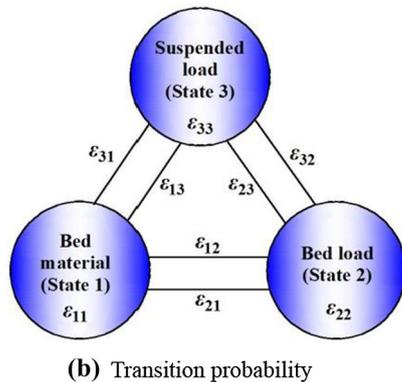
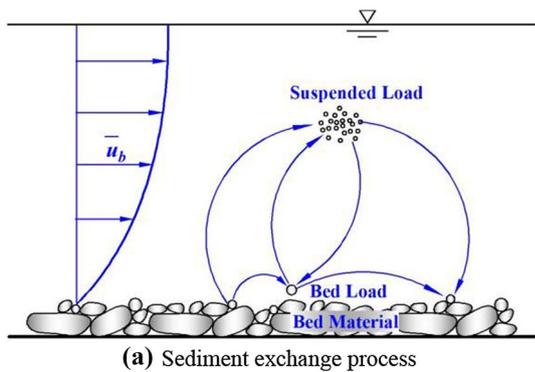


Fig. 5 Sediment exchange process and three-state transition probability in a sand–gravel bed

first determined, and the transition probabilities ($\epsilon_{21i}^n, \epsilon_{31i}^n$) in the first column are then derived. In Eq. (3), the sum of each row of the transition probability matrix equals 1, and other values can be determined according to this premise.

Threshold probabilities of sediment motion

For non-uniform sediment, the suspension probability is independent with the rolling probability due to the hiding-exposure effect among non-uniform sediment (Bialik et al. 2015). As indicated by Han and He (1999), the velocity for stopping motion (U_i) for the i th size fraction was generally smaller than the moving velocity U , and $U_i = aU \approx 6.725d_i^{0.5}$. Accordingly, the probability to stop moving from rolling state can be written as follows (Li et al. 2019):

$$1 - P_{2i}^n = \frac{1}{\sqrt{2\pi}} \int_{-\frac{U_i}{2u_*} - 2.7}^{\frac{U_i}{2u_*} - 2.7} e^{-\frac{t^2}{2}} dt \tag{4}$$

where d_i is the diameter for the i th size fraction and u_* is the shear velocity.

Moreover, the relationship between $(1 - P_{Ri}^n)$, $(1 - P_{Si}^n)$, $(1 - P_{2i}^n)$ and $(1 - P_{3i}^n)$ can be described as:

$$\begin{cases} 1 - P_{Ri}^n \neq 1 - P_{2i}^n \\ 1 - P_{Si}^n \neq 1 - P_{3i}^n \end{cases} \Rightarrow \begin{cases} P_{Ri}^n \neq P_{2i}^n \\ P_{Si}^n \neq P_{3i}^n \end{cases} \tag{5}$$

According to the studies of Han and He (1999), the relationship between $(1 - P_{2i}^n)$ and $(1 - P_{3i}^n)$ can be obtained as follows:

$$1 - P_{3i}^n = (1 - P_{Si}^n)(1 - P_{2i}^n) \tag{6}$$

Furthermore, Eq. (6) can be simplified as:

$$P_{3i}^n = P_{2i}^n + (1 - P_{2i}^n)P_{Si}^n \tag{7}$$

In addition, as indicated by Li et al. (2018), the rolling and lifting probabilities of non-uniform sediment based on the hiding-exposure effect can be written as:

$$P_{Ri}^n = \frac{1}{2} \left\{ \frac{0.21 - A}{|0.21 - A|} \sqrt{1 - \exp \left[-\left(\frac{0.46}{A} - 2.2 \right)^2 \right]} - \frac{0.135 - A}{|0.135 - A|} \sqrt{1 - \exp \left[-\left(\frac{0.295}{A} - 2.2 \right)^2 \right]} \right\} \tag{8}$$

$$P_{Li}^n = \frac{1}{2} \left\{ 1 - \frac{0.21 - A}{|0.21 - A|} \sqrt{1 - \exp \left[-\left(\frac{0.46}{A} - 2.2 \right)^2 \right]} \right\} \tag{9}$$

In Eqs. (8)–(9), A is the comprehensive coefficient, and $A = \sqrt{\xi_i \Theta_i C_L}$; ξ_i = the hiding-exposure factor of non-uniform sediment, and $\xi_i = \sigma_g^{0.25} (d_i/d_m)^{0.5}$, $\sigma_g = \sqrt{d_{84.1}/d_{15.9}}$, d_m = arithmetic mean diameter of the surface bed material; Θ_i = fractional mobility parameter or dimensionless shear stress for the i th size fraction, and $\Theta_i = \rho u_*^2 / [(\gamma_s - \gamma)d_i]$; ρ is the density of water; γ and γ_s are the specific weight of water and sediment, respectively; C_L = coefficient of uplift force, generally $C_L = 0.1$ (Li et al. 2019).

Probabilistic model for the GSD of bed material

As indicated by the relevant studies (Xu et al. 1999; Sun and Sun 2000), the equilibrium equation of sediment quantity in the active layer can be expressed by:

$$EP_{ai}^{n-1} - \Delta H^n P_{bi}^n + \Delta H^n P_{0i} = EP_{ai}^n \tag{10}$$

However, the proposed model in Eq. (10) was based on the two-state continuous-time Markov chain (the interaction between bed material and bed load) and neglected the influence of suspended load. Furthermore, the thickness of the active layer E continuously increased with the bed scour process, and this model assumed that the amount of sediment scouring from the active layer was supplemented by the original bed material in the subsurface layer (Xu et al. 1999; Sun and Sun 2000). Thus, Eq. (10) is not suitable for predicting the armoring process in natural rivers (Armanini 1995; Church and Haschenburger 2017), and the modified probabilistic model becomes:

$$E^n P_{ai}^{n-1} - \Delta H_1^n [P_{bi}^n + P_{si}^n] + \Delta H_2^n [P_{bi}^{n-1} \epsilon_{21i}^n + P_{si}^{n-1} \epsilon_{31i}^n] = E^n P_{ai}^n \tag{11}$$

where E^n = active layer thickness of surface bed material, and for the sand–gravel bed, it is related to the maximum size of sediment motion (d_{cmax}^n); P_{ai}^n = GSD of the surface bed material; P_{bi}^n = GSD of the bed load; P_{si}^n = GSD of the suspended load layer; P_{ai}^n, P_{bi}^n and P_{si}^n represent the corresponding percentages finer than d_i ; ΔH^n = total thickness of bed scour and deposition; P_{0i} = GSD of the original bed material, and $P_{ai}^0 = P_{0i}$.

In the proposed model, it is assumed that the amount of suspended load layer for the n th step is composed of three contributions: the remaining surface bed material, bed load transferred from the suspended load layer and surface bed material, as shown in Fig. 5. Therefore, regarding the surface

bed material, bed load and suspended load layers as a whole, the GSD of bed load layer P_{bi}^n for the n th step can be written as:

$$P_{bi}^n = P_{ai}^{n-1} \epsilon_{12i}^n + P_{bi}^{n-1} \epsilon_{22i}^n + P_{si}^{n-1} \epsilon_{32i}^n \quad (12)$$

Similarly, the grain size gradation of suspended load layer P_{si}^n for the n th step is:

$$P_{si}^n = P_{ai}^{n-1} \epsilon_{13i}^n + P_{bi}^{n-1} \epsilon_{23i}^n + P_{si}^{n-1} \epsilon_{33i}^n \quad (13)$$

Thickness of bed scour and deposition

According to the previous mentioned studies (Xu et al. 1999; Sun and Sun 2000), the riverbed scour depth (ΔH_1^n) for the n th step in a sand–gravel bed river can be derived as follows:

$$\Delta H_1^n = T^n E_1^n \left(\sum_{i=1}^{\max} P_{ai}^{n-1} \epsilon_{13i}^n + \sum_{i=1}^{\max} P_{ai}^{n-1} \epsilon_{12i}^n \right) \quad (14)$$

Similarly, the deposition thickness (ΔH_2^n) for the n th step is determined by:

$$\Delta H_2^n = E_2^n \left(\sum_{i=1}^{\max} P_{si}^{n-1} \epsilon_{31i}^n + \sum_{i=1}^{\max} P_{bi}^{n-1} \epsilon_{21i}^n \right) \quad (15)$$

where E_1^n = the active layer thickness of surface bed for the n th step, and $T^n E_1^n = E^n$ in Eq. (14); T^n = the relative time interval of one exchange process for the n th step, and $T^n = \Delta T / \Delta t$; ΔT is the time interval of one exchange process, and Δt is the time interval of single-step motion; and E_2^n = the coefficient of bed deposition thickness, with $E_2^n = E^n$.

In this study, if the bed scouring occurs in the SEP, then $\Delta H_1^n > 0$; otherwise, if the bed deposition occurs in the SEP,

then $\Delta H_2^n < 0$. Therefore, based on Eqs. (14) and (15), the total thickness of riverbed scour and deposition can be determined by:

$$\Delta H^n = \Delta H_1^n + \Delta H_2^n \quad (16)$$

In conclusion, substituting Eqs. (12)–(15) into Eq. (11), the probabilistic model based on the hiding-exposure effect for calculating GSD in a sand–gravel bed river can be expressed by:

$$P_{ai}^n = P_{ai}^{n-1} - N_1 [P_{bi}^n + P_{si}^n] + N_2 [P_{bi}^{n-1} \epsilon_{21i}^n + P_{si}^{n-1} \epsilon_{31i}^n] \quad (17)$$

where $N_1^n = \sum_{i=1}^{\max} P_{ai}^{n-1} \epsilon_{13i}^n + \sum_{i=1}^{\max} P_{ai}^{n-1} \epsilon_{12i}^n$ and $N_2^n = \sum_{i=1}^{\max} P_{si}^{n-1} \epsilon_{31i}^n + \sum_{i=1}^{\max} P_{bi}^{n-1} \epsilon_{21i}^n$.

Validation and discussion

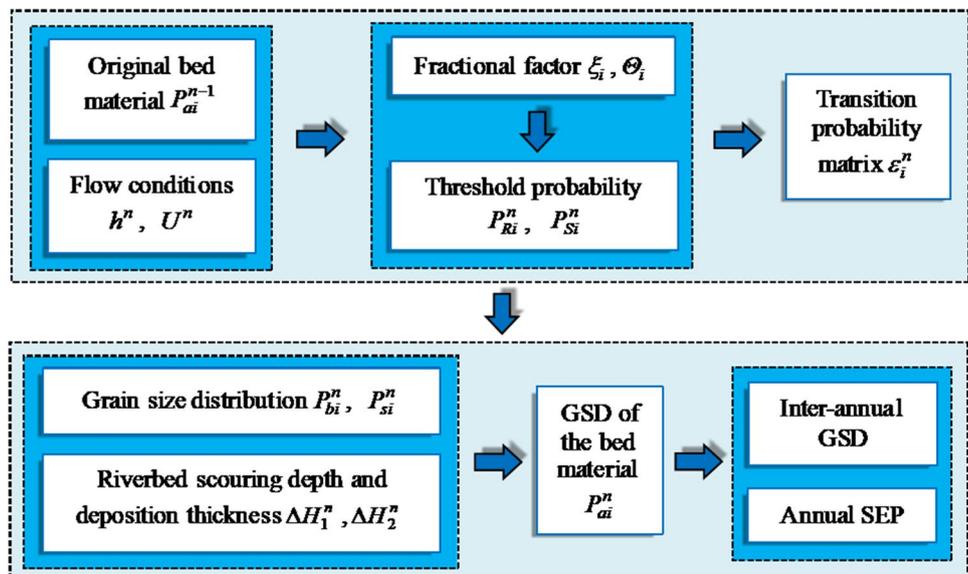
Figure 6 presents the calculation procedures and key elements of the probabilistic model based on the three-state Markov chain, and the proposed model is validated against the field measurements of grain size distribution (GSD) of bed material and bed load, annual SEP, inter-annual bed armoring process at the Zhicheng station of the Middle Yangtze River, China.

Annual SEP

Analysis of the SEP

Water and sediment conditions are changeable in the Middle Yangtze River, and the flood season is generally referred to as the period from May to October (Yu and Lu 2008; Xia et al. 2017), which leads to the SEP more complicated. In

Fig. 6 Flowchart of the proposed model based on the three-state Markov chain



order to investigate the SEP in sand–gravel bed rivers, measured data of surface bed material at Zhicheng from 2003 to 2009 were collected (the field measurements of GSD of bed load and suspended load at Zhicheng from 2010 to 2017 were incomplete in CWRC), and the temporal variation in mean monthly median particle size (d_{50}) was analyzed (the mean monthly d_{50} is the average value of several measurements in a given month), as shown in Fig. 7a–c. Through statistical analysis, the change process of the mean monthly median particle size can be normally divided into three major categories (Fig. 7): (i) equilibrium state–bed material coarsening, before the end of flood season, the SEP was in an approximate equilibrium state and then gradually coarsened with the temporal variation in the water and sediment regime after the flood season; (ii) equilibrium state–bed material fining–bed material coarsening, the SEP was in an

approximate equilibrium state from January to April, and then, the sediment deposition occurred rapidly from April to May, and finally, the bed material coarsened from May to December; (iii) bed material fining–bed material coarsening–bed material fining–bed material coarsening, this process was complicated, and the changing features of each year are quite different, no specific analysis is made.

Validation of the SEP Model

Figure 8a, b shows the comparison between the measured GSD of surface bed material at Zhicheng in 2007 and 2009 and the results calculated using the probabilistic model (OM original bed material, PM present model, MD measured data). As shown in Fig. 8a, the SEP in 2007 approximately obey the process (ii), and the surface bed material gradation on 21 March was similar to the original bed material in 18 January (equilibrium state). The sediment deposition process was completed until the beginning of the flood season, and then, the bed material coarsened gradually, as shown in Fig. 7b. Similarly, in Fig. 8b, the SEP in 2009 approximately obeys the process (iii), and the first process of sediment deposition was completed at the beginning of the flood season. Then, the surface bed material gradually coarsened and completed at the end of the flood season, as shown in Fig. 7c.

In addition, in order to illustrate the temporal variations in median grain size (d_{50}) using the probabilistic model in

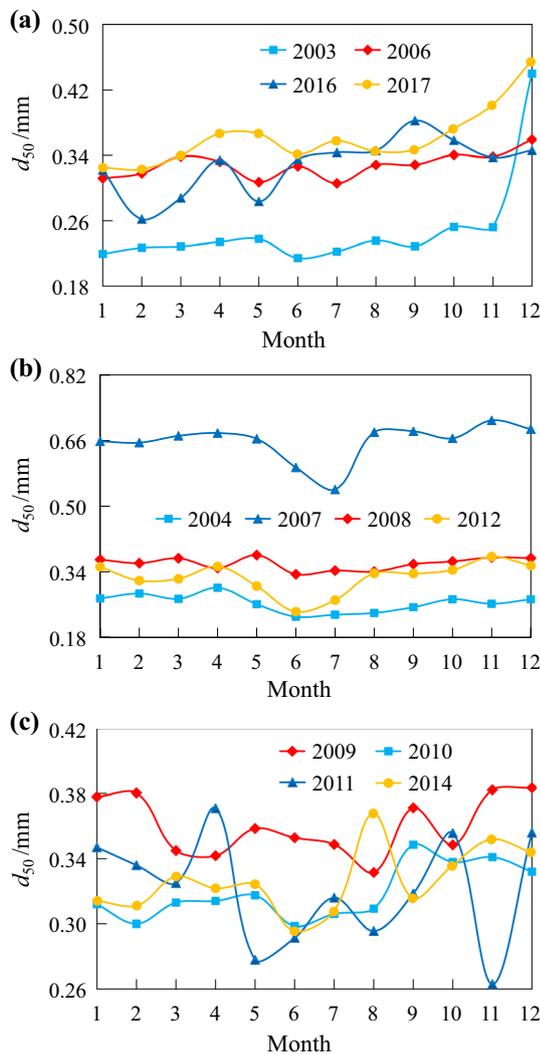


Fig. 7 Temporal variations in bed material from 2003 to 2017 at Zhicheng: **a** the variation process (i); **b** the variation process (ii); **c** the variation process (iii)

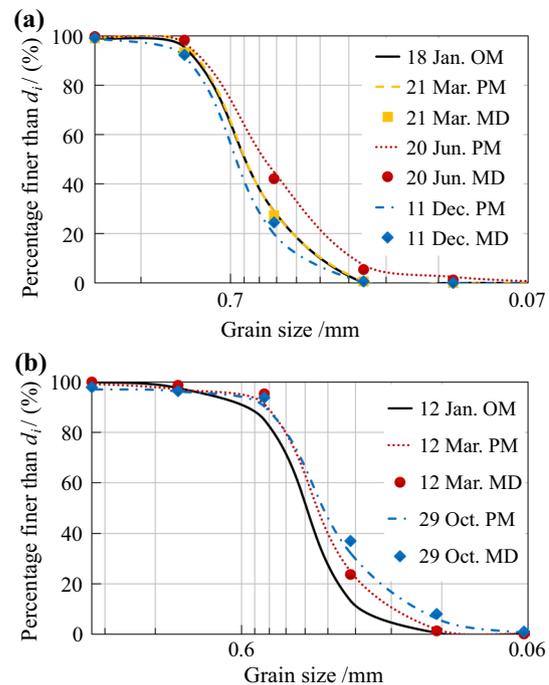


Fig. 8 Temporal variations in GSD at Zhicheng in the years of **a** 2007; **b** 2009

this study, the calculated results from 2003 to 2009 are compared with the measured data, as shown in Fig. 9a–f, and the beginning time of each year is marked in figures; other time nodes can be calculated according to the beginning time. As can be seen from Fig. 9, the annual SEP at Zhicheng was complicated in 2003–2009, and the bed erosion and deposition processes occurred alternately, especially in 2007; in 2003–2009, the calculated maximum grain size (d_{\max}) was about 0.253, 0.336, 0.356, 0.710, 0.381, 0.382 mm, respectively, and slightly different from the measured d_{\max} of 0.260, 0.301, 0.359, 0.709, 0.381, 0.383 mm; in addition, comparisons between the calculated and measured minimum grain size (d_{\min}) also had a similar result. The above analysis confirms that the annual SEP at Zhicheng responded well to the altered flow and sediment regime. Overall, the variation

trend in Fig. 9 is consistent well with the field measurements from 2003 to 2009 at Zhicheng, and the proposed model can well predicts the annual exchange process of surface bed material.

Inter-annual riverbed armoring

The inter-annual riverbed armoring process at Zhicheng is compared with the measured data from 2003 to 2009, as shown in Fig. 10 (the annual GSD of bed material is defined as the arithmetic mean value of daily GSD). It can be seen from Fig. 10 that the grain size range of surface bed material increased from 0.002 to 41 mm in 2003 to 0.016 to 128 mm in 2007; in 2003–2009, the surface bed material coarsened in different degrees, the coarsening rate in 2007–2009 was

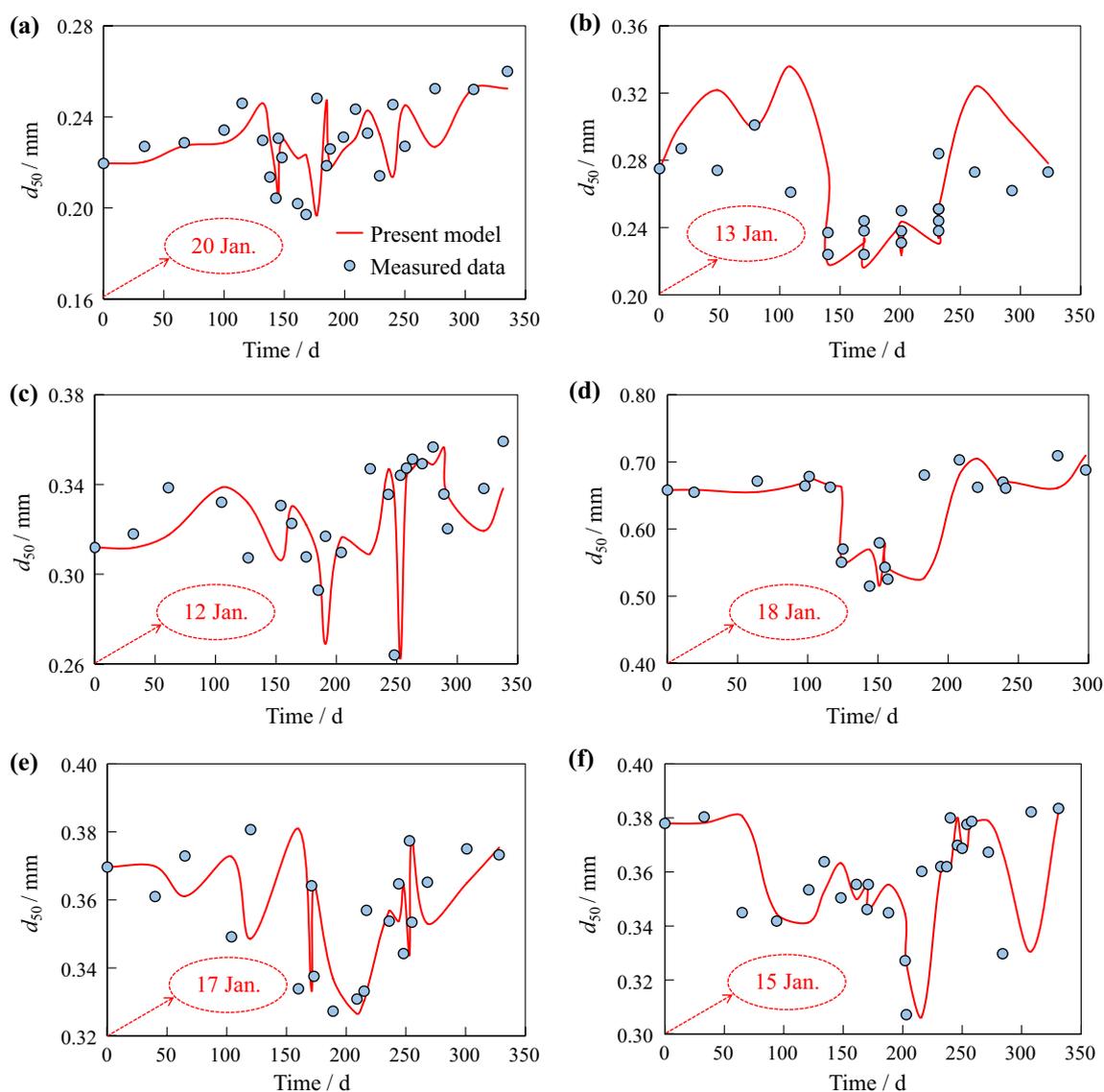


Fig. 9 Temporal variations in median grain size d_{50} at Zhicheng in the years of **a** 2003; **b** 2004; **c** 2006; **d** 2007; **e** 2008; **f** 2009

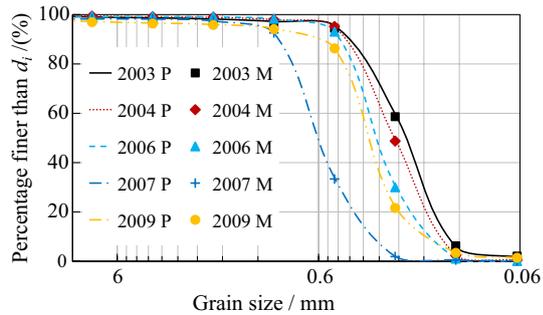


Fig. 10 Comparison of the inter-annual GSD at Zhicheng in 2003–2009

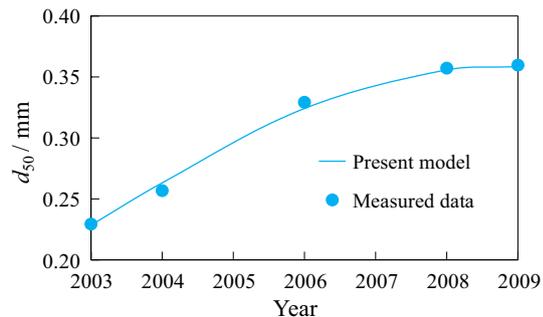


Fig. 11 Comparison of median particle size d_{50} at Zhicheng in 2003–2009

slower than that in 2003–2006, and the inter-annual SEP in sand–gravel beds can be well predicted by the probabilistic model in this study.

In addition, Fig. 11 shows the comparison results of median particle size (d_{50}) at Zhicheng from 2003 to 2009. As can be seen from Fig. 11, the median particle size of surface bed material increased from 0.23 mm in 2003 to 0.36 mm in 2009, but the mutation occurred in 2007; furthermore, the overall variation tendency of median particle size showed an upward increasing trend in Fig. 11, and this is the main reason for the continuous coarsening in the Zhicheng reach. In addition, the proposed model performs well when compared with the field measurements, and the changing trend of bed material is well consistent with the observations.

Conclusions

The flow and sediment regime entering the Middle Yangtze River has been significantly altered as a result of the recent TGP operation, which leads to the riverbed coarsened drastically. Especially in gravel–sand bed reaches, channel adjustments are more complicated owing to bed material coarsening

process. In this study, based on the Markov chain and inter-granular effect between non-uniform sediment, the probabilistic model of SEP and riverbed armoring is established by means of mechanical analysis and probability statistics. In this three-state model, the influences of flow and sediment regime, bed erosion and deposition on the SEP have all been considered, and the improved model can be used to evaluate the inter-annual bed coarsening and annual SEP in a sand–gravel bed river. On this basis, the main conclusions are as follows:

1. Bed material composition at Zhicheng obviously coarsened after the TGP operation, with the GSD curves shifts to the left, and the value ranges of grain size became wider, but the mutation occurred in 2007. In addition, the median grain size (d_{50}) of surface bed material increased from 0.23 mm in 2003 to 0.424 mm in 2017, and it showed an upward increasing trend in 2003–2017. In 2003–2009, the surface bed material coarsened in different degrees, and the coarsening rate in 2007–2009 was slower than that in 2003–2006.
2. Through statistical analysis, the change process of the monthly median particle size (d_{50}) can be normally divided into three major categories: (i) equilibrium state–bed material coarsening; (ii) equilibrium state–bed material fining–bed material coarsening; and (iii) bed material fining–bed material coarsening–bed material fining–bed material coarsening.
3. The annual SEP at Zhicheng was complicated in 2003–2009, and the bed erosion and deposition processes occurred alternately, especially in 2007; in 2003–2009, the calculated maximum grain size (d_{max}) was about 0.253, 0.336, 0.356, 0.710, 0.381, 0.382 mm, respectively, and slightly different from the measured d_{max} ; and comparisons between the calculated and measured minimum grain size (d_{min}) also had a similar result. The above analysis confirms that the annual SEP at Zhicheng responded well to the altered flow and sediment regime. Through validation, the probabilistic model proposed in this study agrees reasonably well with the field measurements at Zhicheng station, and it can be used to predict the riverbed armoring and annual SEP in a sand–gravel bed river.

Acknowledgements This work was supported by the Program of the National Key Research and Development Plan (Grant Nos. 2016YFC0402303/05) and by the National Natural Science Foundation of China (Grant Nos. 51725902; 51579186). And all data and information applied in this paper are available from the corresponding author upon request.

References

- Armanini A (1995) Non-uniform sediment transport: dynamics of the active layer. *J Hydraul Res* 33(5):611–622. <https://doi.org/10.1080/00221689509498560>
- Bialik RJ, Nikora VI, Karpiński M, Rowiński PM (2015) Diffusion of bedload particles in open-channel flows: distribution of travel times and second-order statistics of particle trajectories. *Environ Fluid Mech* 15(6):1281–1292. <https://doi.org/10.1007/s10652-015-9420-5>
- Bose SK, Dey S (2013) Sediment entrainment probability and threshold of sediment suspension: exponential-based approach. *J Hydraul Eng* 139(10):1099–1106. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0000763](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000763)
- Cheng NS, Chiew YM (1999) Analysis of initiation of sediment suspension from bed load. *J Hydraul Eng* 125(8):855–861. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1999\)125:8\(855\)](https://doi.org/10.1061/(ASCE)0733-9429(1999)125:8(855))
- Christina TM, Lai KC (2014) Three-state continuous-time Markov Chain Model for mixed-size sediment particle transport. *J Hydraul Eng* 140(9):04014047. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0000897](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000897)
- Christina TM, Yang FN (2013) Modeling bed-load transport by a three-state continuous-time Markov Chain Model. *J Hydraul Eng* 139(12):1265–1276. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0000764](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000764)
- Church M, Haschenburger JK (2017) What is the “active layer”? *Water Resour Res* 53(1):5–10. <https://doi.org/10.1002/2016WR019675>
- Clayton JA, Pitlick J (2008) Persistence of the surface texture of a gravel-bed river during a large flood. *Earth Surf Proc Land* 33(5):661–673. <https://doi.org/10.1002/esp.1567>
- CWRC (Changjiang Water Resources Commission) (2018) Analysis of channel degradation downstream of the Three Gorges Dam. Scientific Report of CWRC, Wuhan (in Chinese)
- Elhaheem M, Papanicolaou T, Tsakiris AG (2017) A probabilistic model for sediment entrainment: the role of bed irregularity. *Int J Sedim Res* 32(1):137–148. <https://doi.org/10.1016/j.ijsrc.2016.11.001>
- Ettema R (1984) Sampling armor-layer sediments. *J Hydraul Eng* 110(7):992–996. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1984\)110:7\(992\)](https://doi.org/10.1061/(ASCE)0733-9429(1984)110:7(992))
- Gessler J (1965) The beginning of bed-load movement of mixtures investigated as natural armoring in channels. Translation T-5, W. M. Keck Laboratory of Hydraulics and Water Resources, California Institute of Technology, Pasadena, CA
- Guerit L, Barrier L, Liu Y, Narteau C, Lajeunesse E, Gayer E (2018) Uniform grain-size distribution in the active layer of a shallow, gravel-bedded, braided river (the Urumqi River, China) and implications for paleo-hydrology. *Earth Surf Dyn* 6(4):1011–1021. <https://doi.org/10.5194/esurf-6-1011-2018>
- Han QW (2018) The frontier research achievements on stochastic theory of sediment transport. *J Hydraul Eng* 49(9):1040–1054 (in Chinese)
- Han QW, He MM (1999) Study on state probabilities and the ratio of bed load to suspended load. *J Hydraul Eng* 10:7–16 (in Chinese)
- Kuai KZ, Tsai CW (2016) Discrete-time Markov chain model for transport of mixed-size sediment particles under unsteady flow conditions. *J Hydrol Eng* 21(11):04016039(1–12). [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001392](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001392)
- Li JD, Sun J, Lin BL (2018) Bed-load transport rate based on the entrainment probabilities of sediment grains by rolling and lifting. *Int J Sedim Res* 33(1):126–136. <https://doi.org/10.1016/j.ijsrc.2017.12.005>
- Li LL, Zhang GG, Zhang JJ (2019) Formula of bed-load transport based on the total threshold probability. *Environ Fluid Mech* 19(2):569–581. <https://doi.org/10.1007/s10652-018-9638-0>
- Lisle TE (1995) Particle size variations between bed load and bed material in natural gravel bed channels. *Water Resour Res* 31(4):1107–1118. <https://doi.org/10.1029/94WR02526>
- Little WC, Mayer PG (1972) The role of sediment gradation on channel armoring. School of Civil Engineering, Georgia Institute of Technology, Atlanta
- Peirce S, Ashmore P, Leduc P (2019) Evolution of grain size distributions and bed mobility during hydrographs in gravel-bed braided rivers. *Earth Surf Proc Land* 44(1):304–316. <https://doi.org/10.1002/esp.4511>
- Powell DM, Reid I, Laronne JB (2001) Evolution of bed load grain size distribution with increasing flow strength and the effect of flow duration on the caliber of bed load sediment yield in ephemeral gravel bed rivers. *Water Resour Res* 37(5):1463–1474. <https://doi.org/10.1029/2000WR900342>
- Shen HW, Lu JY (1983) Development and prediction of bed armoring. *J Hydraul Eng* 109(HY4):611–629. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1983\)109:4\(611\)](https://doi.org/10.1061/(ASCE)0733-9429(1983)109:4(611))
- Singh A, Guala M, Lanzoni S, Foufoula-Georgiou E (2012) Bedform effect on the reorganization of surface and subsurface grain size distribution in gravel bedded channels. *Acta Geophys* 60(6):1607–1638. <https://doi.org/10.2478/s11600-012-0075-z>
- Sun ZL, Donahue J (2000) Statistically derived bed-load formula for any fraction of nonuniform sediment. *J Hydraul Eng* 126(2):105–111. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2000\)126:2\(105\)](https://doi.org/10.1061/(ASCE)0733-9429(2000)126:2(105))
- Sun ZL, Sun ZF (2000) Experiment and prediction of an armoring layer. *J Hydroelectr Eng* 4:40–48 (in Chinese)
- Sziło J, Bialik RJ (2018) Grain size distribution of bedload transport in a glaciated catchment (Baranowski Glacier, King George Island, Western Antarctica). *Water* 10(4):360. <https://doi.org/10.3390/w10040360>
- Wu FC, Chou YJ (2003) Rolling and lifting probabilities for sediment entrainment. *J Hydraul Eng* 129(2):110–119. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2003\)129:2\(110\)](https://doi.org/10.1061/(ASCE)0733-9429(2003)129:2(110))
- Wu FC, Yang KH (2004) Entrainment probabilities of mixed-size sediment incorporating near-bed coherent flow structures. *J Hydraul Eng* 130(12):1187–1197. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2004\)130:12\(1187\)](https://doi.org/10.1061/(ASCE)0733-9429(2004)130:12(1187))
- Xia JQ, Deng SS, Lu JY, Xu QX, Zong QL, Tan GM (2016) Dynamic channel adjustments in the Jingjiang Reach of the Middle Yangtze River. *Sci Rep Rep*. <https://doi.org/10.1038/srep22802>
- Xia JQ, Deng SS, Zhou MR, Lu JY, Xu QX (2017) Geomorphic response of the Jingjiang Reach to the Three Gorges Project operation. *Earth Surf Proc Land* 42(6):866–876. <https://doi.org/10.1002/esp.4043>
- Xu QX, Zhang XF, Tan GM (1999) Multi-step prediction modeling on riverbeds scouring and armoring. *Adv Water Sci* 10(1):42–47. <https://doi.org/10.14042/j.cnki.32.1309.1999.01.008> (in Chinese)
- Yang FG, Liu XN, Cao SY, Huang E (2010) Bed load transport rates during scouring and armoring processes. *J Mt Sci* 7(3):215–225. <https://doi.org/10.1007/s11629-010-2013-3>
- Yu WC, Lu JY (2008) Bank erosion and protection in the Yangtze River. China Water and Power Press, Beijing (in Chinese)
- Zhang W, Yang YP, Zhang MJ, Li YT, Zhu LL, You XY, Wang D, Xu JF (2017) Mechanisms of suspended sediment restoration and bed level compensation in downstream reaches of the Three Gorges Projects (TGP). *J Geogr Sci* 27(4):463–480. <https://doi.org/10.1007/s11442-017-1387-3>
- Zhou MR, Xia JQ, Deng SS, Lin FF (2018) Channel adjustments in a gravel-sand bed reach downstream of the Three Gorges Dam. *Glob Planet Change* 170:213–220. <https://doi.org/10.1016/j.gloplacha.2018.08.014>



Algorithm for modeling electromagnetic channel of seismo-ionospheric coupling (SIC) and the variations in the electron concentration

Yuriy Rapoport^{1,3,5} · Vladimir Grimalsky² · Andrzej Krankowski³ · Sergey Pulinets^{3,4} · Alla Fedorenko⁵ · Sergei Petrishchevskii¹

Received: 21 December 2018 / Accepted: 28 October 2019 / Published online: 26 November 2019

© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2019

Abstract

We present in detail the algorithm of the electrostatic–quasi-stationary–electromagnetic/MHD approximations and equivalent external sources (EQUEMES method) to develop the quasi-stationary–electromagnetic models of seismo-ionospheric coupling. The penetration of the electromagnetic field created by near-Earth alternative currents of ULF range was simulated by solving equations for the horizontal electric field components E_x , E_y of the second order with respect to the vertical coordinate z . This system of two second-order equations is derived from the system of Maxwell equations. The penetration of rather strong horizontal electric field [of order of (1–10) mV/m] to the ionospheric E and F layers has been modeled. The corresponding variations in the electron concentration in the E and lower F layers of the ionosphere reach a value of order of (1–10)%. Farther increase in these variations can be connected with the related synergetic processes. A possibility of the effective initiation of electron concentration perturbations in the unstable near-equatorial plasma in the F layer of the ionosphere by the packet of atmospheric gravity waves radiated by the near-ground source is illustrated. A good correspondence of the results obtained on the basis of this model to the data of satellite observations is shown.

Keywords Quasi-stationary · Electron concentration · Seismo-ionospheric coupling · TEC · ULF

✉ Yuriy Rapoport
yuriy.rapoport@gmail.com

Vladimir Grimalsky
v_grim@yahoo.com

Andrzej Krankowski
kand@uwm.edu.pl

Sergey Pulinets
pulse1549@gmail.com

Alla Fedorenko
aurora28@i.ua

¹ Physical Faculty, Taras Shevchenko National University of Kyiv, Kiev 03022, Ukraine

² CIICAp, Autonomous University of State Morelos (UAEM), 62209 Cuernavaca, MOR, Mexico

³ Space Radio, Diagnostics Research Centre, University of Warmia and Mazury in Olsztyn, Prawochenskiego 9, 10-720 Olsztyn, Poland

⁴ Space Research Institute of RAS (IKI RAS), Moscow, Russia

⁵ Space Research Institute National Academy of Sciences of Ukraine and State Space Agency of Ukraine, Kiev, Ukraine

Introduction

In the present time, a new model of the electromagnetic channel of seismo-ionospheric coupling (Pulinets 2009; Pulinets et al. 2000, 2015; Rapoport et al. 2004a, b, 2009) is considered. This model eliminates some of the shortcomings of existing models by the consequent unification of electrostatic, quasi-stationary and MHD approximations. Firstly, the rough approximation of an infinite conductivity in the upper ionosphere accepted in the papers (Sorokin et al. 2001; Sorokin and Hayakawa 2013; Kuo et al. 2011) is insufficient to compute the geomagnetic field-aligned current in the electrostatic approximation, as was shown in Grimalsky et al. (1999, 2003), Rapoport et al. (2014a, b). Secondly, it is necessary to mention that a real field sources always, in proper timescale, are sources of alternative non-quasi-stationary electromagnetic fields.

We applied earlier the electrostatic (Grimalskiy et al. 2003; Rapoport et al. 2014b) and quasi-stationary (Grimalskiy and Rapoport 2000) approximations for modeling of the penetration of electric fields into the lower

ionosphere from the atmosphere and electromagnetic/MHD approach for the penetration of the electromagnetic field from the lithosphere through the system “lithosphere–atmosphere–ionosphere–magnetosphere (LAIM)” (Grimalsky et al. 1999).

In the present paper, the penetration into the ionosphere of the electromagnetic fields excited by the alternative ULF seismogenic current sources localized in the atmosphere is investigated. Generally, this needs the self-consistent solving of the Maxwell equations and the equations of motion of charges in the time domain. That is, the current sources should be determined in a self-consistent manner. Here, the new algorithm is proposed that makes possible to simplify the cumbersome problem of finding the full current sources for the electrodynamic problem. Namely, the current sources can be determined on the basis of the self-consistent solving the equation for the ULF electric field within the quasi-stationary approximation and the equation for the density of the electric charge in the time domain in the atmosphere and the lower ionosphere. Then, the total electric current sources for the electromagnetic problem should be computed in the time domain using the distributions of ULF electric field in the atmosphere and lower ionosphere. The next problem is the penetration into the *E* and *F* layers of the electromagnetic fields from the already determined total current sources within the framework of the Maxwell equations. This approach is called as the method of equivalent external sources (EQUEMES method); it is presented in detail here.

We suppose that in distinction to Denisenko and Pomozev (2010), a conclusion of an effective penetration of electric fields into the ionosphere (Pulinets et al. 2000, 2015; Pulinets 2009) will be proven on the basis of the electromagnetic approach. The proposed model could be useful for a comparison between theoretical and experimental results, obtained both on satellite and on ground observatories, and for better understanding of the mechanisms of seismo-ionospheric coupling.

Another possibility of the effective initiation of electron concentration perturbations, for example in the unstable near-equatorial plasma in the *F* layer of the ionosphere, can be connected with the packet of atmospheric gravity waves (AGWs) radiated by the near-ground (seismogenic) source. The paper includes a demonstration of a very effective process of the formation of the charge plasma and neutral concentration structures in the ionosphere. We also show a good correspondence of the results obtained on the basis of this model to the data of satellite observations. Note also that the packet of AGW causes, besides plasma and neutral concentration perturbation/structures, electrostatic field perturbations as well (Rapoport et al. 2009).

There are several reasons, albeit related, for which the synergistic aspect of the penetration of the electromagnetic field and the induction of plasma perturbations in the ionosphere

must be investigated in the context of the research to which this article is devoted, and within the framework of appropriate approximations and developed models.

1. On the one hand, it is already known that there are a number of hydrodynamic, plasma and hydrodynamic–plasma instabilities associated with AGW and the electric field at different altitudes and in different regions of the atmosphere–ionosphere system. In particular, there are: (1) thermal instability of AGW in the lower atmosphere (at the altitudes 0–15 km) (Kotsarenko et al. 1994); we note that AGWs can be connected to an electric field through the conduction of charged aerosols, which increase hydration (Pulinets and Boyarchuk 2005; Pulinets et al. 2015), which in turn is associated with a possible instability of AGWs. (2) Rapoport et al. (2004a, b) noted that it is worthwhile to consider the possibility of thermal instability (Gurevich 1978) in the region of the mesosphere (at the altitudes 55–65 km) in the presence of electric fields of seismogenic origin. Moreover, preliminary evaluations show that such an instability similar to negative differential conductivity in semiconductors is possible, but we will not stay on this in the present paper and separate work will be dedicated to this question. (3) At the E region (at altitude of order of 100 km), electric-heating AGW instability is possible (Sorokin et al. 1998). (4) In the ionospheric F region, Rayleigh–Taylor or Perkins instability, depending on the longitude of the ionospheric region, may lead to the formation of large-scale plasma bubbles.
2. On the other hand, as will be shown below, even with significant values of the surface current source of the order of $1 \mu\text{A}/\text{m}^2$, the corresponding variations in the concentration of electrons in the F region of the ionosphere reach a maximum of about (1–10)%, which is only partly enough to explain, the observed before the most powerful earthquakes, TEC perturbations. The last in some case are in the range of 10% (Oikonomou et al. 2016), while in other case they reach a value of order of tens of percent (Pulinets 2011; Pulinets et al. 2011, 2014, 2015). For the formation of appropriate structures, developing plasma instabilities is necessary, for example Rayleigh–Taylor or Perkins instability, which can lead to the formation of plasma bubbles, just accompanied by disturbances of the plasma concentration of the value of order of dozens percent. However, even if there is a corresponding instability in the ionosphere, its development requires the presence of a seeding factor in the form of an electric field of the order of a few units up to ten mV/m (Kuo et al. 2011) or AGW disturbances with speeds of the order of tens to hundred m/s.
3. In accordance with Oikonomou et al. (2016), the perturbations with periods 20 and 2–5 min. of characteristic

for gravity wave branch AGW are observed in TEC few days before the strongest earthquakes, such as in Nepal ($M=7.8$) and Chile ($M=8.3$) in 2015. Emphasize also that besides the perturbations in TEC, the same strongest seismogenic sources can cause the disturbances in VLF signals propagating in or radiated from the waveguide “earth–ionosphere” (WGEI) as well (Stangl et al. 2011; Sanchez-Dulcet et al. 2015). Then, both the strongest seismogenic and meteorological and other geophysical sources (cyclones, typhoons, tsunamis, etc.) cause oscillations of the VLF signals in the range of gravity waves (Nina and Čadež’ 2013; Rozhnoi et al. 2014a, b, 2015; Solovieva et al. 2015). These observable effects prove, again, a possible importance of AGW/gravity wave branch as an agent of the plasma concentration variations, altogether with electric/electromagnetic field.

4. Hypothetically, the sequence of the above (and other possible) hydrodynamic–photochemical–plasma instabilities at different heights in the atmosphere–ionosphere system can form an active “synergistic” channel in the form of a chain of connected “active lenses” for distribution over an area with powerful seismogenic or meteorological sources in the lower atmosphere, down to the ionosphere. This article does not aim to develop such a hypothesis, to which a special work should be devoted. However, we believe that a synergistic approach is needed to explain the observed seismogenic disturbances in the ionosphere, including variations of TEC of the order from ten to tens of percent (Pulinets 2011; Pulinets et al. 2011, 2014, 2015; Nenovski et al. 2015; Oikonomou et al. 2016), and the corresponding instabilities can develop at different altitudes, although such instabilities should be considered, generally speaking, within the framework of a self-consistent approach accounting for an integrated plasma–hydrodynamic processes in atmosphere–ionosphere system. Ultimately, the corresponding electrodynamic perturbations can be described as ones excited by effective current sources, which for a real system should be determined on the basis of an appropriate integrated/self-consistent approach. Nevertheless, in this article, a self-consistent definition of current sources is not done. But we can, within the framework of the model of given external currents adopted in our article, investigate ionospheric disturbances depending on the altitude, where such an external current is located in the system “atmosphere–ionosphere.” Note also that in accordance with Nenovski et al. (2015), before a very strong earthquake, local perturbations in TEC can be connected with the processes occurring in the lower ionosphere, namely at the altitudes below 160 km. Therefore, not only TEC as a whole, but a thin structure of the altitude dependence of the perturbations in plasma concentration is important. We do not answer the question

in this article, how such external currents are formed, based on the development of the above-mentioned instabilities. But we answer another question: how does the efficiency of penetration of the electromagnetic field into the ionosphere depend on the altitude localization of the corresponding effective current sources, presented in our model by external currents (Vainshtein 1988; Collin 1991), without consideration of the processes leading to the formation of such currents.

Thus, the following questions are included into this paper:

1. Electrostatic–quasi-stationary–electromagnetic approximations and equivalent external sources (EQUEMES method) to develop the quasi-stationary–electromagnetic algorithms of seismo-ionospheric coupling;
2. The penetration of the electromagnetic field created by near-Earth alternative currents of ULF range $\omega < 10 \text{ s}^{-1}$; a possibility of the variations in the electron concentration in the E and lower F layers of the ionosphere reaching a value of order from ten to few tens of percent, as it follows from the data of observations (Pulinets 2011; Pulinets et al. 2011, 2014, 2015; Nenovski et al. 2015; Oikonomou et al. 2016).
3. The study of the dependence of the efficiency of penetration of the electromagnetic field into the ionosphere on the altitude of the localization of the corresponding effective current sources, which in this context are considered as external currents.
4. A possibility of the initiation of electron concentration perturbations by means of AGW packet radiated from the near-ground gas/heat source; in particular, the demonstration of the correspondence between the results, obtained using the model (Rapoport et al. 2009), and the data of the observations of the seismogenic structures in the unstable ionospheric F -layer plasma, initiated by the AGW of the lithospheric origin, and detected on the board of the satellite “Atmosphere Explorer-E” (Fedorenko et al. 2005).
5. Finally, the discussion and conclusions will be presented.

Algorithms and modeling of the electromagnetic field penetration from the lithosphere/lower atmosphere to the ionosphere and perturbations in the electron concentration

The idea of the proposed approach is to extend the scope of the purely electrostatic method, which is currently used in the modeling of seismo-ionospheric coupling (Sorokin et al. 2001; Sorokin and Hayakawa 2013; Denisenko and

Pomozov 2010). The need for this and hence for the development of a new method is connected mainly with the following factors. (1) An increase in the conductivity with altitude creates negligibly small perturbations of electrostatic fields arising at the bottom, growing up with heights; (2) there is a necessity to account for the finite values of currents (in the directions, both longitudinal and transverse, respectively, to the geomagnetic field direction) for the upper ionosphere and even the magnetosphere, which is impossible under an approximation of the infinite longitudinal—along the geomagnetic field—conductivity (Sorokin et al. 2001; Sorokin and Hayakawa 2013; Denisenko and Pomozov 2010); (3) then the dynamic nature of real sources of fields and currents in the atmosphere and ionosphere should be accounted for.

Let us consider algorithms for modeling of the penetration of electromagnetic fields from the lower atmosphere or the lithosphere–atmosphere interface into the ionosphere.

The goal is to investigate the penetration of the electric field into the ionosphere created by the electric current sources of ULF range at the frequencies $\omega = (0–10) \text{ s}^{-1}$, localized near the Earth's surface in the atmosphere. Thus, the problem is generally dynamic and needs to solve directly the Maxwell equations within the time domain, due to a wide electromagnetic spectrum. But in the time domain the Maxwell equations should be, generally speaking, added by the differential equations for the motion of the complex anisotropic medium, in particular the E and F layers of the ionosphere, where the values of characteristic frequencies vary essentially. In another words, the anisotropic medium possesses the frequency dispersion there and can be considered simpler, namely in the frequency domain. So, the subdivision of such a cumbersome problem into the sequence of simpler problems may be proposed. Such a subdivision seems useful for a better physical interpretation of simulation results and also would simplify the simulations. Namely, in the atmosphere and lower ionosphere the 3D quasi-stationary problem is considered in the time domain, where the values of the total electric current are obtained in the self-consistent manner in the whole ULF range.

For a penetration of the electric fields into the ionosphere E and F layers at relatively low-frequency part of ULF frequency range, $\omega < 0.1 \text{ s}^{-1}$, the solution of the quasi-stationary problem becomes important. At these frequencies, the quasi-stationary approximation is valid in the lower and middle atmosphere in the first approximation, because the corresponding electromagnetic wavelengths are $\lambda > 1000 \text{ km}$. The quasi-stationary approximation is valid in this case, because all the characteristic frequencies, i.e., plasma, cyclotron and collision ($\nu_{i,e}$) ones both for electrons and for ions, in the lower ionosphere are much higher than the frequency in ULF range. Thus, the conductivity tensor does not depend on frequencies, i.e., the dispersion is absent. At the higher frequency, part of the ULF range dispersion is practically absent at the lower atmosphere

and ionosphere, where $\omega < \nu_i$, but it is present and should be accounted for at the E and F regions of the ionosphere, where $\omega \geq \nu_i$ (Alperovich and Fedorov 2007). In any cases, to determine the perturbations in the electron concentration and, finally, TEC, the MHD approximation or even the solution of the Maxwell equations could be applied as the next step/approximation used in the EQUEMES method.

On the other hand, coming to the frequency domain, the obtained values of the electric currents are used as the full current sources to solve the Maxwell equations, or the equivalent set of equations for the horizontal components of the variable electric field, to compute the electromagnetic fields in E and F layers of the ionosphere. This approach yields good results at the frequencies $\omega \geq 0.1 \text{ s}^{-1}$, as our simulations have shown. Proper Fourier components of the current sources should be used there. At these frequencies, the corresponding electromagnetic wavelengths λ are comparable or smaller than the typical spatial scales of the ionosphere $\lambda < 1000 \text{ km}$. This approach gives a possibility to use the simple stable numerical algorithm called the elimination method with 3-diagonal matrices/matrix sweep method (Samarskii 2001; Samarskii and Nikolaev 1989), or the bypassing method.

Also, the problem of the penetration of the MHD wave(s) from the magnetosphere into the ionosphere, with the proper boundary conditions on the ionosphere–magnetosphere boundary, can be realized using the algorithm similar to the EQUEMES one.

EQUEMES algorithm

The proposed EQUEMES algorithm includes the following three consequent steps.

Step 1, Electrostatics

Initially, the atmosphere ($0 < z < L_i = 60 \text{ km}$) is characterized by the unperturbed conductivity $\sigma_0(z)$. The ionosphere ($L_i < z < L_z = 120 \text{ km}$) is characterized by the anisotropic unperturbed conductivity $\hat{\sigma}_0(z)$. The determination of the elements of this conductivity tensor is described in “Appendix 1”. In the modeling presented below, the three-component plasma model is used.

The initial electric field φ_0 , E_0 , and the initial charge density ρ_0 are calculated from:

$$0 < z < L_i : \frac{d}{dz} \left(\sigma_{00}(z) \cdot \frac{d\varphi_0}{dz} \right) = 0; \quad (1a)$$

$$\rho_0 \equiv -\frac{1}{4\pi} \frac{d^2\varphi_0}{dz^2}; \quad E_0 \equiv -\frac{d\varphi_0}{dz};$$

$$L_i < z < L_z : \frac{d}{dz} \left((\sigma_0)_{33} \cdot \frac{d\varphi_0}{dz} \right) = 0; \quad \sigma_{00}(L_i) = (\sigma_0)_{33}(L_i). \quad (2)$$

The boundary conditions are:

$$\varphi_0(z = 0) = U_0; \quad \varphi_0(z = L_z) = 0. \tag{3}$$

where the ionospheric potential, by the order of value, is $U_0 \approx 3 \times 10^5 V \approx 900$ abs. units (Sedunov et al. 1991).

Step 2, Quasi-stationary model

Consider the following 3D problem: within the atmosphere ($0 < z < L_i = 60$ km) both the perturbations of the conductivity $\delta\sigma(\vec{r})$ and the external (with respect to conduction) current density (Vainshtein 1988; Collin 1991) \vec{j}^{ext} exist.

$$\vec{E}^{grad} = \vec{E}_0 - \vec{\nabla}\varphi; \vec{j} = \sigma(\vec{r})\vec{E}^{grad}; \tag{4}$$

$$\sigma(\vec{r}) \equiv \sigma_{00}(z) + \delta\sigma(\vec{r}).$$

$\vec{E}_0 = E_0\vec{e}_z$ is the initial (unperturbed) field for unperturbed conductivity $\hat{\sigma}_{00}(z)$, φ is a perturbation that should be computed. For the electric field, the equation is:

$$\text{div}\vec{E}^{grad} = 4\pi\rho. \tag{5}$$

Here ρ is the total density of charge. Equation for ρ is:

$$\frac{\partial\rho}{\partial t} + \text{div}\vec{j} + \text{div}\vec{j}^{ext} = 0. \tag{6}$$

We use here the notation:

$$\frac{\partial\rho^{ext}}{\partial t} \equiv -\text{div}\vec{j}^{ext} \tag{7}$$

Therefore, the set of equations is:

$$\Delta\varphi = -4\pi(\rho - \rho_0) \equiv -4\pi \cdot \delta\rho; \tag{8}$$

$$\frac{\partial\delta\rho}{\partial t} + \frac{\rho}{\tau_M} = \frac{\partial\rho^{ext}}{\partial t} - \vec{E}^{grad} \cdot \vec{\nabla}\sigma; \quad \vec{E}^{grad} \equiv \vec{E}_0 - \vec{\nabla}\varphi; \quad \tau_M = \frac{1}{4\pi\sigma}. \tag{9}$$

Note that the system of Eqs. (6)–(9), in fact, determines the connection between external current source \vec{j}^{ext} and the perturbations of the conductivity $\delta\sigma(\vec{r})$, as effective “source terms” on the one hand and perturbation of the charge density $\delta\rho$ and electric potential φ on the other one. Repeating that the initial electric potential and field φ_0, E_0 , and the initial charge density ρ_0 are calculated from:

$$\frac{d}{dz} \left((\sigma_0)_{33}(z) \cdot \frac{d\varphi_0}{dz} \right) = 0 (0 < z < L_i); \tag{1b}$$

$$\rho_0 \equiv -\frac{1}{4\pi} \frac{d^2\varphi_0}{dz^2}; \quad E_0 \equiv -\frac{d\varphi_0}{dz};$$

As it is seen from (7), the presence of quasi-static external current is equivalent to the presence of the external

non-stationary charge density. The set of Eqs. (7) and (8) is valid at $0 < z < L_i \approx 60$ km, at the atmosphere. The effective source of perturbations is $\text{div}\vec{j}^{ext}$.

Within the ionosphere $L_i < z < L_z = 120$ km, the conductivity $\hat{\sigma}$ is anisotropic. The following equation for φ is used for this region in the case $\vec{j}^{ext} = 0; \partial\rho/\partial t = 0; \delta\sigma = 0$:

$$\text{div}(\hat{\sigma}_0 \cdot \vec{\nabla}\varphi) = 0. \tag{10}$$

$\hat{\sigma}_0 \equiv \hat{\sigma}_0(z)$ is anisotropic unperturbed ionospheric conductivity.

In the presence of the external current source, $\text{div}\vec{j}^{ext} \neq 0$, one can get for the ionosphere, instead of (10):

$$\text{div}(\hat{\sigma}_0 \cdot \vec{\nabla}\varphi) = \text{div}\vec{j}^{ext}. \tag{11}$$

In this case, the perturbations of the anisotropic conductivity $\delta\hat{\sigma}(\vec{r})$ should be included into \vec{j}^{ext} : $\vec{j}^{ext} = \delta\hat{\sigma}\vec{E}$.

The boundary conditions for the perturbation of the potential are:

$$\varphi(z = 0) = 0; \quad \varphi(z = L_z) = 0. \tag{12}$$

So, the density of charge ρ is used within the atmosphere only, whereas Eq. (11) for the potential φ are used within the whole region.

Equations (8, 7), (11) for φ are solved by the double Fourier transform $(x, y) \sim \exp(-ik_x x - ik_y y) \cdot \Phi$:

$$- \left(\sigma_{11}k_x^2 + \sigma_{22}k_y^2 - ik_y \frac{d\sigma_{23}}{dz} + ik_x \frac{d\sigma_{13}}{dz} \right) \Phi \tag{13}$$

$$- 2ik_x \sigma_{13} \frac{d\Phi}{dz} + \frac{d}{dz} \left(\sigma_{33} \frac{d\Phi}{dz} \right) = F; \quad (L_i \leq z < L_z);$$

$$\frac{d^2\Phi}{dz^2} - \left(k_x^2 + k_y^2 \right) \Phi + \frac{1}{\sigma_{00}} \frac{d\sigma_{00}}{dz} \Phi = \frac{1}{\sigma_{00}} F; \quad (0 < z < L_i). \tag{14}$$

In (13), (14), $k_{x,y}$ are the wavenumbers of the corresponding Fourier transform components, Φ is amplitude of the corresponding potential Fourier mode, F is the corresponding amplitude of the Fourier transform of $\text{div}\vec{j}^{ext}$. Equation (14) is equivalent in the adopted approximation to the following equation:

$$\frac{d^2\Phi}{dz^2} - \left(k_x^2 + k_y^2 \right) \Phi = -4\pi(\delta\rho)^{FT}; \quad (0 < z < L_i)$$

Upper index “FT” in the right hand of this equation is the corresponding amplitude of the Fourier transform of the corresponding value. The boundary conditions for the Fourier components are:

$$\Phi(z = 0) = \Phi(z = L_z) = 0. \tag{15}$$

Then, Eqs. (13), (14) with z only are approximated by finite differences and are solved by the factorization. Equation (9) for ρ is solved by finite differences:

$$\frac{\delta\rho^{p+1} - \delta\rho^p}{\tau} + \frac{\rho^{p+1}}{\tau_M} = -\text{div}(\vec{j}^{\text{ext}}) - \left(E_z \frac{d\sigma}{dz}\right)^{p+1}. \quad (16)$$

Note that the methods proposed in the present paper in this and the following paragraphs and in “Appendix 2” are based on the combined analytical–numerical approaches. Such an approach is dictated in fact by the subject of the investigation—electrostatic, quasi-static, and (dynamic) electromagnetic fields penetrating from the lower atmosphere/lithosphere into the magnetized ionospheric plasma possessing inhomogeneity and gyrotropy, and corresponding perturbations in plasma concentration. The numerical algorithm is based on the spectrum finite-difference method (Samarskii 2001; Samarskii and Nikolaev 1989). This analytical–numerical approach including spectrum finite-difference algorithms has proven its great effectiveness for different kinds of problems on the linear and nonlinear wave processes in the media of different physical nature with the characteristic time and spatial scales which differ from each other in many orders of values. This concerns, in particular, our modeling the electromagnetic (Grimalsky et al. 1999), electrostatic-photochemistry (Grimalsky et al. 2003; Rapoport et al. 2004a, 2006) and AGW (Rapoport et al. 2004b, 2017) coupling in the LAIM system, nonlinear vortex planetary electromagnetic wave perturbations in the ionosphere (Rapoport et al. 2012a, 2014c), and also nonlinear wave processes in the ferrite films (Rapoport et al. 2005), controllable nonlinear active metamaterial gyrotropic waveguides (Rapoport et al. 2014b), hyperbolic layered structures (Boardman et al. 2017), strongly nonlinear isotropic (Rapoport et al. 2014e) and hyperbolic (Rapoport et al. 2012c) field concentrators, graphene metamaterials (Boardman et al. 2019), etc. Note also that in all above-mentioned cases, Fourier transform in the form of fast Fourier transform [in the present paper in particular as double complex fast Fourier transform (Press et al. 1997)] has been used successfully for the very effective spectral field presentation.

The calculated profiles of ULF electric field should be used to compute the total current sources for the electromagnetic problem of the penetration of the electromagnetic field into the ionosphere E and F layers at the Step 3.

Step 3, Electromagnetics

As mentioned above, the solving of the general problem of the penetration of the electromagnetic field excited by the near-Earth’s current sources into the ionosphere can be subdivided into several sequential steps. From the mathematical point of view, this is based on a representation of the

electric field as the sum of the potential, or gradient, part, Eq. (4), and the solenoidal one (Vainshtein 1988; Collin 1991): $\vec{E} = \vec{E}^{\text{grad}} + \vec{E}_c$. This method was used for excitation of resonators by external current sources.

The potential electric field has been calculated at the Step 2, from the scalar potential φ : $\vec{E}^{\text{grad}} = -\vec{\nabla}\varphi$. At the final step, it is possible to find the solenoidal electromagnetic field from the set of the Maxwell equations:

$$\begin{aligned} [\vec{\nabla} \times \vec{H}] &= \frac{1}{c} \frac{\partial \vec{E}_c}{\partial t} + \frac{1}{c} \left(\frac{\partial \vec{E}^{\text{grad}}}{\partial t} + 4\pi(\vec{j} + \vec{j}^{\text{ext}}) \right) \equiv \frac{1}{c} \frac{\partial \vec{D}_c}{\partial t} + \frac{4\pi}{c} \vec{j}^{\text{tot}}; \\ \vec{j}^{\text{tot}} &\equiv \vec{j}^{\text{ext}} + \frac{1}{4\pi} \frac{\partial \vec{D}^{\text{grad}}}{\partial t}; \\ \frac{\partial \vec{D}_c}{\partial t} &\equiv \frac{\partial \vec{E}_c}{\partial t} + 4\pi\hat{\sigma}\vec{E}_c; \\ \frac{\partial \vec{D}^{\text{grad}}}{\partial t} &\equiv \frac{\partial \vec{E}^{\text{grad}}}{\partial t} + 4\pi\hat{\sigma}\vec{E}^{\text{grad}}; \\ [\vec{\nabla} \times \vec{E}_c] &= -\frac{1}{c} \frac{\partial \vec{H}}{\partial t}; \\ \text{div} \vec{D}_c &= 0; \quad \text{div} \vec{H} = 0. \end{aligned} \quad (17)$$

Here \vec{D}_c is the electric induction associated with the solenoidal electric field. The equation $\text{div} \vec{D}_c = 0$ is the condition for the separation of the solenoidal electric field (Vainshtein 1988; Collin 1991). It means that the solenoidal electric field \vec{E}_c is excited by the total alternative electric current source \vec{j}^{tot} (Eq. 17), while the alternative part of the gradient field \vec{E}^{grad} is excited by the alternative external charge ρ^{ext} (Eqs. 8, 9). The boundary conditions at the Earth’s surface are zero tangential components of the electric field: $z=0$: $\vec{E}_{c\tau} = 0$. At the upper boundary, the radiation conditions would be the most adequate ones.

It is seen from Eq. (17) that the electric current source for the electrodynamic problem is \vec{j}^{tot} , which includes both the external density of the electric current and one excited by the quasi-stationary electric field. The value of \vec{j}^{tot} has been computed accounting for the solution of the quasi-stationary problem, Eqs. (4)–(12). Therefore, it is possible to solve the non-quasi-stationary problem with the solenoidal electric field \vec{E}_c and \vec{H} .

The middle and upper ionosphere possesses the effective anisotropic permittivity $\hat{\sigma}(\omega)$ that depends on frequency in ULF range, whereas in the lower atmosphere the conductivity does not depend on frequency because of great collision frequencies of ions compared with the frequency ω (Sorokin et al. 2001; Sorokin and Hayakawa 2013). The solution of the Maxwell equations can be realized, in particular in the frequency domain. The expression for the electric induction \vec{D}_c is within the frequency domain $\vec{D}_c = \hat{\varepsilon}(\omega) \cdot \vec{E}_c$, $\hat{\varepsilon}(\omega) = 1 - 4\pi i\hat{\sigma}(\omega)/\omega$; the dependence of the Fourier components is $\sim \exp(i\omega t)$. Within the frequency domain $\omega \neq 0$, the dynamic equations with curls are important, whereas the equations with divergences

should be satisfied automatically and may be used only for checking the numerical results.

The radiation boundary conditions in the magnetosphere $z = 600\text{--}800$ km also are formulated in a simpler manner within the spectral domain than in the time domain. So, the spectral components of the electric current source \vec{j}^{tot} should be used. They have been obtained within the time domain from the quasi-stationary problem, Step 2, so then the Fourier transform is used. Note that the transition from the dynamic Maxwell equations, $\omega \neq 0$, to the stationary case, $\omega \rightarrow 0$, is non-trivial in a media with the anisotropic conductivity that depends on frequency (Alexandrov et al. 1984).

The problem is with the upper boundary conditions for \vec{H} , because the magnetic field penetrates highly $z > 200$ km. It is possible to use the radiation conditions at the magnetosphere $z = L_f \geq 600$ km: at $z > L_f$ there are the outgoing waves only.

Algorithm for the TEC modeling

Now we will restrict ourselves with “non-self-consistent” approximation for TEC modeling. In other words, TEC will be determined in a supposition that electric field is already known, in our case from the modeling based on an algorithm described in “EQUIMES algorithm” section. In case of sufficiently powerful processes, this approximation is expected to give a remarkable inaccuracy or even to be failed. It is possible to develop a self-consistent theory in further investigations.

Computation of TEC “in the approximation for ambipolar diffusion”/TEC in F2 layer

Neglecting the thermodiffusion, one can write equations of motions for ions and electrons in the form (Bryunelli and Namgaladze 1988):

$$\begin{aligned}
 N_j m_j \frac{d\vec{V}_j}{dt} &= -\vec{\nabla} p_j + N_j m_j \vec{g} + (q_j N_j) \left(\vec{E} + \frac{1}{c} [\vec{V}_j \times \vec{H}] \right) \\
 &\quad - \sum_{k(j \neq k)} N_j m_j \nu_{jk} (\vec{V}_j - \vec{V}_k); \tag{18} \\
 \frac{\partial N_j}{\partial t} + \vec{\nabla} (m_j N_j \vec{V}_j) &= (P_j - L_j) m_j;
 \end{aligned}$$

where the index $j = e, i$ corresponds to electrons and ions, respectively, $q_{e,i} = \mp e$, $k = e, i, n$, index n corresponds to neutral particles in the ionosphere, $p_j = N_j k_B T_j$, $p_j, m_j, T_j, \vec{V}_j, \vec{E}, \vec{H}, c, \vec{g}, P_j, L_j, e$, denote a pressure, mass, temperature in energetic units, velocity, electric and magnetic fields, light speed, free-fall acceleration, production

and loss rates and charge of electron, respectively, ν_{jk} is the frequency of collisions between corresponding particles, and k_B is Boltzmann constant.

For the F2 layer, what can be applied approximately in the range of altitudes (200–500) km (Bryunelli and Namgaladze 1988), the following equation could be used to determine $N_e \approx N_i = N$:

$$\begin{aligned}
 \frac{\partial N}{\partial t} &= Q - L - \text{div} (N \vec{V}_n^{\parallel}) - \text{div} (N \vec{V}_i^{\perp}) \\
 &\quad + \text{div} \left(\frac{D_a}{k_B T_P} [\vec{\nabla} (p_e + p_i) - N m_i \vec{g}^{\parallel}] \right). \tag{19}
 \end{aligned}$$

In (19), $D_a = k_B T_P / (m_i \nu_{in} + m_e \nu_{en})$, $T_P = T_i + T_e$, while (Bryunelli and Namgaladze 1988) $N \vec{V}_i^{\parallel} = N \vec{V}_n^{\parallel} - (D_a / 2k_B T_i) [\vec{\nabla} (p_e + p_i) - N m_i \vec{g}^{\parallel}]$, $V_i^{\perp} = -(c/H^2)(\vec{H} \times \vec{E})$; indices “ \parallel ”, “ \perp ” denote components parallel and perpendicular to magnetic field, respectively; production and loss rates are equal to (Bryunelli and Namgaladze 1988) $Q = q_O$, $L = \beta N$, respectively; and q_O , β are the rate of ionization of atomic oxygen and linear coefficient of losses, respectively.

In the approximation of ambipolar diffusion along a field line and transverse electromagnetic drift and accounting for the vertical transport, an equation for electron concentration takes the form (Bryunelli and Namgaladze 1988):

$$\frac{\partial N}{\partial t} = \frac{\partial}{\partial Z} \left(D_a \sin^2 I \cdot \left(\frac{\partial N}{\partial Z} + \frac{N}{T_P} \frac{\partial T_P}{\partial Z} + \frac{N}{H_p} \right) - wN \right) - \beta N + q. \tag{20}$$

In (20), $w = (V_{n\parallel})_Z + (V_{i\perp})_Z$, I -magnetic field inclination, H_p is plasma scale high. Finally, N is integrated in the region $Z_1 \leq Z \leq Z_2$, $Z_1 = 200$ km, $Z_2 = 500$ km. Corresponding to Eq. (20), upper and lower boundary conditions for N are presented in “Boundary conditions for electron concentration” section by relations (24) and (25a, 25b, 25c), respectively.

In this paper, the simplified algorithm is realized numerically for the estimations of the modulation of the electron concentration (see “The estimation of the modulation of the electron concentration: dynamic mechanism of modulation, Penetration of electromagnetic and electrostatic field into the ionosphere from the lithosphere–atmosphere boundary or lower atmosphere” sections). The comprehensive realization of the proposed new algorithm will need the special paper.

Algorithm for computation of TEC including E and F layers

We present also to show a perspective, more detailed algorithm for computing TEC, with inclusion into consideration both E and F layers. To account for these layers altogether, consider the equation of motions derived from the first equations from (18) (Kendall and Pickering 1967):

$$N(m_i v_{in} + m_e v_{en}) (\vec{V}_i - \vec{V}_e) = -\vec{\nabla} (p_e + p_i) + Nm_i \vec{g} + (m_e v_{en} / e) \vec{J} + (1/c) (\vec{J} \times \vec{H}); \tag{21}$$

where $\vec{J} = eN(\vec{V}_i - \vec{V}_e)$. Using the relations (Rapoport et al. 2012a, 2014a see Eq. (4) in this paper) $\vec{J} = \sigma_{\parallel} \vec{E}_{\parallel} + \sigma_P \vec{E}_{\perp} + \sigma_H (\vec{h}_0 \times \vec{E}_{\perp})$, $(\vec{h}_0 \times \vec{J}) = \sigma_P (\vec{h}_0 \times \vec{E}_{\perp}) - \sigma_H \vec{E}_{\perp}$, $\vec{h}_0 = \vec{H}_0 / H_0$ [in the linear approximation, used in the present model, where \vec{H}_0 is geomagnetic field (Rapoport et al. 2014a)], and (20), one can present (19) in the form:

$$\begin{aligned} \partial N / \partial t &= Q - L - \text{div} (N \cdot (\vec{v}_n + \vec{v}_{DTg} + \vec{v}_{DEH})); \\ \vec{v}_{DTg} &= -D_a (\vec{\nabla} N / N + \vec{\nabla} T_P / T_P) + D_a (1/H_i) \vec{g}_0; \\ \vec{v}_{DEH} &= \frac{1}{N(m_i v_{in} + m_e v_{en})} \left(\frac{m_e v_{en}}{e} (\sigma_{\parallel} \vec{E}_{\parallel} + \sigma_P \vec{E}_{\perp} + \sigma_H [\vec{h}_0 \times \vec{E}_{\perp}]) \right. \\ &\quad \left. + \frac{H_0}{c} (\sigma_H \vec{E}_{\perp} - \sigma_P [\vec{h}_0 \times \vec{E}_{\perp}]) \right); \vec{g}_0 = \frac{\vec{g}}{g}; \quad H_i = \frac{k_B T_P}{m_i g}. \end{aligned} \tag{22}$$

In accordance with Gurevich and Shvartzburg (1973), the relations for ion rates of production and losses, valid both for *E* and *F* layers of the ionosphere, are:

$$\begin{aligned} Q &= q_{iO_2} + q_{iO} + q_{iNO}; \quad L = (\alpha_1 N_{NO^+} + \alpha_2 N_{O_2^+}) N; \\ N &= N_{O^+} + N_{O_2^+} + N_{NO^+}; \\ N_{NO^+} &\approx \frac{q'_{iO} \beta_1 N_{N_2}}{\alpha_1 N (\beta_1 N_{N_2} + \beta_2 N_{O_2})} + \frac{q'_{iNO}}{\alpha_1 N}; \\ N_{O_2^+} &\approx \frac{q'_{iO_2} + q'_{iO} \beta_2 N_{O_2} (\beta_1 N_{N_2} + \beta_2 N_{O_2})^{-1}}{\alpha_2 N} \end{aligned} \tag{23a}$$

$$q'_{iO} = q_{iO} + p_1 N_O; \quad q'_{iO_2} = q'_{iO_2} + p_2 N_{O_2}; \tag{23b}$$

$$q'_{iNO} = p_3 N_O; \quad q_{iO_2} + q_{iO} + q_{iNO} = q'_{iO_2} + q'_{iO} q'_{iNO}. \tag{23c}$$

In (23), denotations like “ N_{N_2} ”, “ N_{NO^+} ” mean concentrations of corresponding neutral and ion components, the sense of the photochemical coefficients included into (23) is revealed from the following photochemical reactions and corresponding balance equations (Gurevich and Shvartzburg 1973):

$$\begin{aligned} N_2 + O^+ &\rightarrow NO^+ + N; O_2 + O^+ \rightarrow O_2^+ + O; \\ \frac{\partial N_{O_2^+}}{\partial t} &= q'_{iO_2} + \beta_2 N_{O_2} N_{O^+} - \alpha_2 N_{O_2^+} N \\ \frac{\partial N_{O^+}}{\partial t} &= q'_{iO} - \beta_1 N_{N_2} N_{O^+} - \beta_2 N_{O^+} N_{O_2} \\ \frac{\partial N_{NO^+}}{\partial t} &= q'_{iNO} + \beta_1 N_{N_2} N_{O^+} - \alpha_1 N_{NO^+} N. \end{aligned} \tag{23d}$$

Primes with production ion rates $q'_{iO_2, iO, iNO}$, included into (23b–d) correspond to the redefined effective coefficients of ionizations of corresponding atoms $q_{iO_2, iO, iNO}$. In accordance with Gurevich (Gurevich and Shvartzburg 1973), this redefinition (23c) includes fast reactions of ions N_2^+ with molecules O_2, O , namely $N_2^+ + O \rightarrow N_2 + O^+$, $N_2^+ + O_2 \rightarrow N_2 + O_2^+$, $N_2^+ + O \rightarrow N + NO^+$, characterizing by coefficients $\gamma_{1,2,3}$, as it is seen from the equation (Gurevich and Shvartzburg 1973) $\partial N_{N_2^+} / \partial t = q'_{iN_2} - \gamma_1 N_O N_{N_2^+} - \gamma_2 N_{O_2} N_{N_2^+} - \gamma_3 N_O N_{N_2^+}$, while the values $p_{1,2,3}$ included into (23c) are determined as follows: $p_k = \gamma_k q_{iN_2} / [(\gamma_1 + \gamma_3) N_O + \gamma_2 N_{O_2}]$, $k = 1, 2, 3$. Note that (23b) describe the steady-state solution (without accounting drift for) of the system (23d).

Therefore, all values included in Eq. (22) are described. Equation (22) determines the number of ions/electrons and allows to determine TEC after integration of N in the range of altitudes from $Z = Z_1 \sim 100$ km to $Z = Z_2 \sim 500$ km including both *E* and *F* layers. Corresponding boundary conditions are described in “Boundary conditions for electron concentration” section. (Upper and lower boundary conditions are presented by relations (24) and (25a, 25b, 25c), respectively).

Boundary conditions for electron concentration

In the approximation of prevailed vertical ion transport at upper altitude $Z = Z_2 \sim 500$ km of the ranges used for the integration of N for the determination of TEC, the following upper boundary condition for the electron (ion) concentration is formulated for a given vertical flow $\Phi|_Z = [NV_{iz}]_Z$ of ions, for Eq. (20) (Bryunelli and Namgaladze 1988):

$$\left[\frac{\partial N}{\partial Z} + \left(\frac{1}{H_p} + \frac{1}{T_p} \frac{\partial T_p}{\partial Z} - \frac{w}{D_a \sin^2 I} \right) N \right]_{Z=Z_2} = - \frac{\Phi}{D_a \sin^2 I} \Big|_{Z=Z_2} \tag{24}$$

In the similar way, the upper boundary condition at $Z = Z_2 \sim 500$ km for Eq. (22), in the limiting case of magnetized ions and electrons (Bryunelli and Namgaladze 1988), can be used in the form (24) as well.

The lower boundary condition at $Z = Z_1$, for the given electron concentration $N = N_{01}$, has the form

$$N|_{Z=Z_1} = N_{01}. \tag{25a}$$

For Eq. (20), $Z = Z_1 = 200$ km (*F* layer), the lower boundary conditions, relation (25a), accounting for the steady-state solution of (23d), can be presented in the form (Gurevich and Shvartzburg 1973):

$$N|_{Z=Z_1=200 \text{ km}} = N_{01} \approx N_{O^+} \approx (\beta_1 N_{N_2} + \beta_2 N_{O_2})^{-1} q'_{iO}. \tag{25b}$$

For Eq. (22), $Z = Z_1 = 80$ km (lower E layer), the lower boundary conditions, relation (25a), accounting for the steady-state solution of (23d), reduces to (Gurevich and Shvartzburg 1973):

$$N|_{Z=Z_1=80\text{ km}} = N_{01} \approx N_{O_2^+} + N_{NO^+} \approx \left\{ \left(q'_{io}/\alpha_2 \right) + \left(q'_{iNo}/\alpha_1 \right) + (\beta_1 N_{N_2} + \beta_2 N_{O_2})^{-1} [(\beta_1 N_{N_2}/\alpha_1) + (\beta_2 N_{O_2}/\alpha_2)] q'_{io} \right\}^{1/2}. \tag{25c}$$

The estimation of the modulation of the electron concentration: dynamic mechanism of modulation

The simplified estimation is presented of modulation of the electron concentration in the ionosphere plasma due to the penetration of the ULF electric field from below, considering the general case of the inclined geomagnetic field \vec{H}_0 . The continuity equations for electrons and ions are:

$$n_i \approx n_e \approx n; \quad \frac{\partial n}{\partial t} + \vec{\nabla} \cdot (n\vec{v}_e) \approx 0; \quad \frac{\partial n}{\partial t} + \vec{\nabla} \cdot (n\vec{v}_i) \approx 0. \tag{26}$$

where $v_{e,i}$ are velocities of electrons and ions that depend on the variable electric field \vec{E} . From (26), the following relation can be written down for the perturbation of the electron concentration \tilde{n} (it is assumed that $|\tilde{n}| < 0.3n_0$, where n_0 is the equilibrium concentration):

$$\frac{1}{n_0} \frac{\partial \tilde{n}}{\partial t} \approx -\vec{\nabla} \cdot \vec{v}_e. \tag{27}$$

The coordinate frame x', y', z' is aligned along the geomagnetic field \vec{H}_0 , and the relation between the coordinates x, y, z and x', y', z' is:

$$x' = x \cos \theta + z \sin \theta; \quad y' = y; \quad z' = z \cos \theta - x \sin \theta. \tag{28}$$

Here θ is the angle between the geomagnetic field \vec{H}_0 and vertical direction. In this coordinate frame, one can get:

$$\begin{aligned} \vec{\nabla} \cdot \vec{v}_e &\equiv \frac{\partial v_{ex'}}{\partial x'} + \frac{\partial v_{ey'}}{\partial y'} + \frac{\partial v_{ez'}}{\partial z'}; \\ v_{ex'} &= -\frac{e}{m_e \omega_{Be}^2 + (v_e + i\omega)^2} [(v_e + i\omega)\tilde{E}_{x'} - \omega_{Be}\tilde{E}_{y'}]; \\ v_{ey'} &= -\frac{e}{m_e \omega_{Be}^2 + (v_e + i\omega)^2} [(v_e + i\omega)\tilde{E}_{y'} + \omega_{Be}\tilde{E}_{x'}]; \\ v_{ez'} &= -\frac{e}{m_e(v_e + i\omega)} \tilde{E}_{z'}; \\ \tilde{E}_{x'} &= \tilde{E}_x \cos \theta + \tilde{E}_z \sin \theta; \quad \tilde{E}_{y'} = \tilde{E}_y; \\ \tilde{E}_{z'} &= \tilde{E}_z \cos \theta - \tilde{E}_x \sin \theta; \quad \tilde{E}_{x,y,z} \sim \exp(i\omega t). \end{aligned} \tag{29}$$

For the electrons within the ionosphere F layer, the following relations between the circular frequency ω , the collision frequency ν_e , and the cyclotron frequency $\omega_{Be} \equiv eH_0/m_e c$ are valid: $\omega \sim 0.1 \text{ s}^{-1} \ll \nu_e \sim 10^3 \text{ s}^{-1} \ll \omega_{Be} \sim 5 \times 10^6 \text{ s}^{-1}$.

Finally, using (27)–(30), the following estimation for the perturbation of the concentration is:

$$\frac{|\tilde{n}|}{n_0} \approx \frac{|\vec{\nabla} \cdot \vec{v}_e|}{\omega}. \tag{30}$$

Note that only the direct dynamic mechanism of modulation of the concentration has been considered, and the storage effects are not considered here, like heating of the electron–ion plasma. As shown in Bryunelli and Namgaladze (1988), the storage effects can strengthen the modulation of the electron concentration under the values of the electric field obtained in our simulations.

Results of modeling

Penetration of electromagnetic and electrostatic fields into the ionosphere from the lithosphere–atmosphere boundary or lower atmosphere

The penetration of the electromagnetic field created by near-Earth alternative currents of ULF range $\omega \leq 10 \text{ s}^{-1}$ was simulated in the case of the harmonic electric current source $\sim \exp(i\omega t)$ localized near the Earth’s surface.

Equation (17) can be represented as:

$$\begin{aligned} [\vec{\nabla} \times \vec{H}] &= \frac{1}{c} (i\omega \vec{E} + 4\pi \hat{\sigma}(\omega, z) \vec{E}) \\ &+ \frac{4\pi}{c} \vec{j}^{\text{tot}}(x, y, z); \quad [\vec{\nabla} \times \vec{E}] = -\frac{i\omega}{c} \vec{H}. \end{aligned} \tag{17a}$$

Then, Eq. (17a) has been rewritten as a set of equations for the horizontal electric field components E_x, E_y of the second order with respect to the vertical coordinate z (see Eqs. (39), (40) in “Appendix 2”). These equations have been solved numerically by finite differences. The Fourier transform was applied with respect to the coordinates x, y . The solid Earth was assumed as highly conductive. Within the magnetosphere $z = 800$ km, the boundary conditions of the absence of wave reflection, i.e., only outgoing waves present, were applied (see relations (41), (42) in “Appendix 2”). This is equivalent to the condition that some linear combination between $E_{x,y}$ and their derivatives $dE_{x,y}/dz$ is equal to zero. Some details of the derivation and the equations used for modeling as well as corresponding upper boundary condition are presented in “Appendix 2”. The exciting density of the total current j_z is directed upwards and has the form:

$$j_z = j_0 \cdot \exp\left\{-\left(\frac{x - 0.5L_x}{x_0}\right)^2 - \left(\frac{y - 0.5L_y}{y_0}\right)^2\right\} \cdot \exp\left\{-\left[\frac{z - z_1}{z_0}\right]^2\right\}, z \geq 0.$$

In the electrostatic simulations, Eqs. (7), (8) for the electric potential have been applied. The boundary condition for the electric potential $\varphi = 0$ is given in the magnetosphere at $L_z = 1000$ km. The results change slightly when the boundary condition has been applied at $L_z > 120$ km. The region of simulations in the horizontal plane is the same as for the dynamic simulations, i.e., $L_x = L_y = 10,000$ km. The exciting current is similar to one used in the dynamic model. The penetration of the electrostatic field is poor at the altitudes $z > 80$ km. The conductivity at the Earth's surface is $\sim 10^{-3} \text{ s}^{-1}$ without the Radon release and second order higher under the release.

In numerous papers, e.g., (Kuo et al. 2011), the penetration of the electrostatic field from the Earth's surface into the ionosphere seems much more effective than in our simulations. But in those papers the approximation of the infinite conductivity σ along the magnetic field lines was used. This is equivalent to zero value of $E| = 0$. It was assumed that the value of the electric field potential is transferred *upwards* from the lower boundary of the ionosphere to above. But it is physically impossible to use the infinite values of any component of the conductivity and to transfer the values of the electric potential from one boundary to another one, because the correct electrostatic problem should be added by the boundary conditions for the electric potential at all the boundaries, where conductivity jumps from the finite to infinite values. The authors, in fact, used the problem with initial conditions for the electrostatics. Really the boundary conditions should be used. If one formally uses the infinite conductivity within some region, the conditions of the continuity of the electric potential and the continuity of the normal component of the electric induction should be satisfied at the junction between the domains of the different conductivities. But again the electric potential should be computed as a result of the solution of the new electrostatic problem which includes the boundary conditions, both at the lower boundary and at the upper one. Such a consideration, as far as we know, is absent in the papers where the approximation of the infinite conductivity had been used, such as in (Kuo et al. 2011). Then in the paper (Kuo et al. 2011), the electrostatic potential and, respectively, electric field are continued inside the region where conductivity is supposed to be infinite, what is incorrect, to our opinion. In distinction to this, there are also papers (Denisenko et al. 2008, 2013; Grimalskiy et al. 2003) with exploration of the electrostatic approximation, where special boundary conditions are derived at the boundary between the lower layer with a finite conductivity and upper layer (upper ionosphere/magnetosphere) with a conductivity, which is supposed to be infinite. Note that in the models used in above-mentioned

papers (Denisenko et al. 2008, 2013; Grimalskiy et al. 2003), (1) the electric field is determined in the region with the finite conductivity only and (2) neither electric field nor current does not penetrate from the lower region, where the sources (of electric field or external current) are placed, into an upper space plasma region, where the conductivity of the ionosphere/magnetosphere is considered infinite. In the models presented in the papers (Grimalskiy et al. 2003) and (Denisenko et al. 2008, 2013), the electrostatic potential/field is presented as a sum of Fourier modes in a horizontal direction(s), while only an inhomogeneity in the vertical direction is included for the atmosphere–ionosphere plasma. The amplitudes are depended on the vertical coordinate. In the papers mentioned above, the upper boundary conditions for the Fourier amplitudes have, generally, the same form, namely the sum of the each amplitude and its vertical derivative, with proper coefficients equal to zero at the upper boundary of the computational region. For high enough upper boundary, these boundary conditions reduce, in fact, to the zero values of the potential (Grimalskiy et al. 2003) or its vertical derivative (Denisenko et al. 2008, 2013). It is possible to say that the upper boundary conditions in Grimalskiy et al. (2003) and Denisenko et al. (2008, 2013) correspond to the minimal and maximal possible penetration of the horizontal field components into the ionosphere, respectively. Nevertheless, in both cases, it is supposed that the field/potential does not penetrate into the ionosphere at the altitudes, larger than one, where upper boundary condition is formulated. In our opinion, the electrostatic upper boundary condition is still ambiguous. In distinction to this, the dynamic upper boundary condition, namely the condition of a radiation, is quite unambiguous, because it is dictated by the causality principle, absent unfortunately in the electrostatic approximation.

Therefore, from our point of view, more correct approach is dynamic one, with the finite conductivity in the ionosphere. The final goal is to estimate the induced variations of the electron concentrations at different altitudes $z \geq 200$ km. In the electrodynamics simulations, these variations are proportional to ω^{-1} . Thus, at lower frequencies it is possible to expect to get higher variations. But the gap between the frequency domain simulations with periods $T < 60$ s and the quasi-stationary simulations with the characteristic times $T > 10$ min remains not to be filled, due to necessity to realize the electrodynamic simulations in the time domain. To provide this, a special algorithm should be realized, which is planned for the further research.

The illustration of the ionospheric parameters used for the computations, results of the penetration of electromagnetic field, generated by the external (seismogenic) near-ground current source with a magnitude of maximum of order of $1 \mu\text{A/m}^2$ (Kuo et al. 2011) and effects in the ionosphere, in

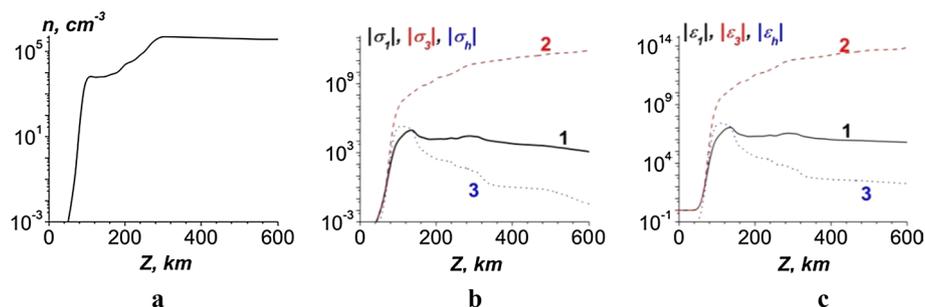


Fig. 1 The altitude dependences of the set of parameters and corresponding characteristics of the ionosphere used in the farther modeling; part **a** is the electron concentration; part **b** are the absolute values of the elements of the conductivity tensor in the system of coordinates, connected with the geomagnetic field; $\sigma_{1,3,h}$ are the corresponding diagonal components of the conductivity tensor in the directions, transverse and parallel to the geomagnetic field and non-

diagonal component, respectively; **c** are the absolute values of the dielectric permittivity tensor in the system of coordinates, connected with the geomagnetic field; $\epsilon_{1,3,h}$ are corresponding diagonal components in the directions, transverse and parallel to the geomagnetic field and non-diagonal component, respectively. Curves 1, 2 and 3 correspond to $|\sigma_1|$, $|\sigma_3|$ and $|\sigma_h|$, respectively (**b**) and $|\epsilon_1|$, $|\epsilon_3|$ and $|\epsilon_h|$, respectively (**c**)

particular the variations in the electron concentration, are presented in Figs. 1, 2, 3, 4, 5, respectively.

In particular, in Fig. 1, the altitude dependences of the set of parameters and corresponding characteristics of the ionosphere are used for the modeling (Al'pert 1972; Alperovich and Fedorov 2007; Kelley 2009; Schunk and Nagy 2010; Jursa 1985).

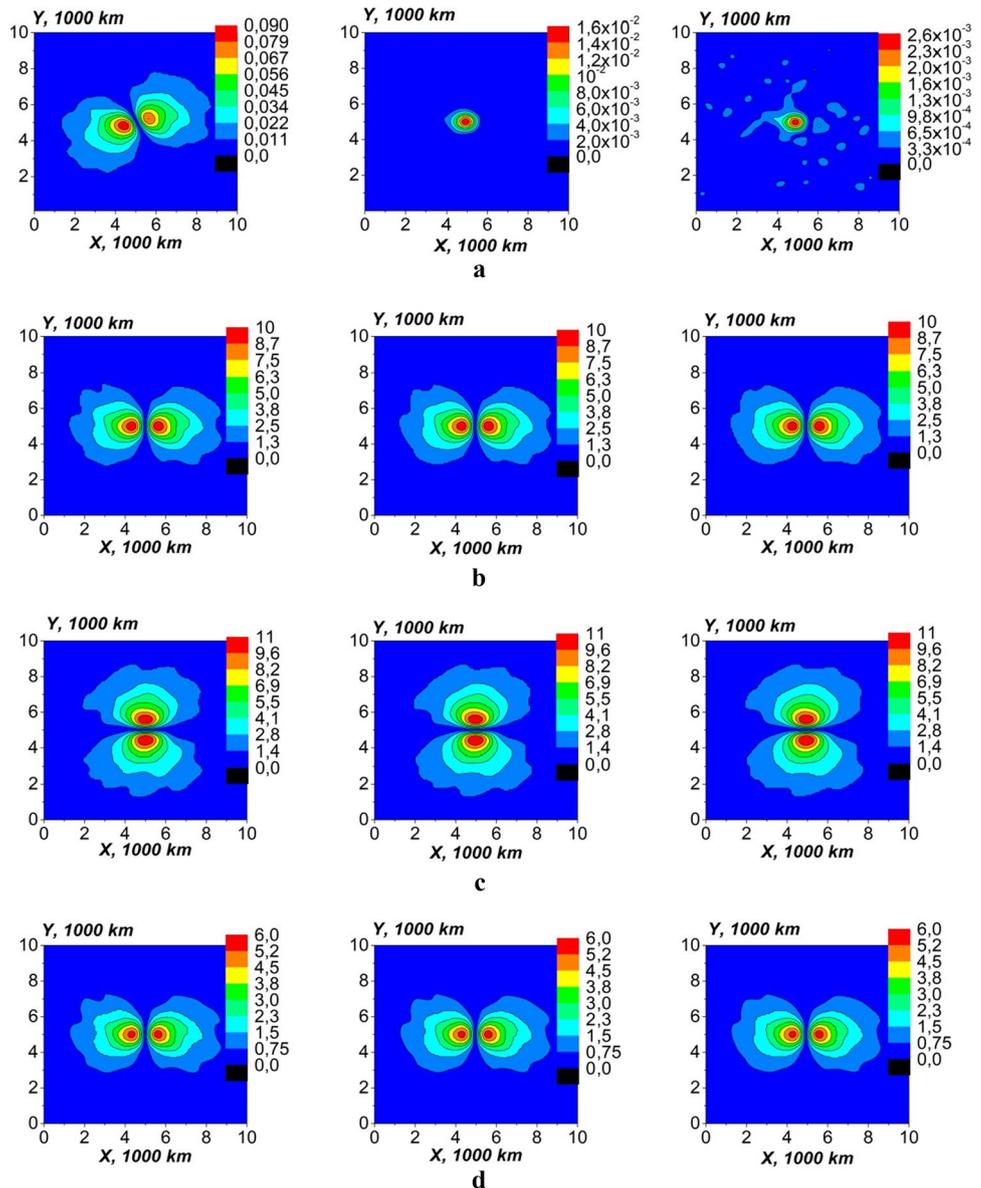
At the altitudes $z \geq 150$ km, there are two maxima of $|E_x|$, $|E_y|$ (Figs. 2, 4, 5). This is due to the inclination of the geomagnetic field and is valid, in particular, for $\theta = 60^\circ$. At the frequencies $\omega \leq 0.2$ s⁻¹, the estimated variations of n are $> 10\%$. The results of simulations for electromagnetic field penetration from the lower atmosphere/lithosphere–atmosphere boundary layer to the ionosphere depend weakly on the frequency ω , when the frequency is chosen as 0.05 s⁻¹ $< \omega < 1$ s⁻¹. At smaller frequencies $\omega \leq 0.05$ s⁻¹, the period of the oscillations is big $T = 2\pi/\omega > 120$ s, and the simulations within the time domain seem more adequate, as well as estimations of the perturbations of the electron concentration.

Spatial distributions of the relative perturbations in the electron concentration $|\tilde{n}|/n_0$ (a) and the absolute values of all the electromagnetic field components, $|E_x|$, $|E_y|$, $|E_z|$ (mV/m), are shown in Fig. 2. The exciting density of the total current j_z is directed upwards and has the form: $j_z = j_0 \cdot \exp\{-(x - 0.5L_x)/x_0\}^2 - ((y - 0.5L_y)/y_0)^2\} \cdot \exp\{-[(z - z_1)/z_0]^2\}$, $z \geq 0$. Here $0 \leq x \leq L_x$, $0 \leq y \leq L_y$ is the region of calculations in the horizontal plane. The frequency of the external current source is $\omega = 0.1$ s⁻¹. The altitude of the current maximum is $z_1 = 0$. The horizontal and vertical scales of the spatial distributions of external current are $x_0 = y_0 = 500$ km and $z_0 = 10$ km, respectively. In Fig. 3, dependences of (a) the maximal values of relative variations of the electron concentration at the three different altitudes

in the ionosphere on the altitude of the maximum of external current source z_1 and (b) the absolute values of the electromagnetic fields components on the altitude z at the ionosphere are shown. The frequency of the external current source is equal to $\omega = 0.1$ s⁻¹. Figure 4 illustrates spatial distributions of the relative perturbation of the electron concentration and the absolute values of the electromagnetic field components for the ionospheric altitude $z = 200$ km, different values of external current frequency and different values of the altitude of the external current maximum. Simulations presented in Figs. 2, 3, 4 correspond to the angle between the geomagnetic field direction and vertical axis z , equal to $\theta = 30^\circ$. In Fig. 5, the special distributions of the relative perturbation of the electron concentration and the absolute values of the electromagnetic field components, $|E_y|$, are shown for the frequencies $\omega = 0.2$ s⁻¹, 0.1 s⁻¹ and 0.05 s⁻¹ and $\theta = 60^\circ$.

As it is seen from the comparison of Figs. 2 and 5, for $\theta = 30^\circ$ and 60° , corresponding electric field components are of the same order of magnitude, while corresponding values of relative perturbations of the electron concentration in the ionosphere practically coincide with each other. For example, for the ionospheric altitude $z = 200$ km and current source frequency $\omega = 0.1$ s⁻¹, this is illustrated by the central pictures in Fig. 2, panels b and c and the central picture in Fig. 5, panels b and c for the corresponding components $|E_x|$ and $|E_y|$, and, for the relative perturbations in the electron concentration in plasma, by the central pictures in panel a, Fig. 2, and in panel a, Fig. 5, respectively. Note that, due to increasing the angle θ between the geomagnetic field \vec{H}_0 , which lies in the xz plane, and the vertical direction, from $\theta = 30^\circ$ to 60° , the asymmetry in the x direction in the values of $|\tilde{n}|/n_0$, $|E_x|$ and $|E_z|$ increases, respectively, as it is follows from the comparison between the central pictures in the panels a

Fig. 2 Spatial distributions of the relative perturbation of the electron concentration $\delta n/n_0$ (a) and the absolute values of the electromagnetic field components, $|E_x|$, $|E_y|$, $|E_z|$ (mV/m) (b, c, d, respectively); frequency of the external current source is $\omega = 0.1 \text{ s}^{-1}$; altitude of the current maximum is $z_1 = 0$; external current maximum value is $j_0 = 1 \text{ } \mu\text{A/m}^2$; horizontal and vertical scales of the external current spatial distributions are $x_0 = y_0 = 500 \text{ km}$ and $z_0 = 10 \text{ km}$, respectively; the angle between the geomagnetic field and vertical direction $\theta = 30^\circ$; in each horizontal panel, left, central and right figures corresponds to the altitude at the ionosphere, equal to $z = 150, 200, 250 \text{ km}$, respectively



of Figs. 2 and 5, between central pictures in panels b of Figs. 2 and 5, and between central pictures in panels d of Figs. 2 and 5, respectively. Note also that the asymmetry in x direction in the spatial distributions of $|E_x|$ and $|E_z|$ (see panels b and d of Fig. 5, respectively) decreases with decreasing frequency from 0.2 to 0.05 s^{-1} . Then, with increasing value of θ from 30° to 60° , the relations between horizontal and vertical components of the electric field change. Namely, for $\theta = 30^\circ$, for the corresponding maximum values, the relation $|E_z| < |E_{x,y}|$ is fulfilled, while for $\theta = 60^\circ$, $|E_z| \sim |E_y| > |E_x|$, see panels b, c and d for the values $|E_x|$, $|E_y|$ and $|E_z|$ at Fig. 2 (for $\theta = 30^\circ$) and Fig. 5 (for $\theta = 60^\circ$). For the other equal parameters, the values $|E_x|$, $|E_y|$ and $|E_z|$ depend only very slightly on the altitude, in the altitude ranges (100–400) km, see Fig. 3b, curves 1,

2 and 3. The same effects are also illustrated by the comparison between the spatial field distributions, see panels b, c and d in Fig. 2. Note that the maximum values in the spatial distributions of E_x , E_y in the horizontal plane x , y and the modulation of the electron concentration are the same, by the order of value, in the range of the angles $10^\circ \leq \theta \leq 60^\circ$.

Note the following drawback of the present model based on the concept of the external current sources. The logic of this concept, used in the present paper, inevitably implies the ignoring of any processes, such as electron and ion transport (drift), formation of charged aerosols, plasma and/or hydrodynamic instabilities, photochemical processes, etc., which in a real “atmosphere–ionosphere” system leads to a formation of the corresponding effective current source. Naturally,

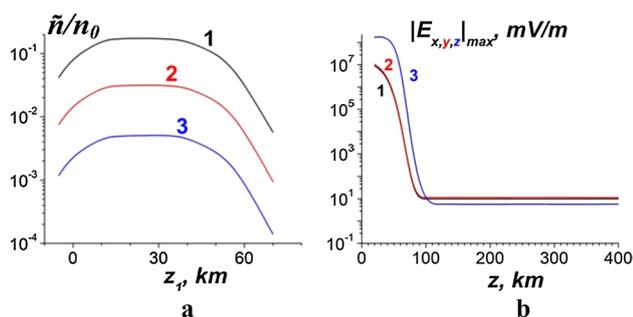


Fig. 3 Dependences of maximal values of relative variations of the electron concentration \tilde{n}/n_0 at the three different altitudes in the ionosphere on the altitude of the maximum of external current source z_1 external current maximum value are $j_0 = 1 \mu\text{A}/\text{m}^2$; **a** the absolute values of the electromagnetic fields components, $|E_x|$, $|E_y|$, $|E_z|$ (mV/m), on the altitude z at the ionosphere and atmosphere (**b**); the angle between the geomagnetic field and vertical direction $\theta = 30^\circ$; frequency of the external current source is equal to $\omega = 0.1 \text{ s}^{-1}$; curves 1, 2 and 3 correspond to the ionospheric altitudes $z = 150, 200, 250 \text{ km}$ (**a**) and to the x, y and z components of the electric field, respectively

in reality, the above-mentioned transport (plasma) processes and the corresponding electric field in the region(s) of such effective current sources are formed and therefore should be determined in a self-consistent manner. Naturally, because we ignore in the approximation adopted in the present article the details of the external source formation, we refuse also from any possibility of the adequate determination of the realistic electric field in the source region. We refuse from any pretension on the adequate determination of the electric field in the region of a current source location. In particular, by this reason it is principally senseless to show an electric field in the lower atmosphere region of the seismogenic source formation, in particular in the first 10 kms over the ground in Fig. 3b. Therefore, when we are talking about the generation of electric field by any hypothetical external current sources placed at the different altitudes ($z_1 \sim (0-55) \text{ km}$, see Fig. 3b) in the atmosphere, we mean that an adequate determination of the corresponding field is possible only at the altitudes lying well above the regions of the location of the corresponding current sources, in particular in the ionosphere at 70–250 km (see Fig. 3b).

Let us emphasize that, with decreasing frequency of an external current source with a maximum placed at the ground ($z_1 = 0$), from 0.2 to 0.05 s^{-1} , for $j_0 = 1 \mu\text{A}/\text{m}^2$, the maximum value of the relative perturbation in the electron concentration at the altitude $z = 200 \text{ km}$ in the ionosphere increases from 0.007 to 0.035 (see panel a in Fig. 5). As it was mentioned in Introduction, the placement (of the maximum) of an external current source at the different altitudes z_1 (Fig. 3a) mimics corresponding effective sources, formed hypothetically due to a possible hydrodynamic-plasma-photochemical instability (not considered directly

in the present paper) in the “atmosphere–ionosphere” system. Alternatively, a small raise in the location of an external current source from the ground level to an altitude up to, say, 10 km, may be connected with the presence of the convective currents (Sorokin et al. 2001). As it is seen from Fig. 3a, the maximum in the relative perturbations in the electron concentration in the range of the ionospheric altitudes $z = (150-250) \text{ km}$ may be reached for z_1 laying in the ranges $\sim (15-55) \text{ km}$, while this maximum level of the plasma perturbation decreases from a value larger than 10% to a value of order of 3×10^{-3} , when the ionospheric altitude increases from 150 to 250 km (see curves 1, 2, 3 in Fig. 3a). For $\omega = 0.1 \text{ s}^{-1}$ and the values $z_1 = 25 \text{ km}$ and $z_1 = 55 \text{ km}$, spatial distributions of the relative perturbation in the electron concentration at the ionospheric altitude $z = 200 \text{ km}$ are presented in the first pictures in the panels a and b of Fig. 4. Maxima of these spatial distributions correspond to the proper points in curve 2 shown in Fig. 3a. Emphasize also an interesting peculiarity in the shapes of spatial distribution of the electron concentration in the ionosphere. Namely, with increasing ionospheric altitude from $z = 150 \text{ km}$ to $z = 200 \text{ km}$, corresponding distribution with two spatial maxima is replaced by the spatial distribution with one maximum (see first and second pictures, respectively, in the panel a of Fig. 2).

The simulations with extremely high values of near-ground external current, of order of $10 \mu\text{A}/\text{m}^2$ (Kuo et al. 2011), give the following results < summarized below without including corresponding figures. These results are obtained for the current sources with frequencies $\omega = 0.1 \text{ s}^{-1}$, $\omega = 0.05 \text{ s}^{-1}$ and $\omega = 0.2 \text{ s}^{-1}$. It can be seen that a corresponding value of the horizontal component of the electric field penetrating into E and F layers of the ionosphere reaches a value of order of 100 mV/m. Corresponding value of relative perturbation of electron concentration reaches, at the altitudes from E to F layers, a value of order of (14–18) %, (16–30) % and up to 35% for the frequencies of the current source equal to $\omega = 0.2 \text{ s}^{-1}$, $\omega = 0.1 \text{ s}^{-1}$ and $\omega = 0.05 \text{ s}^{-1}$, respectively. The electric field of order of 100 mV/m and corresponding perturbation of electron concentration of order of few dozens of percent can play a role of the initiating and seeding factors for the development of plasma instabilities, formation of plasma bubbles and formation of TEC perturbations of order of dozens of percent observed, in particular during the earthquake preparation processes (Pulinets 2011; Kuo et al. 2011).

But the volume seismogenic current density, $10 \mu\text{A}/\text{m}^2$, possible, in accordance with Kuo et al. (2011), is one order larger than the corresponding density of the stationary magnetospheric convection current (Lyatsky and Maltsev 1983). Hypothetically (and the checking of such a hypothesis will need a special paper and is not a subject of the present one), an increased current with the volume density of order of 10

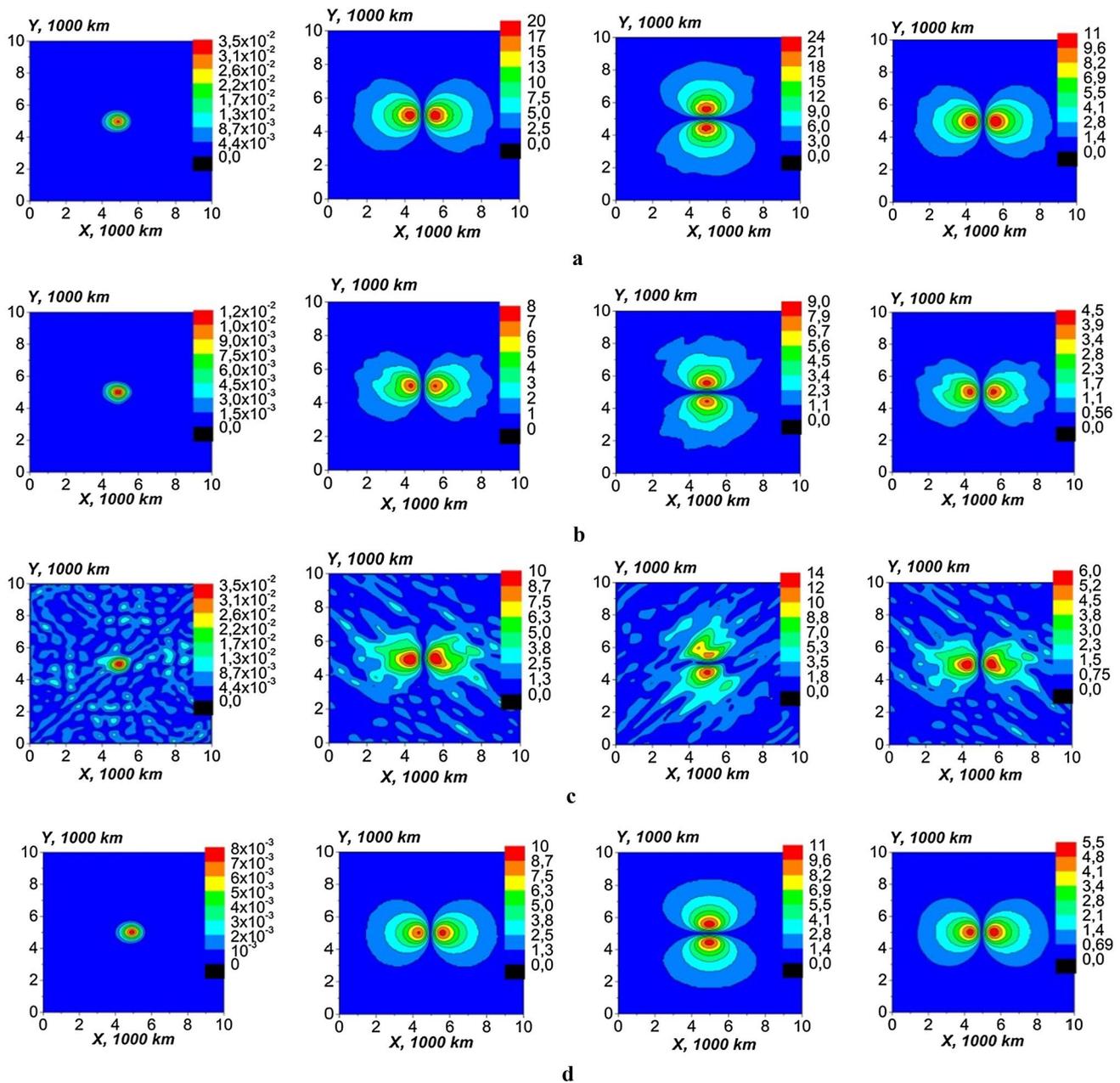


Fig. 4 Spatial distributions of the relative perturbation of the electron concentration $|\delta n|/n_0$ (first figure in each horizontal panel) and the absolute values of the electromagnetic fields components, $|E_x|$, $|E_y|$, $|E_z|$ (mV/m) (second, third and fourth figures in each horizontal panel, respectively) for the ionospheric altitude $z=200$ km, different values

of external current frequency ω and different values of the altitude of the external current maximum z_1 ; external current maximum value is $j_0 = 1 \mu\text{A}/\text{m}^2$; the angle between the geomagnetic field and vertical direction $\theta = 30^\circ$; **a** for $\omega=0.1 \text{ s}^{-1}$, $z_1=25$ km; **b** for $\omega=0.1 \text{ s}^{-1}$, $z_1=55$ km; **c** for $\omega=0.05 \text{ s}^{-1}$, $z_1=0$; **d** for $\omega=0.2 \text{ s}^{-1}$, $z_1=0$

$\mu\text{A}/\text{m}^2$ can be considered qualitatively as the effective current source, which, as a result of the developing one of the possible hydrodynamic-plasma instability, has risen up to a level, one order larger than a value characterizing corresponding sources, not supported by a possible system instabilities. As a result of an increasing in magnitude of j_0 maximum by one order of value with leaving all other parameters

identical to these used for the simulation illustrated in the right figure in panel a of Fig. 5, the relative perturbation in the electron concentration would be tens of percent (in particular $\sim 35\%$). Such a hypothetical value corresponds by the order of magnitude to the variation in a plasma concentration occurred, in particular before L'Aquila (Italy) 2009 earthquake (Stangl et al. 2011; Pulnits et al. 2011, 2014).

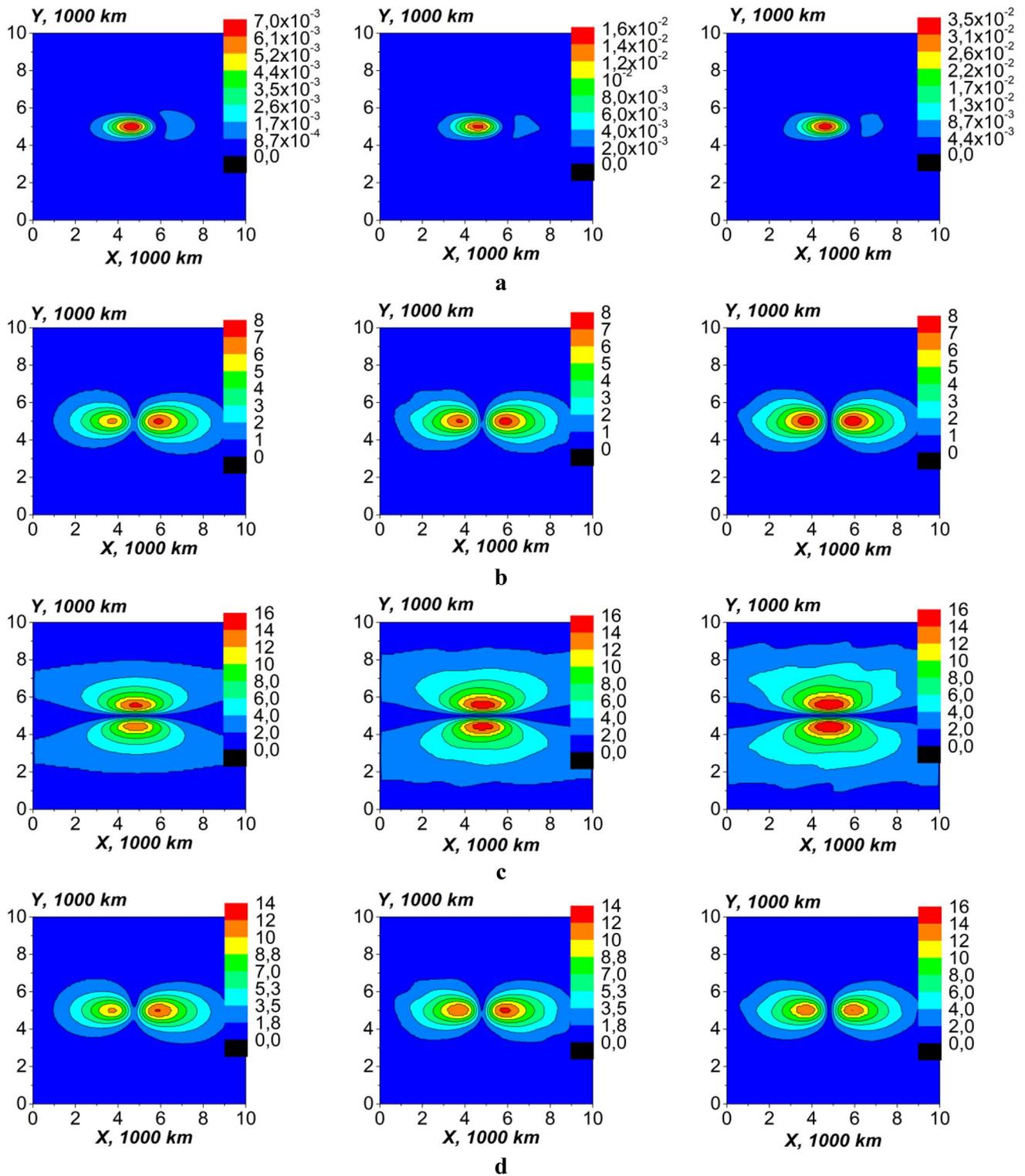


Fig. 5 Spatial distributions of the relative perturbation of the electron concentration $|\tilde{n}|/n_0$ (a) and the absolute values of the electromagnetic field components, $|E_x|$, $|E_y|$ and $|E_z|$, (mV/m), (b)–(d), respectively; altitude of the current maximum is $z_1=0$; horizontal and vertical scales of the external current spatial distributions are $x_0=y_0=500$ km and $z_0=10$ km, respectively; altitude at the ionosphere, equal to

$z=200$ km; external current maximum value is $j_0 = 1 \mu\text{A}/\text{m}^2$; the angle between the geomagnetic field and vertical direction is $\theta = 60^\circ$; in each horizontal panel, left, central and right figures correspond to the frequencies of the external current source are $\omega=0.2 \text{ s}^{-1}$, 0.1 s^{-1} and 0.05 s^{-1} , respectively

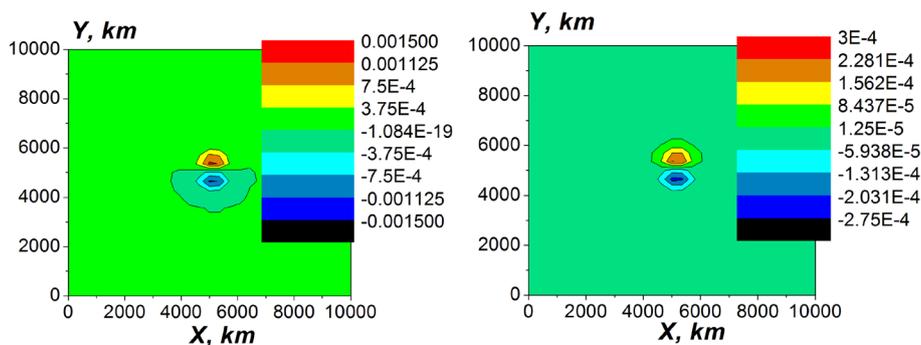


Fig. 6 Purely electrostatic simulations. The electric field components E_y (mV/m) at $z=150$ km (left figure) and $z=200$ km (right figure). External current maximum value is $j_0 = 1 \mu\text{A}/\text{m}^2$. The angle between

the geomagnetic field and vertical direction is $\theta = 30^\circ$. The boundary condition for the electric potential $\varphi=0$ is given in the upper ionosphere at $L_z=1000$ km

As it is shown in Fig. 6, the electric field excited by a stationary external current source penetrates into the ionosphere with much lower effectiveness and the corresponding values of the field are much (a few orders of magnitude) smaller than for the dynamic external current sources. To see this, compare, for example, Fig. 6b with the central panels of Fig. 2. The same conclusion may be drawn also based on the comparison between the maxima in the spatial distributions of the stationary component $|E_y|$ (Fig. 6a, b) with the corresponding (to $z=150$ km and $z=200$ km) points in the curve 2 shown in Fig. 3b. The results shown in Fig. 6 are obtained using Eqs. (13), (14) and the boundary conditions (15). This model is similar to one described in Grimalskiy et al. (2003).

An example of a process with one of the ionospheric instabilities and the seeding factor, other than electric field, for the developing ionospheric structures, is outlined briefly in the next subsection.

Evaluation of the initiation of variations in electron concentration by atmospheric gravity waves of the seismogenic origin

In addition to the initiation of perturbations in the electron concentration by the electric field penetrating into the ionosphere from the current source placed in the lower atmosphere/lithosphere, let us note a possibility of the corresponding seismogenic effect in the F layer of the unstable ionospheric plasma, initiated by AGW of the seismogenic origin, as a seeding factor (Bryunelli and Namgaladze 1988; Rapoport et al. 2004b; Fedorenko et al. 2015; 2018). This is the instability of the collisional plasma with the gradient of the concentration, directed oppositely to the gravity force and the motion of the ions in crossing electric and magnetic fields (Huang and Kelley 1996a, b; Huang et al. 1994; Kelley 2009; Treumann and Baumjohann 1997). This instability is called the combined Rayleigh–Taylor- $E \times B$

instability. In particular, such a possibility, alternative, respectively, to the effects provided by the electromagnetic seismo-ionospheric coupling, will be demonstrated, based on the linear model (Rapoport et al. 2009) and the data of the observations (Fedorenko et al. 2015, 2018). We will not stay on all the details of this model, described in Rapoport et al. (2009) and partly in Rapoport et al. (2004b) and based on the approaches described in these papers and the references in these papers (Huang and Kelley 1996a, b; Huang et al. 1994; Sekar and Raghavarao 1987 etc.), reflecting in particular the long story about the modeling Rayleigh–Taylor instability in the ionosphere and its seeding by AGW. Nevertheless, we outline briefly here only some basic points of the model (Rapoport et al. 2009) and then present an example supported theoretically and experimentally, a possibility of the formation of seismogenic plasma structure in the ionosphere.

This model includes AGW wave packet excited by means of the near-ground seismogenic source of finite width (see also the caption to Fig. 7 with the AGW source shape specification); full spectrum of AGW with reactive modes and the effect of “spatial resonance under the excitation of AGW packet” (Rapoport et al. 2009); simplified isotropic and linear model of the atmosphere; linear model of the development of plasma structure in the equatorial plasma with combined Rayleigh–Taylor- $E \times B$ instability (Kelley 2009; Rapoport et al. 2009); accounting for AGW forcing of the combined Rayleigh–Taylor- $E \times B$ instability due to both vertical and horizontal components of AGW velocities and density. The model has been verified (very successfully) on the basis of comparison between the results of our theoretical approach and “Atmosphere Explorer-E” satellite measurements (Rapoport et al. 2009) of post-earthquake structures in the ionosphere and neutral component in the F layer.

A possibility of the initiation of electron and neutral ionospheric component’s perturbations by means of AGW packet radiated from the near-ground seismogenic source and the

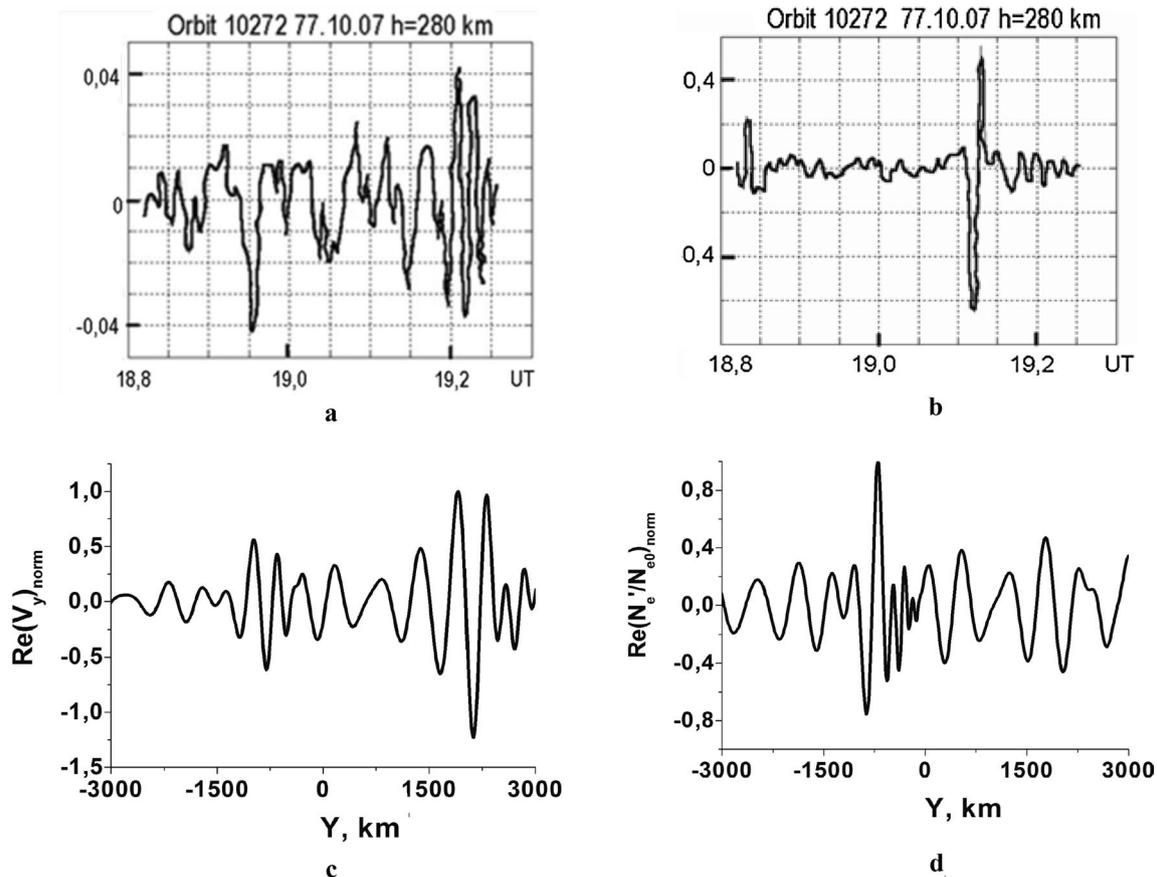


Fig. 7 Experimental (a), (b) and theoretical (c), (d) results for the seismogenic perturbations of the near-equatorial ionosphere; a, b experimental dependences of the relative values of the concentration on the time based on the data of the satellite “Atmospheric Explorer-E” (results have been obtained and processed by Fedorenko et al. (2005); c, d results of theoretical modeling based on the model (Rapoport et al. 2009) for the normalized spatial distributions of the horizontal velocity of AGW and perturbation of the electron concentration. Perturbations are connected with the seismogenic effect from the earthquake with $M=6.5$. Distribution of the concentrations of neutrals (N_2) (a) and ions O_+ (b), cm^{-3} on the “seismoactive” part of the satellite path; relative variation after the “exclusion” of the slow trend of the obtained data. Period AGW $\tau \sim 30$ min; b spatial distribution of the normalized relative perturbations of the electron concentration. Altitude $z=250$ km; d spatial distribution of

the normalized horizontal velocity of the AGW, connected with the perturbations of the concentration of neutral particles; c, d horizontal velocities of the wind and ions are equal to 60 and 20 m/s, respectively. Axis Y in figs. (c, d) is directed to the east. Satellite moves practically from the west to the east. To the position of epicenter, the point of the trajectory of the satellite, for which $UT=19.15$ in figures. (a), (b) and point $Y=0$ in the figure. (c), (d) does correspond. For the modeling, the boundary conditions with ground AGW source (Gokhberg et al. 1996) of finite width have been adopted, namely vertical component of AGW velocity has been proportional to the function $f(t, y) = \exp(2\pi t/\tau) \text{ch}^{-2}(y/l_y)$, where $l_y = 100$ km. The altitude is $z=250$ km, and the parameters are (Sekar and Raghavarao 1987; Genkin et al. 1987; Huang and Kelley 1996a, b) $v_{in0} = 1 \text{ s}^{-1}$, $\alpha_0 = 0$ [which reflects qualitatively the effect of the (local) night conditions (Basu 2002)], $U_{i0z} = -10$ m/s, $U_{i0x} = 20$ m/s, $U_{0x} = 60$ m/s, $U_{0z} = 0$

results of the corresponding observations of seismogenic effects in the near-equatorial plasma (Fedorenko et al. 2005) are illustrated in Fig. 7.

Dynamics of the developing of combined Rayleigh–Taylor- $E \times B$ instability in the presence of spatial packet of AGW and stationary drift of ions (caused by corresponding crossed stationary–quasi-stationary electric fields and geomagnetic field) is described by the system of linearized equations for Fourier modes of electron concentration. These equations have been derived, on the basis of the corresponding system of equations for the low-frequency linear

perturbations in quasi-neutral three-component plasma of the ionospheric region F in the inertialess approximation, in Rapoport et al. (2009) [see, for further details, Rapoport et al. (2009) and references (Huang and Kelley 1996a; Basu 2002; Hooke 1968), and other references, cited in this connection in Rapoport et al. (2009)]. This system for the Fourier modes has the form:

$$n = n_0(t) e^{i(\omega_{AGW} - k_y y - k_z z)}; \quad \frac{\partial n_0}{\partial t} = \gamma n_0 - i \Delta \omega n_0 + F; \quad n \sim n_0 \sim \frac{N'_e}{N_{e0}} \quad (26a)$$

In Eq. (26a), $n_0(t)$ is dimensionless amplitude of the normalized perturbation n of the electron concentration in the ionosphere, N_{e0} is a stationary electron concentration in the ionosphere, $\Delta\omega = \omega_{AGW} - \omega'_{RTI}$ is the frequency difference between AGW mode, $\omega_{AGW} \equiv \omega$ and unstable plasma mode, $\omega\omega'_{RTI}$; $k_{x,y}$ are the horizontal wavenumber components of a Fourier mode included into AGW wave packet and electron concentration perturbation; F plays a role an effective “external force” for the corresponding effective oscillator describing the perturbations of the electron concentration, connected with AGW packet. The value F has a dimensionality $[t^{-1}]$. The form of the second equation from (26a) for $n_0(t)$ is the same as one in (Huang and Kelley 1996a, b). As it is seen from (26a), $n_0(t)$ plays a role of the envelope of the electron concentration, while the first relation from (26a) includes carrier frequency and wavenumber of the corresponding AGW mode. The effects of plasma response to the corresponding AGW mode are included into $n_0(t)$. The final result on the plasma excitation is obtained after the summation by all the Fourier modes (k_x, k_y) (Rapoport et al. 2011; Huang and Kelley 1996a, b). An increment γ of combined Rayleigh–Taylor- $E \times B$ instability for a definite Fourier mode and the real part of frequency of the corresponding unstable plasma mode are determined by the relations (here Y is directed to the East, Z is directed upwards):

$$\gamma = \omega''_{RTI} \approx \left(\frac{\frac{k_y}{L_{ef}} \left[k_y \left(-\frac{1}{L_{ef}} \frac{k_B(T_i+T_e)}{m_i} + g - v_{in0} U'_{0z} \right) + k_z v_{in0} U'_{0y} \right]}{v_{in0} \left[(k_y^2 + k_z^2)^2 + \frac{k_z^2}{L_{ef}^2} \right]} \right) \times (k_y^2 + k_z^2) - \alpha_0; \tag{27a}$$

$$\omega'_{RTI} \approx \vec{k} \cdot \vec{V}_{i0} + \frac{2}{L_{ef}} \frac{1}{\omega_{Hi}} k_y \frac{T_i}{m_i} - \left(\frac{\frac{k_y}{L_{ef}} \left[k_y \left(-\frac{1}{L_{ef}} \frac{k_B(T_i+T_e)}{m_i} + g - v_{in0} U'_{0z} \right) + k_z v_{in0} U'_{0y} \right]}{v_{in0} \left[(k_y^2 + k_z^2)^2 + \frac{k_z^2}{L_{ef}^2} \right]} \right) \frac{k_z}{L_{ef}}; \tag{27b}$$

where g is free-fall acceleration, k_z is effective vertical wavenumber (mode) component of the wave packet,

$$U'_{0y} = U_{0y;eff} = U_{0y} + \frac{e}{m_i \omega_{Hi}} |E_{0z}|; \tag{27c}$$

$$U'_{0z} = U_{0z;eff} = U_{0z} - \frac{e}{m_i \omega_{Hi}} |E_{0y}|.$$

Here $L_{ef}^{-1} = N_{e0}^{-1} \cdot dN_{e0}/dz$ (Sekar and Raghavarao 1987), $v_{in} = v_{in0}(1 + \rho'/\rho_0)$ (Sekar and Raghavarao 1987; Genkin et al. 1987), ρ_0 and ρ' -stationary density and its perturbations in neutral atmosphere in the presence of AGW; $T_{i,e}$ are the

electron and ion temperatures, respectively; $v_{in0}, \omega_{Hi}, \alpha_0$ are ion–neutral collision frequency, electron cyclotron frequency and recombination rate in the F layer, respectively (Rapoport et al. 2009, 2012, b, c); \vec{E}_0 and \vec{U}_0 are stationary electric field and the wind velocity, respectively.

The effective “external force” F , connected with AGW, is ($U_{x,z}$ are AGW velocity components, $U_{0x,z}$ are wind velocity components):

$$F \approx F_{U_z} U_z + F_{U_y} U_y + F_\rho \frac{\rho'}{\rho_0} - \alpha_0 \left(\frac{T'}{T_0} + \frac{P'}{P_0} \right). \tag{28a}$$

Here T', P' and T_0, P_0 are alternating and stationary components of temperature and pressure, respectively (variations of which are connected with AGW),

$$F_{U_z} \approx ik_y \left(2 \frac{v_{in0}}{\omega_{Hi}} - \frac{\frac{1}{L_{ef}} ik_y}{\left[k_y^2 + k_z \left(k_z + \frac{i}{L_{ef}} \right) \right]} \right); \tag{28b}$$

$$F_{U_y} \approx i \left(k_z + \frac{i}{L_{ef}} \right) \left(\frac{\frac{1}{L_{ef}} ik_y}{\left[k_y^2 + k_z \left(k_z + \frac{i}{L_{ef}} \right) \right]} - 2 \frac{v_{in0}}{\omega_{Hi}} \right); \tag{28c}$$

$$F_\rho \approx \left(2i \frac{v_{in0}}{\omega_{Hi}} + \frac{\frac{1}{L_{ef}} k_y}{\left[k_y^2 + k_z \left(k_z + \frac{i}{L_{ef}} \right) \right]} \right) \left(k_y (U_{0z} - V_{i0z}) - \left(k_z + \frac{i}{L_{ef}} \right) (U_{0y} - V_{i0y}) \right) + \alpha_0. \tag{28d}$$

The first, second and third terms in Eq. (28a) are the components of the effective “external force” connected with the vertical and horizontal components of AGW velocity and density perturbations, respectively. The corresponding coefficients F_{U_z}, F_{U_y} and F_ρ have the dimensionalities $[L^{-1}], [L^{-1}]$ and $[t^{-1}]$, respectively. Consider combined Rayleigh–Taylor- $E \times B$ instability in the presence of ion stationary motion and the wind. This motion can be caused by the presence of crossed geomagnetic and electric quasi-stationary fields. Electric fields can be either of ionospheric origin, or be brought into the ionosphere by quasi-stationary electromagnetic processes of seismogenic origin, such as radon emanation or appearance of near-ground current, as it is considered in Sect. 3.1. Therefore, the processes of development of instability (and corresponding developing plasma bubbles (Huang and Kelley 1996a, b) and strong seismogenic variations of TEC) can be caused by AGW (Huang and Kelley 1996a, b) and/or quasi-stationary electric field of the seismogenic origin Kuo et al. (2011). Instability is determined by a velocity of ions relatively to a media,

$$U_{0x,z;\text{eff}} = U_{0x,z} - U_{i0x,z}; \quad (29)$$

where $U_{0x,z}$ are horizontal and vertical components of wind velocity, respectively, $U_{i0x,z}$ are horizontal and vertical components of the velocity of ion motion, respectively. The other details of the model and algorithm can be found in the papers (Rapopot et al. 2009) and the references included in this paper, in particular in Huang et al. (1994), Huang and Kelley (1996a, b), Genkin et al. (1987), Sekar and Raghavarao (1987), etc. Results of the modeling, namely spatial distribution of ionospheric concentration response in the presence of horizontal wind and ion motion, and corresponding experimental data (Fedorenko et al. 2005) are shown in Fig. 7.

The correspondence between the results, obtained using the model (Rapopot et al. 2009), and the data of the observations of the seismogenic structures in the unstable ionospheric F layer plasma, initiated by the AGW of the lithospheric origin, obtained on the board of the satellite “Atmosphere Explorer-E” (Fedorenko et al. 2005), is demonstrated, namely (Fig. 7): (1) the region of plasma perturbations is shifted to a distance of about 2500 km from the eastern maximum of the field AGW in the presence of an east wind (Fig. 7a–d); (2) the region of the plasma instability is localized and has typical dimensions of the order of 1500 km (Fig. 7b, d); (3) in perturbations of the plasma and neutrals, quasi-periodic components with a spatial period \sim (600–800) km are allocated (Fig. 7a–d). Note also that relative value of plasma concentration disturbances can reach tens of percent during a time from tens of minutes to several hours after the onset of RTI.

As it is seen from Fig. 7, the spatial asymmetry in unstable plasma is connected with a direction of ion motion or corresponding (vertical, in this case) electric field. This asymmetry in N'_e/N_{e0} spatial distribution relatively to the directions «+ Y» and «- Y» (i.e., «East» and «West») is caused by an anisotropy of equatorial plasma instability in the presence of ion motion in “east–west” direction (or vertical electric field in upward–downward directions). The modeling demonstrates (in a linear approximation) that a characteristic time of plasma structure formation is of order of few dozens of minutes since “switching on” an influence of AGW and developing combined Rayleigh–Taylor- $E \times B$ instability in a near-equatorial plasma. Nevertheless, in spite of the fact that this theoretical result also formally corresponds to the observational results, it should be noted the following. To get a comprehensive quantitative correspondence between the theory and experiment, it is necessary to develop nonlinear theory accounting for the saturation of ionospheric plasma instability. In particular, plasma bubbles (Kelley 2009) could form before earthquake in near-equatorial region, and the formation and grows of the plasma bubbles may result in the corresponding TEC change (Pulinets 2011; Kuo et al. 2011; Pulinets and Boyarchuk 2005).

Nevertheless, the results, obtained in the linear approximation for the unstable plasma and neutral structures, formed in the ionospheric F layer under the influence of the AGW packet of the seismogenic origin correspond qualitatively to the data of satellite observations, as described above.

Discussions and conclusions

The new quasi-stationary–electromagnetic *algorithms* of seismo-ionospheric coupling using the method of “successive electrostatic–quasi-stationary–electromagnetic approximations-equivalent external sources (EQUEMES)” are proposed.

Both the algorithm for TEC computation based on the “ambipolar diffusion” approximation, which is valid approximately for the range of (200–500) km, and the more detailed algorithm, which provides an inclusion of E and F layers in the “region of integration” for the determination of TEC, are described.

The simulations of the penetration of the electric field from near-Earth sources into the ionosphere have demonstrated that this penetration should be considered within the dynamic model, where *all the components* of the conductivity are considered as finite. Mathematically, putting to infinity any coefficient in a differential equation leads to the loss of information about other terms in the equation. The penetration of ULF electromagnetic field into the ionosphere becomes more effective when the frequency decreases at $\omega < 1 \text{ s}^{-1}$. The purely electrostatic problem of penetration with all finite components of the conductivity demonstrates low values of the penetrated electric field, in particular for the zero upper boundary condition for the potential. But, at $\omega < 0.05 \text{ s}^{-1}$, i.e., at the big periods, the simulations in the frequency domain possess some problems, which will be overcome in a further work, while a transition to a time domain is expected. Therefore, an existence of some optimum value of the ULF frequency/timescale can be expected, which would provide the most effective penetration of the electromagnetic field into the ionosphere. The optimization should be realized in the temporal domain; the corresponding periods (or characteristic times) are $T > 100 \text{ s}$.

A possibility of the formation of the seismogenic structures of the charge and neutral components in the near-equatorial F region plasma is demonstrated. Such structures can be developed in the presence of the combined Rayleigh–Taylor- $E \times B$ instability and a packet of AGW of the seismogenic origin. The correspondence between the results of theoretical modeling and the data of the satellite (“Atmospheric Explorer-E”) observations is shown. Namely: (i) the region of plasma perturbations is shifted to a distance of about 2500 km from the eastern maximum of the field AGW in the presence of an east wind; (ii) the region of

the plasma instability is localized and has typical dimensions of the order of 1500 km; (iii) in perturbations of the plasma and neutrals, the quasi-periodic components with a spatial period \sim (600–800) km are allocated. To get a comprehensive quantitative correspondence between the theory and experiment, the nonlinear saturation of ionospheric plasma instability should be accounted for. The formation and growth of the plasma bubbles may result in the corresponding TEC change (Pulinets 2011; Kuo et al. 2011; Pulinets and Boyarchuk 2005). Such a seismogenic change in TEC may be rather strong (of order of tens percent) and can be initiated either by AGW packet or by electric field of seismogenic origin, or both, penetrating into the ionosphere from the lower atmosphere.

In the frames of the proposed model, the penetration of the electromagnetic field into the ionosphere from the lower atmosphere could be considered in the most adequate way. Such a model could be useful for comparison between theoretical and experimental results, obtained both on a satellite and in ground observatories, and for better understanding mechanisms of seismo-ionospheric coupling.

A further development of the present model paves a way for searching so-called trigger phenomena. The last are connected with physical consequences of synergetic properties (Pulinets 2011) of opened and unstable system “lithosphere–atmosphere–ionosphere–magnetosphere (LAIM)”. A set of possible instabilities are connected with a heating of the ionosphere which can cause a developing of atmospheric gravity waves (Kotsarenko et al. 1994). In this perspective, this model can be developed in a self-consistent manner, including the self-interaction of electromagnetic waves in the system LAIM due to a modification of the media (conductivity) under the influence of the electric field. For example, a possibility of developing some sort of the “negative differential conductivity” with corresponding heating-photochemistry current instability could be expected (Sorokin and Hayakawa 2013; Rapoport et al. 2004a, b, 2005; Gurevich 1978).

Before some of the catastrophic phenomena such as earthquakes, an infrasound is generated, and this phenomenon is the physical basis of the acoustic channel of seismo-ionospheric coupling. The monitoring of disasters precursors is important. Using the original parametric ground-based sound generator, which intensity is comparable with a flying up jet, the theory has been constructed and the series of experiments have been conducted to model acoustic action of disasters on the ionosphere (Koshovyy et al. 2005; Kotsarenko et al. 1999; Cheremnykh et al. 2014, 2015). The analysis of these experiments leads as well to the assumption that the synergy processes reveal itself in the experiments on the penetration and nonlinear transformation of the artificially generating sound wave through the atmosphere and ionosphere. As it was pointed out in Emelyanov et al. (2015),

the energy release due to water vapor stored in the atmosphere and definite photochemistry processes may be a reason caused by the reaction of the ionosphere to the artificial acoustic waves launched by the parametric sound generator (Emelyanov et al. 2015). Probably, this also concerns the experiments (Koshovyy et al. 2005; Kotsarenko et al. 1999; Cheremnykh et al. 2014, 2015). The synergetic approach (Pulinets 2011) also incorporates atmospheric water vapor as an important factor.

It was shown in Zettergren and Snively (2013) that acoustic waves generated by tropospheric sources may reach significant amplitudes in the ionosphere, with temperature and vertical wind perturbations on the order of tens of Kelvins and m/s, respectively, in the *E* and *F* layers. The perturbations of the TEC are detectable by ground-based radar and GPS receivers. Acoustic waves can drive field-aligned currents, detectable by in situ magnetometers (Iyemori et al. 2015). Measurements of GPS-derived TEC (Zettergren and Snively 2015) reveal acoustic wave periods from \sim 1 to 4 min in the *F*-layer ionosphere following natural hazard events, including earthquakes, severe weather fronts and volcanoes.

We assume that to search the seismogenic effects in the ionosphere, it is necessary to unify the well-developed model of the electromagnetic channel to seismo-ionospheric coupling (Molchanov et al. 1995; Rapoport et al. 2004b; Grimalsky et al. 1999; Pulinets and Boyarchuk 2005; Sorokin and Hayakawa 2013) with the model of the acoustic channel (Gokhberg and Shalimov 2000; Rapoport et al. 2004a, 2009; Koshevaya et al. 2002; Grimalskiy et al. 2003; Zettergren and Snively 2015; Iyemori et al. 2015). Until recently, AGWs and electromagnetic waves have been considered on the basis of two main competitive mechanisms of seismo-ionospheric coupling (Sorokin and Hayakawa 2013; Klimenko et al. 2011; Pulinets and Boyarchuk 2005).

Finally, the necessity of the discussed synergetic approach is supported also by the following experimental data. (1) The perturbations with periods 20 and 2–5 min. of characteristic for gravity wave branch AGW are observed in TEC few days before the strongest earthquakes, such as in Nepal ($M=7.8$) and Chile ($M=8.3$) in 2015. (2) Besides the perturbations in TEC, the same strongest seismogenic sources can cause the disturbances in VLF signals propagating in or radiated from the waveguide “earth–ionosphere” (WGEI) as well (Stangl et al. 2011; Sanchez-Dulcet et al. 2015). (3) Both the strongest seismogenic and meteorological and other geophysical sources (cyclones, typhoons, tsunamis, etc.) cause oscillations of the VLF signals in the range of gravity waves (Nina and Čadež 2013; Rozhnoi et al. 2014a, b, 2015; Solovieva et al. 2015). (4) The modification of the ionosphere plasma can be caused by either electric/electromagnetic field, or AGW, or both (Pulinets 2011; Rapoport et al. 2004a, b, 2005, 2011, 2014b, 2017; Klimenko et al. 2011). Therefore, the importance of AGW/gravity wave branch as an agent of

the variations in the plasma concentration, altogether with the electric/electromagnetic field, becomes evident.

The presence of the effective external electric current source at the different altitudes in the atmosphere–ionosphere system mimics the result of the developing some plasma-photochemical-electric or hydrodynamic-plasma-electric-thermal instabilities as the physical basic elements of the above-mentioned general synergetic picture. These external sources are characterized in the present paper by the dependence of the effectiveness of the corresponding electric field penetration into the ionosphere. Our results reveal the preferable position of the corresponding external sources, providing the maximal effectiveness of the penetration into the ionosphere of the excited electric fields and the plasma perturbations caused by these fields. Namely, the most effective in the sense discussed above is the placement of these effective external currents in the (10–55) altitude range.

The following results of the present paper should be emphasized.

1. We have proposed and described in details electrostatic–quasi-stationary–electromagnetic approximations and equivalent external sources (EQUEMES method) to develop the quasi-stationary–electromagnetic algorithms of seismo-ionospheric coupling. In this paper, we present these new algorithms in detail. The electromagnetic field does penetrate through the atmosphere–ionosphere system more effectively than the static one. In general, the dynamic approach is more adequate, than static one, which can be considered as a corresponding limiting case. To realize such a limiting pass, the EQUEMES method may be used.
2. The penetration of the electromagnetic field created by near-Earth alternative currents of ULF range $\omega < 10 \text{ s}^{-1}$ was simulated by solving the set of equations for electric field components E_x, E_y , of the second order with respect to the vertical coordinate z . The penetration of rather strong horizontal electric field (of order from few to 10 mV/m (Kuo et al. 2011)) to the ionospheric E and F layers has been modeled with current source maximum $1 \mu\text{A}/\text{m}^2$. It was shown that the corresponding variations in the electron concentration in the E and lower F layers of the ionosphere reach a value of order of (1–10) %, what corresponds to the data of the observations of the TEC perturbations before some of the strongest earthquakes (Oikonomou et al. 2016).
3. In the case (Pulinets 2011; Pulinets et al. 2011, 2014, 2015), TEC and corresponding perturbations in the ionospheric plasma concentration reach a value of order of tens percent, what exceeds remarkably the possible plasma perturbations obtained in the frames of the linear and non-self-consistent theory of the electromagnetic channel of coupling developed in the present paper. Such values of the ionospheric plasma perturbations can be achieved if the effective external current sources could reach a value of order of $10 \mu\text{A}/\text{m}^2$, with a corresponding electric field penetrating into the ionosphere of order of 100 mV/m. Such an extraordinary value of the effective current source in the atmosphere/lower ionosphere could mimic hypothetically a result of the developing some of the possible hydrodynamic-plasma-electric-thermal instabilities.
4. For the other equal parameters, the values $|E_x|, |E_y|$ and $|E_z|$ depend only very slightly on the altitude, in the altitude ranges (100–400) km. The maximum values in the spatial distributions of E_x, E_y in the horizontal plane x, y and the modulation of the electron concentration are the same, by the order of value, in the range of the angles $10^\circ \leq \theta \leq 60^\circ$.
5. A possibility of the initiation of electron concentration perturbations by means of AGW packet radiated from the near-ground gas/heat source is illustrated. The correspondence between the results, obtained using our model, and the data of the observations of the seismogenic structures in the unstable ionospheric F layer plasma initiated by the AGW of the lithospheric origin, obtained on the board of the satellite “Atmosphere Explorer-E” (Fedorenko et al. 2005), is demonstrated.
6. To provide the most effective penetration into the ionosphere of the electric field excited by the effective external current sources mimicking the synergetic processes connected with a possible instabilities at the different altitudes of the atmosphere–ionosphere system, the corresponding sources should be placed at the altitude range (10–55) km.

The following problems for the next investigation to understand an electromagnetic channel of seismo-ionospheric coupling are proposed.

1. Developing the algorithm for self-consistent modeling of electromagnetic channel of seismo-ionospheric coupling including atmospheric and ionospheric dynamics of both electromagnetic field and plasma, primarily plasma bubbles that cause variation of $\sim(30\text{--}80)\%$ in TEC (through combined Rayleigh–Taylor- $E \times B$ or Perkins instabilities), but with non-stationary electric fields in the atmosphere, in distinction to (Kuo et al. 2011).
2. Searching for confirmation of the hypothesis of the “funnel effect,” while funnel will include both some instabilities in an atmosphere (near-lithospheric layer/lower atmosphere, D and F layers, including effects of heat release, photochemistry and aerosol effects) and F layer (combined Rayleigh–Taylor- $E \times B$ or Perkins instabilities) of the ionosphere.

3. The self-consistent model of the electromagnetic channel of TEC disturbances formation including the developing possible hydrodynamic–photochemistry–electric instabilities at the different altitudes of the atmosphere–ionosphere system should be developed.
4. We assume that a unified description of electromagnetic and acoustic channels is necessary as the basis for the synergy theory of the mechanism of seismo-ionospheric coupling (Pulinets 2011; Rapoport et al. 2014a, b; Pulinets and Boyarchuk 2005).

In a perspective, a nonlinear AGW model (Rapoport et al. 2017) and non-stationary self-consistent electromagnetic channel model will be united, to establish the synergetic mechanism of seismogenic TEC variations, and formation of plasma bubbles and strong variations in TEC.

Acknowledgements Two of authors (Andrzej Krankowski and Sergey Pulinets) thank the Ministry of Science and Higher Education (MSHE) of the Republic of Poland for granting fund: MSHE decision no. DIR/WK//2016/05) and MSHE decision no. 220815/E-383/SPUB/2016/2), which supported in part the present work.

Compliance with ethical standards

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

Appendix 1

In this “Appendix”, we present the expression for the conductivity tensor $\hat{\sigma}_0(z)$ in the magnetized ionospheric plasma corresponding to the stationary problem, when the frequency of the electromagnetic wave is $\omega = 0$. The expressions for the components of the effective permittivity of the ionospheric plasma in the coordinate frame $X'YZ'$ where OZ' axis is aligned along the geomagnetic field \vec{H}_0 are:

$$\begin{aligned} \hat{\sigma}' &= \begin{pmatrix} \sigma_P & -\sigma_H & 0 \\ \sigma_H & \sigma_P & 0 \\ 0 & 0 & \sigma_{\parallel} \end{pmatrix}, \quad \sigma_P = e^2 n_0 \left[\frac{\nu_e}{m_e(\nu_e^2 + \omega_{He}^2)} + \frac{\nu_i}{m_i(\nu_i^2 + \omega_{Hi}^2)} \right]; \\ \sigma_H &= e^2 n_0 \left[\frac{\omega_{He}}{m_e(\nu_e^2 + \omega_{He}^2)} + \frac{\omega_{Hi}}{m_i(\nu_i^2 + \omega_{Hi}^2)} \right], \quad \sigma_{\parallel} = \left(\frac{1}{m_e \nu_e} + \frac{1}{m_i \nu_i} \right); \\ \omega_{pe}^2 &= \frac{4\pi e^2 n_0}{m_e}, \quad \omega_{pi}^2 = \frac{4\pi e^2 n_0}{m_i}, \quad \omega_{He} = \frac{eH_0}{m_e c}, \quad \omega_{Hi} = \frac{eH_0}{m_i c} \end{aligned} \tag{31}$$

Here $\omega_{pe}, \omega_{pi}, \omega_{He}, \omega_{Hi}$ are plasma and cyclotron frequencies for electrons and ions, respectively; $m_e, m_i, \nu_e, \nu_i, H_0$ and c are the masses and the collision frequencies of the electrons and ions, geomagnetic field and light

speed, respectively. The expressions of the components of $\hat{\sigma}_0(z)$ are obtained from (31) by means of multiplication with the corresponding rotation matrices (Spiegel 1959), and the angle between the axes OZ and Oz' is equal to θ . In the case of a medium with a scalar conductivity σ_{00} , like the lower ionosphere or atmosphere, the effective permittivity (1) reduces to the scalar: $\hat{\sigma}_0(z) = \sigma_{00} \hat{I}$, where $\hat{\sigma}_{00} \equiv \hat{\sigma}_P(\omega_{Hi,e} = 0), \hat{I}$ is unit matrix. As a result, tensor $\hat{\sigma}_0(z)$ has all nine nonzero components $\hat{\sigma}_{ij}(i, j = 1 \div 3)$, while $\sigma_{12} = -\sigma_{21}, \sigma_{23} = -\sigma_{32}, \sigma_{13} = \sigma_{31}$.

Appendix 2

Consider the details of the derivation of the dynamic equation solved to obtain Figs. 1, 2, 3, 4 presented in the paper. Consider electromagnetic field excited by external current in the system “lithosphere–atmosphere–ionosphere.” We use the fictitious lateral walls (FLWs). Distances between FLW are supposed to be much larger than typical horizontal scales of current “external source” in the lower atmosphere \vec{J}^{ext} (Fig. 8). We suppose that an electromagnetic field excited by a source \vec{J}^{ext} is zero at FLW. Because the parameters of the medium is supposed to be dependent only on vertical coordinates z , the spectrum electromagnetic fields excited by a source \vec{J}^{ext} is discrete with the horizontal Fourier wavenumbers (k_x, k_y) and the amplitude of Fourier components $F_{x,y,z}^{(k_x,k_y)}(z)$ which depends on Z . Fourier amplitudes of external current are components $E_{x,y,z} = \sum_{k_x,k_y} e^{i(\omega t - k_x x - k_y y)} E_{x,y,z}^{(k_x,k_y)}(z), H_{x,y,z} = \sum_{k_x,k_y} e^{i(\omega t - k_x x - k_y y)} H_{x,y,z}^{(k_x,k_y)}(z),$ while the current source can be presented in the form $J_{x,y,z} = \sum_{k_x,k_y} e^{i(\omega t - k_x x - k_y y)} J_{x,y,z}^{(k_x,k_y)}(z).$

The Maxwell equation can be written in the form

$$\begin{aligned} \vec{\Delta}_x \vec{H} &= ik_0 \vec{D} + \frac{4\pi}{c} \vec{j}; \quad k_0 \equiv \frac{\omega}{c}; \vec{D} = \hat{\epsilon} \vec{E} \\ \vec{\Delta}_x \vec{E} &= -ik_0 \vec{H} \end{aligned} \tag{32}$$

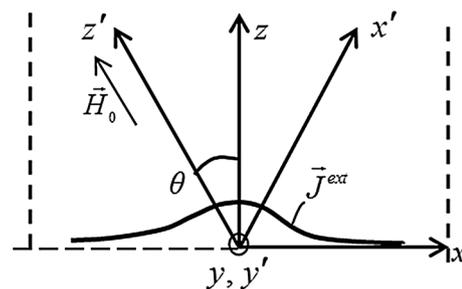


Fig. 8 System “lithosphere–atmosphere–ionosphere” with external current J_{ext} . Geomagnetic field \vec{H}_0 is parallel to z' axis. The fictitious lateral walls are shown by dashed lines

$$\vec{\varepsilon} = \begin{pmatrix} \varepsilon_{11}(z) & \varepsilon_{12}(z) & \varepsilon_{13}(z) \\ \varepsilon_{21}(z) & \varepsilon_{22}(z) & \varepsilon_{23}(z) \\ \varepsilon_{31}(z) & \varepsilon_{33}(z) & \varepsilon_{33}(z) \end{pmatrix} \equiv \hat{\varepsilon}(z) \tag{33}$$

Using (32), one can get

$$\begin{aligned} \vec{\nabla}_x \vec{\nabla}_k \vec{E} &= -ik_0 \vec{\nabla}_x \vec{H} = -ik_0 \left(-ik_0 \vec{D} + \frac{4\pi \vec{J}}{c} \right) = k_0^2 \vec{D} - \frac{4\pi ik_0 \vec{J}}{c} \\ - \vec{\nabla} \operatorname{div} \vec{E} + D\vec{E} + k_0^2 \vec{D} - \frac{4\pi ik_0 \vec{J}}{c} &= 0; \end{aligned} \tag{34}$$

Then, we use the combined spectrum finite-difference method (Boardman et al. 2005; Rapoport et al. 2012a, b, 2014a). Namely, we make Fourier transform of Eq. (34), accounting for (33). As a result, we obtain the relations for Fourier components of the fields. Then, writing

corresponding equation for the Fourier components, we omit the upper indices (k_x, k_y). The same concerns Fourier components of the external source \vec{J}^{ext} . All Fourier amplitudes depend on Z . Writing the formulas for the Fourier components of fields and currents below, we omit the dependences on Z , such as $E_x^{(k_x, k_y)}(z) \rightarrow E_x, J_x^{(k_x, k_y)}(z) \rightarrow J_x$.

In components x, y, z , Eq. (34) takes the form:

$$(x) \quad \frac{\partial^2 E_x}{\partial z^2} - k_y^2 E_x + k_x k_y E_y + ik_x \frac{\partial E_z}{\partial z} + k_0^2 D_x - \frac{4\pi ik_0}{c} J_x = 0; \tag{35}$$

$$(y) \quad \frac{\partial^2 E_y}{\partial z^2} - k_x^2 E_y + k_x k_y E_x + ik_y \frac{\partial E_z}{\partial z} + k_0^2 D_y - \frac{4\pi ik_0}{c} J_y = 0; \tag{36}$$

$$(z) \quad - (k_x^2 + k_y^2) E_z + k_0^2 D_z + ik_x \frac{\partial E_x}{\partial z} + ik_y \frac{\partial E_y}{\partial z} - \frac{4\pi ik_0}{c} J_z = 0. \tag{37}$$

Denote $k_t^2 \equiv k_x^2 + k_y^2$. Using (37), one can get

$$E_z = - \left(\frac{ik_x}{k_0^2} \frac{1}{\varepsilon_{33} - \frac{k_x^2}{k_0^2}} \frac{\partial E_x}{\partial z} + \frac{ik_y}{k_0^2} \frac{1}{\varepsilon_{33} - \frac{k_y^2}{k_0^2}} \frac{\partial E_y}{\partial z} + \frac{\varepsilon_{33} E_x}{\varepsilon_{33} - \frac{k_x^2}{k_0^2}} + \frac{\varepsilon_{32} E_y}{\varepsilon_{33} - \frac{k_y^2}{k_0^2}} - \frac{4\pi i J_z}{\omega} \right). \tag{38}$$

Put (38) into (35) and (36). Denote $\varepsilon_{33t} \equiv \varepsilon_{33} - \frac{k_t^2}{k_0^2}$. One can get:

$$\begin{aligned} \frac{\partial}{\partial z} \left[\frac{\varepsilon_{33} - k_y^2/k_0^2}{\varepsilon_{33t}} \frac{\partial E_x}{\partial z} \right] + \frac{k_x k_y}{k_0^2} \frac{\partial}{\partial z} \left(\frac{1}{\varepsilon_{33t}} \frac{\partial E_y}{\partial z} \right) - ik_x \frac{\partial}{\partial z} \left(\frac{\varepsilon_{31} E_x}{\varepsilon_{33t}} \right) - ik_x \frac{\partial}{\partial z} \left(\frac{\varepsilon_{31} E_x}{\varepsilon_{33t}} \right) - ik_x \frac{\partial}{\partial z} \left(\frac{\varepsilon_{32} E_y}{\varepsilon_{33t}} \right) \\ - \frac{ik_x \varepsilon_{13}}{\varepsilon_{33t}} \frac{\partial E_x}{\partial z} - \frac{ik_y \varepsilon_{13}}{\varepsilon_{33t}} \frac{\partial E_y}{\partial z} + \left[k_0^2 \left(\varepsilon_{11} - \frac{\varepsilon_{13} \varepsilon_{31}}{\varepsilon_{33t}} \right) - k_y^2 \right] E_x + \left[k_0^2 \left(\varepsilon_{12} - \frac{\varepsilon_{13} \varepsilon_{32}}{\varepsilon_{33t}} \right) - k_x^2 \right] E_y \\ + \left[k_0^2 \left(\varepsilon_{11} - \frac{\varepsilon_{13} \varepsilon_{31}}{\varepsilon_{33t}} \right) - k_y^2 \right] E_x + \left[k_0^2 \left(\varepsilon_{12} - \frac{\varepsilon_{13} \varepsilon_{32}}{\varepsilon_{33t}} \right) + k_x k_y \right] E_y - \frac{4\pi k_x}{\omega} \frac{\partial}{\partial z} \left(\frac{J_z}{\varepsilon_{33t}} \right) + \frac{4\pi ik_0 \varepsilon_{13}}{\omega} \frac{J_z}{\varepsilon_{33t}} - \frac{4\pi ik_0}{c} J_x = 0; \end{aligned} \tag{39}$$

$$\begin{aligned} \frac{\partial}{\partial z} \left[\frac{\varepsilon_{33} - k_x^2/k_0^2}{\varepsilon_{33t}} \frac{\partial E_y}{\partial z} \right] + \frac{k_x k_y}{k_0^2} \frac{\partial}{\partial z} \left(\frac{1}{\varepsilon_{33t}} \frac{\partial E_x}{\partial z} \right) - ik_y \frac{\partial}{\partial z} \left(\frac{\varepsilon_{31} E_x}{\varepsilon_{33t}} \right) - \frac{ik_x \varepsilon_{23}}{\varepsilon_{33t}} \frac{\partial E_x}{\partial z} - \frac{ik_y \varepsilon_{23}}{\varepsilon_{33t}} \frac{\partial E_y}{\partial z} \\ + \left[k_0^2 \left(\varepsilon_{22} - \frac{\varepsilon_{23} \varepsilon_{32}}{\varepsilon_{33t}} \right) - k_x^2 \right] E_x + \left[k_0^2 \left(\varepsilon_{21} - \frac{\varepsilon_{23} \varepsilon_{31}}{\varepsilon_{33t}} \right) + k_x k_y \right] E_y - \frac{4\pi k_y}{\omega} \frac{\partial}{\partial z} \left(\frac{J_z}{\varepsilon_{33t}} \right) + \frac{4\pi k_0 \varepsilon_{23}}{\omega} \frac{J_z}{\varepsilon_{33t}} - \frac{4\pi ik_0}{c} J_y = 0; \end{aligned} \tag{40}$$

Suppose that the medium above upper boundary $z=L_z$ is homogeneous layer F with parameters which do not depend on z) and the appropriation $\varepsilon_3 \rightarrow \infty$ (where ε_3 is a component of tensor along axis z').

$$\frac{\partial E_x}{\partial z} + m_{11}E_x + m_{12}E_y = 0; \quad (41)$$

$$\frac{\partial E_y}{\partial z} + m_{21}E_x + m_{22}E_y = 0; \quad (42)$$

where

$$m_{11} = \frac{i}{\Delta}(k_{zar} - \alpha_r \beta_r k_{zsr}); \quad m_{12} = \frac{i\beta_r}{\Delta}(k_{zsr} - k_{zar}); \\ m_{21} = \frac{i\alpha_r}{\Delta}(k_{zar} - k_{zsr}); \quad m_{22} = \frac{i}{\Delta}(k_{zsr} - \alpha_r \beta_r k_{zar}). \quad (43)$$

In (43)

$$\beta_r = -\frac{k_y}{k_x + k_{zMSW} \operatorname{tg} \theta}; \quad \alpha_i = \frac{k_y}{k_x + k'_{zAW} \sin \theta}; \quad k''_{zAW} > 0; \\ k_{zMSW} = \sqrt{k_0^2 \varepsilon_1 - k_r^2}; \quad k''_{zMSW} < 0; \\ k'_{zAW} \simeq k_0 \sqrt{\varepsilon_1}; \quad \Delta = 1 - \alpha_r \beta_r; \quad \alpha_r = \frac{k_y}{k_x + k'_{zAW} \sin \theta}; \quad k''_{zAW} < 0; \quad (44)$$

In (44), the choice of the signs of the imaginary parts (denoted with using of two primes) of the wavenumbers $k_{AW,MSW}$ in the media with finite losses determines the corresponding wavenumbers as a whole. These signs are chosen in a way, corresponding to the radiation/losses of the corresponding waves in the direction “+ z ”, in other words to the magnetosphere from the upper boundary $z=L_z$. The layer F is anisotropic and non-gyrotropic. The tensor $\hat{\varepsilon}$ in a layer F in a coordinate system x', y', z' has a form (Bryunelli and Namgaladze 1988):

$$\hat{\varepsilon} = \begin{pmatrix} \varepsilon_1 & 0 & 0 \\ 0 & \varepsilon_1 & 0 \\ 0 & 0 & \varepsilon_3 \end{pmatrix} \quad (45)$$

and the value ε_1 included into relations (44) is an element of the tensor (45). Finally, one held to solve the system of Eqs. (39), (40) with the boundary conditions (41), (42) with the coefficients (43). This solution is performed using matrix sweep method (Samarskii 2001; Samarskii and Nikolaev 1989).

References

- Al'pert YL (1972) Propagation of electromagnetic waves and the ionosphere. Nauka Publishing House, Moscow, p 563 (in Russian)
- Alexandrov AF, Bogdankevich LS, Rukhadze AA (1984) Principles of plasma electrodynamics. Springer, Berlin
- Alperovich LS, Fedorov EN (2007) Hydromagnetic waves in the magnetosphere and the ionosphere. Springer, Dordrecht, p 426
- Basu B (2002) On the linear theory of equatorial plasma instability: comparison of different description. J Geophys Res. <https://doi.org/10.1029/2001ja000317>
- Boardman AD, Rapoport YG, Grimalsky VV, Ivanov BA, Koshevaya SV, Velasco L, Zaspel CE (2005) Excitation of vortices using linear and nonlinear magnetostatic waves. Phys Rev E. <https://doi.org/10.1103/physreve.71.026614>
- Boardman AD, Alberucci A, Assanto G, Grimalsky VV, Kibler B, McNiff J, Nefedov I, Rapoport YG, Valagiannopoulos CA (2017) Waves in hyperbolic and double negative metamaterials including rogues and solitons. Nanotechnology 28(44):444001. <https://doi.org/10.1088/1361-6528/aa6792>
- Boardman AD, Rapoport YG, Aznakayeva DE, Aznakayev EG, Grimalsky VV, Mei Z (eds) (2019) Graphene metamaterial electron optics: excitation processes and electro-optical modulation: handbook of graphene, vol 3, Chapter 9, pp 264–296. Wiley Publications, Hoboken, p 492
- Bryunelli BE, Namgaladze AA (1988) Physics of the ionosphere. Nauka, Moscow, p 527 (in Russian)
- Cheremnykh OK, Klimov SI, Korepanov VE, Koshovy VV, Melnik MO, Ivantyshyn OL, Mezentsev VP, Nogach RT, Rapoport YG, Selivanov YA, Semenov LP (2014) Ground-space experiment for artificial acoustic modification of ionosphere. Some preliminary results. Space Sci Technol 20:60–74. <https://doi.org/10.15407/knit2014.06.060> (in Russian)
- Cheremnykh OK, Grimalsky VV, Ivchenko VN, Koshoviy VV, Mezentsev VP, Melnik ME, Ivantishin OL, Nogach RT, Rapoport YuG, Selivanov YuA (2015) Experimental and theoretical investigations of an artificial acoustic modification of the atmosphere and ionosphere. Space Sci Technol 21(1):48–53 (in Ukrainian)
- Collin RE (1991) Field theory of guided wave. IEEE Press, New York, p 852
- Denisenko VV, Pomozov EV (2010) Penetration of electric fields from the atmospheric boundary layer into the ionosphere. Solnechno-Zemnaya Fiz 16:70–75
- Denisenko VV, Boudjada MY, Horn M, Pomozov EV, Biernat HK, Schwingschuh K, Lammer H, Prattes G, Cristea E (2008) Ionospheric conductivity effects on electrostatic field penetration into the ionosphere. Natl Hazards Earth Syst Sci 8(5):1009–1017. <https://doi.org/10.5194/nhess-8-1009-2008>
- Denisenko VV, Ampferer M, Pomozov EV, Kitaev AV, Hausleitner W, Stangl G, Biernat HK (2013) On electric field penetration from ground into the ionosphere. J Atmos Solar Terr Phys 102:341–353. <https://doi.org/10.1016/j.jastp.2013.05.019>
- Emelyanov LY, Zivolup TG, Soroka SA, Cheremnykh OK, Chernogor LF (2015) Ground acoustic influence on the ionosphere: the results of observations by means of incoherent scattering and vertical sounding the methods. Radio Phys Radio Astron 20:37–47 (in Russian)
- Fedorenko AK, Lizunov GV, Rotkel X (2005) Satellite observations of quasiwave perturbations of the atmosphere at the altitudes of the region F caused by strong earthquakes. Geomagn Aeron 45(3):403–410 (in Russian)
- Fedorenko AK, Bespalova AV, Cheremnykh OK, Kryuchkov EI (2015) A dominant acoustic-gravity mode in the polar thermosphere. Ann Geophys 33:101–108
- Fedorenko AK, Kryuchkov EI, Cheremnykh OK, Klymenko YO, Yampolski YM (2018) Peculiarities of acoustic-gravity waves in inhomogeneous flows of the polar thermosphere. J Atmos Solar Terr Phys. <https://doi.org/10.1016/j.jastp.2018.05.009>
- Genkin LG, Erukhomov LM, Myasnikov EN, Shvartz MM (1987) To the question of formation and floating of isothermic ionospheric

- and chromospheric bubbles. In: Proceedings of the higher education institutes. Radiophysics, vol. 30, pp 567–577 (in Russian)
- Gokhberg MB, Shalimov SL (2000) Lithosphere–ionosphere coupling and its modeling. *Russ J Earth Sci* 2:95–108 (in Russian)
- Gokhberg MB, Nekrasov AK, Shalimov SL (1996) On the influence of unstable release of green-house gases in the seismically active regions on the ionosphere. *Phys Earth* 8:52–55 (in Russian)
- Grimalsky VV, Rapoport YG (2000) Penetration of the electrostatic field from the near-Earth sources into the lower ionospheric layers. *Kinemat Phys Celest Bodies* 16(1):7–13
- Grimalsky VV, Kremenetsky IA, Rapoport YuG (1999) Excitation of electromagnetic waves in the lithosphere and their penetration into ionosphere and magnetosphere. *J Atmos Electr* 19(2):101–117
- Grimalsky VV, Hayakawa M, Ivchenko VN, Rapoport YuG, Zadoroznii VI (2003) Penetration of electrostatic field from the lithosphere into the ionosphere and its effect on the D-region before earthquake. *JASTP* 65(4):391–407
- Gurevich AV (1978) *Nonlinear phenomena in the ionosphere*. Springer, New York, p 356
- Gurevich AV, Shvartzburg AB (1973) *Nonlinear theory of propagation of radiowaves in the ionosphere*. Nauka, Moscow, p 272 (in Russian)
- Hooke WH (1968) Ionospheric irregularities produced by internal atmospheric gravity waves. *J Atmos Terr Phys* 30:795–823
- Huang CS, Kelley C (1996a) Nonlinear evolution of equatorial spread of F1. On the role of plasma instabilities and spatial resonance associated with gravity wave seeding. *J Geophys Res* 101(A1):283–292
- Huang CS, Kelley C (1996b) Nonlinear evolution of equatorial spread of F2. Gravity wave seeding of Rayleigh–Taylor instability. *J Geophys Res* A101(A1):293–302
- Huang CS, Miller CA, Kelley MC (1994) Basic properties and gravity wave initiation of the midlatitude F region instability. *Radio Sci* 29:395–405
- Iyemori T, Nakanishi K, Aoyama T, Yokoyama Y, Koyama Y, Luhr H (2015) Confirmation of existence of the small-scale field-aligned currents in middle and low latitudes and an estimate of time scale of their temporal variation. *Geophys Res Lett* 42:22–28. <https://doi.org/10.1002/2014gl062555>
- Jursa AS (ed) (1985) *Handbook of geophysics and the space environment*. Air Force Geophysics Laboratory, Ohio, p 1042
- Kelley MC (2009) *The Earth's ionosphere: plasma physics and electrodynamics*. Elsevier, Amsterdam, p 551
- Kendall PC, Pickering WM (1967) Magnetoplasma diffusion at F2-region altitudes. *Planet Space Sci* 15:825–833
- Klimenko MV, Klimenko VV, Zakharenkova IE, Pulinets SA, Zhao B, Tsidilina MN (2011) Formation mechanism of great positive TEC disturbances prior to Wenchuan earthquake on May 12, 2008. *Adv Space Res* 48:488–499
- Koshevaya S, Burlak G, Grimalsky V, Perez-Enriquez R, Kotsarenko A (2002) Nonlinear transformation of seismic waves into ULF atmospheric acoustic waves, systems analysis modeling simulations. *Gordon Breach J* 42:261–268
- Koshovy V, Nazarchuk Z, Romanyshyn I, Ivantyshyn O, Lozynskyy A, Soroka S, Alyokhina IL (2005) Acousto-electromagnetic investigations of an acoustical channel of the lithosphere–ionosphere interaction. In: Proceedings of the XXVIIIth general assembly of international union of radio science (URSI), October 23–29, 2005, New Delhi, India
- Kotsarenko NY, Rapoport YG, Redcoborody YN, Shvidkij AA (1994) Instability of acoustic-gravity waves in an atmosphere with impurities which release and absorb heat. *Kinemat Phys Celest Bodies* 10(1):61–64
- Kotsarenko N, Soroka S, Koshevaya S, Koshovy V (1999) Increase of the transparency of the ionosphere for cosmic radiowaves caused by a low frequency wave. *Phys Scr* 59:174–181
- Kuo CL, Huba JD, Joyce G, Lee LC (2011) Ionosphere plasma bubbles and density variations induced by pre-earthquake rock currents and associated surface charges. *J Geophys Res* 116:A10317. <https://doi.org/10.1029/2011ja016628>
- Lyatsky VB, Maltsev YP (1983) *The magnetosphere–ionosphere interaction*. Nauka, Moscow, p 192 (in Russian)
- Molchanov OA, Hayakawa M, Rafalsky VA (1995) Penetration characteristics of electromagnetic emissions from an underground seismic source into the atmosphere, ionosphere, and magnetosphere. *J Geophys Res* 100A:1691–1712
- Nenovski PI, Pezzopane M, Ciraolo L, Vellante M, Villante U, De Lauretis M (2015) Local changes in the total electron content immediately before the 2009 Abruzzo earthquake. *Adv Space Res* 55(1):243–258
- Nina A, Čadež VM (2013) Detection of acoustic-gravity waves in lower ionosphere by VLF radio waves. *Geophys Res Lett* 40(18):4803–4807. <https://doi.org/10.1002/grl.5093>
- Oikonomou C, Haralambous H, Muslim B (2016) Investigation of ionospheric TEC precursors related to the M7.8 Nepal and M8.3 Chile earthquakes in 2015 based on spectral and statistical analysis. *Nat Hazards* 83(S1):97–116. <https://doi.org/10.1007/s11069-016-2409-7>
- Press WH, Teukolsky SA, Vetterling WT, Flannery BA (1997) *Numerical recipes in Fortran 77*. Press Syndicate University of Cambridge, New York, p 973
- Pulinets S (2009) Physical mechanism of the vertical electric field generation over active tectonic faults. *Adv Space Res* 44:767–773
- Pulinets S (2011) The synergy of earthquake precursors. *Earth Sci* 24:535–548
- Pulinets S, Boyarchuk K (2005) *Ionospheric precursors of earthquakes*. Springer, Berlin, p 315P
- Pulinets SA, Boyarchuk KA, Hegai VV, Kim VP, Lomonosov AM (2000) Quasielectrostatic model of atmosphere–thermosphere–ionosphere coupling. *Adv Space Res* 26(8):1209–1218
- Pulinets S, Ouzounov D, Papadopoulos G, Rozhnoi A, Kafatos M, Taylor P, Anagnostopoulos G (2011) Multi-parameter precursory activity before L'Aquila earthquake revealed by joint satellite and ground observations American Geophysical Union Fall Meeting, AGU2011, 5–9 December, San Francisco, USA
- Pulinets SA, Ouzounov D, Davidenko D (2014) Whether an earthquake prediction is possible? Integral technologies of multiparametric monitoring of geoeffective phenomena in the frames of integral model of coupling in the earth atmosphere and ionosphere, Trovant. https://doi.org/10.17686/sced_rusnauka_2014-1211. https://www.researchgate.net/publication/279058620_Prognoz_zemletrasenij_vozmozen_Integralnye_tehnologii_mnogoparametricheskogo_monitoringa_geoeffektivnyh_avlenij_v_ramkah_kompleksnoj_modeli_vzaimosvazej_v_litofere_atmosfere_i_ionosfere_Zemli (in Russian)
- Pulinets SA, Ouzounov DP, Karelin AV, Davidenko DV (2015) Physical bases of the generation of short-term earthquake precursors: a complex model of ionization-induced geophysical processes in the lithosphere–atmosphere–ionosphere–magnetosphere system. *Geomag Aeron* 55(4):540–558
- Rapoport Yu, Hayakawa M, Ivchenko VN, Juarez-R D, Koshevaya S, Gotynyan O (2004a) Change of ionospheric plasma parameters under the influence of electric field which has lithospheric origin and due to radon emanation. *Phys Chem Earth* 29:579–587
- Rapoport YuG, Gotynyan OE, Ivchenko VN, Kozak LV, Parrot M (2004b) Effect of acoustic-gravity wave of the lithospheric origin on the ionospheric F region before earthquakes. *Phys Chem Earth* 29:607–616

- Rapoport YG, Gotynyan OE, Ivchenko VN (2005) Ionosphere response and instabilities in the presence of transformation of energy from the lower atmosphere through the quasioleostatic and acoustic-gravity channels. In: Fifth Ukrainian conference on space research, 4–11 September 2005, Eupatorium, p 53 (in Russian)
- Rapoport YG, Gotynyan OE, Ivchenko VN, Hayakawa M, Grimalsky VV, Koshevaya SV, Juarez RD (2006) Modeling electrostatic-photochemistry seismoionospheric coupling in the presence of external currents. *Phys Chem Earth Parts A/B/C* 31(4–9):437–446. <https://doi.org/10.1016/j.pce.2006.02.010>
- Rapoport YuG, Hayakawa M, Gotynyan OE et al (2009) Stable and unstable plasma perturbations in the ionospheric F region, caused by spatial packet of atmospheric gravity waves. *Phys Chem Earth* 34:508–515
- Rapoport Y, Selivanov Y, Ivchenko V, Grimalsky V, Fedun V (2011) UK-Ukraine-spain meeting on solar physics and space science, 2011, Alushta, Crimea; Ukraine, August 29-Sept. 2, 2011, p 55
- Rapoport YG, Cheremnykh OK, Selivanov YA, Fedorenko AK, Ivchenko VM, Grimalsky VV, Tkachenko EN (2012a) Modeling AGW and PEMW in inhomogeneous atmosphere and ionosphere. In: 14th International conference on mathematical methods in electromagnetic theory, MMET 2012, August 2012, Kharkiv; Ukraine. IEEE, pp 28–30
- Rapoport Y, Boardman A, Grimalsky V, Selivanov Y, Kalinich N (2012b) Metamaterials for space physics and the new method for modeling isotropic and hyperbolic nonlinear concentrators. In: 14th International conference on mathematical methods in electromagnetic theory, MMET 2012, August 2012, Kharkiv, Ukraine. IEEE, pp 76–78
- Rapoport Y, Boardman A, Grimalsky V, Selivanov Y, Kalinich N (2012c) Metamaterials for space physics and the new method for modeling isotropic and hyperbolic nonlinear concentrators. In: 2012 international conference on mathematical methods in electromagnetic theory. <https://doi.org/10.1109/mmet.2012.6331154>
- Rapoport Yu, Selivanov Yu, Ivchenko V, Grimalsky V, Tkachenko E, Fedun V (2014a) Excitation of planetary electromagnetic waves in the inhomogeneous ionosphere. *Ann Geophys* 32:1–15
- Rapoport YG, Cheremnykh OK, Grimalsky VV, Selivanov YA, Ivchenko VN, Milinevsky GP, Tkachenko EN, Melnik MO, Mezentsev VP, Karataeva LM, Nogach RT (2014b) Ionosphere as a sensitive indicator of powerful processes in the lower atmosphere/lithosphere, artificial acoustic influence and space weather. In: EMSEV 2014 international conference on electromagnetic phenomena associated with seismic and volcanic activities, Konstancin Jeziorna, Poland, on September 21–27, 2014, Conference Paper book, pp 140–142
- Rapoport Y, Selivanov Y, Ivchenko V, Grimalsky V, Tkachenko E, Rozhnoi A, Fedun V (2014c) Excitation of planetary electromagnetic waves in the inhomogeneous ionosphere. *Ann Geophys* 32(4):449–463. <https://doi.org/10.5194/angeo-32-449-2014>
- Rapoport YG, Grimalsky VV, Boardman AD, Malnev VN (2014d) Controlling nonlinear wave structures in layered metamaterial, gyrotropic and active media. In: 2014 IEEE 34th international scientific conference on electronics and nanotechnology (ELNANO). <https://doi.org/10.1109/elnano.2014.6873959>
- Rapoport YG, Boardman AD, Grimalsky VV, Ivchenko VM, Kalinich N (2014e) Strong nonlinear focusing of light in nonlinearly controlled electromagnetic active metamaterial field concentrators. *J Opt* 16(5):055202. <https://doi.org/10.1088/2040-8978/16/5/055202>
- Rapoport YuG, Cheremnykh OK, Koshoviy VV, Melnik MO, Ivantshyn OL, Nogach RT, Selivanov YuA, Grimalsky VV, Mezentsev VP, Karataeva LM, Ivchenko VM, Milinevsky GP, Fedun VN, Tkachenko EN (2017) Ground-based acoustic parametric generator impact on the atmosphere and ionosphere in an active experiment. *Ann Geophys* 35:53–70
- Rozhnoi A, Solovieva M, Levin B, Hayakawa M, Fedun V (2014a) Meteorological effects in the lower ionosphere as based on VLF/LF signal observations. *Natl Hazards Earth Syst Sci* 14(10):2671–2679. <https://doi.org/10.5194/nhess-14-2671-2014>
- Rozhnoi A, Shalimov S, Solovieva M, Levin B, Shevchenko G, Hayakawa M, Fedun V (2014b) Detection of tsunami-driven phase and amplitude perturbations of subionospheric VLF signals following the 2010 Chile earthquake. *J Geophys Res Space Phys* 119(6):5012–5019. <https://doi.org/10.1002/2014ja019766>
- Rozhnoi A, Solovieva M, Parrot M, Hayakawa M, Biagi P-F, Schwingschuh K, Fedun V (2015) VLF/LF signal studies of the ionospheric response to strong seismic activity in the Far Eastern region combining the DEMETER and ground-based observations. *Phys Chem Earth Parts A/B/C* 85–86:141–149. <https://doi.org/10.1016/j.pce.2015.02.005>
- Samarskii AA (2001) The theory of difference schemes. Marcel Dekker, New York, p 761
- Samarskii AA, Nikolaev ES (1989) Numerical methods for grid equations, pp 1–2. Birkhäuser (Translated from Russian)
- Sanchez-Dulcet F, Rodríguez-Bouza M, Silva HG, Herraiz M, Bezzeghoud M, Biagi PF (2015) Analysis of observations backing up the existence of VLF and ionospheric TEC anomalies before the Mw6.1 earthquake in Greece, January 26, 2014. *Phys Chem Earth Parts A/B/C* 85–86:150–166. <https://doi.org/10.1016/j.pce.2015.07.002>
- Schunk RW, Nagy AF (2010) Ionospheres. Physics, Plasma physics, and chemistry. CUP, Cambridge, p 628
- Sedunov YS, Avdiushin SI, Borisenkov EP et al (eds)(1991) Atmosphere. Handbook, Gidrometeoizdat, Leningrad, p 509 (in Russian)
- Sekar R, Raghavarao R (1987) Role of vertical winds on the Rayleigh-Taylor mode instabilities of the night-time equatorial ionosphere. *J Atmos Terr Phys* 49:981–985
- Solovieva M, Rozhnoi A, Fedun V, Schwingschuh K, Hayakawa M (2015) Ionospheric perturbations related to the earthquake in Vrancea area on November 22, 2014, as detected by electromagnetic VLF/LF frequency signals. *Ann Geophys* 58(5):A0552. <https://doi.org/10.4401/ag-6827>
- Sorokin V, Hayakawa M (2013) Generation of seismic-related DC electric fields and lithosphere–atmosphere–ionosphere coupling. *Mod Appl Sci* 7:1–25
- Sorokin V, Chmyrev V, Isaev N (1998) A generation model of small-scale geomagnetic field-aligned plasma inhomogeneities in the ionosphere. *J Atmos Solar Terres Phys* 60(13):1331–1342. [https://doi.org/10.1016/s1364-6826\(98\)00078-9](https://doi.org/10.1016/s1364-6826(98)00078-9)
- Sorokin VM, Chmyrev VM, Yashenko AK (2001) Electrodynamical model of the lower atmosphere and the ionosphere. *JASTP* 63:681–1691
- Spiegel MR (1959) Theory and problems of vector analysis and an introduction to tensor analysis. Shnaum Publications, New York, p 223
- Stangl G, Boudjada MY, Biagi PF, Krauss S, Maier A, Schwingschuh K, Al-Haddad E, Parrot M, Voller W (2011) Investigation of TEC and VLF space measurements associated to L'Aquila (Italy) earthquakes. *Natl Hazards Earth Syst Sci* 11:1019–1024
- Treumann RA, Baumjohann W (1997) Advanced space plasma physics. Imperial College Press, London, p 381
- Vainshtein LA (1988) Electromagnetic waves. Nauka (in Russian), Moscow
- Zettergren MD, Snively JB (2013) Ionospheric signatures of acoustic waves generated by transient tropospheric forcing. *Geophys Res Lett* 40:5345–5349. <https://doi.org/10.1002/2013gl058018>
- Zettergren MD, Snively JB (2015) Ionospheric response to infrasonic-acoustic waves generated by natural hazard events. *J Geophys Res (Space Phys)*. <https://doi.org/10.1002/2015ja021116>



Acknowledgement of Reviewers for 2019

Petra van Steenbergen¹

Published online: 29 January 2020

© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2020

We thank all those scientists and experts in the various fields represented in *Acta Geophysica* for devoting their time and effort to review the papers that we have sent them. The Editor in Chief, and the Publisher acknowledge the colleagues listed below for their most helpful reviews of papers which have been provided during the period January 2019 to December 2019.

Mubarik Ali	Mohammad Khaled Akhtar	Alireza Arabameri
Victor Aarre	John-Felix Akinbami	Mustafa Aral
Ivan Abakumov	Halil Akinci	Frederic Arand
Abdulaziz M. Abdulaziz	Mohammed Al Yafei	David Archer
Karimatu Lami Abdullahi	Hakan Alçık	Mehran Arian
Sumiyoshi Abe	Maria Joao Alcoforado	Feza Arikan
Jahangir Abedi-Koupai	Lorenzo Alfieri	Keisuke Ariyoshi
Jochen Aberle	Syed Shahid Ali	Endrias Asgedom
O. Abiola	Tariq Alkhalifah	Siavash Ashoori
Raymond Abma	Mostafa AllamehZadeh	Aysegul Askan
Ahmed E. Abouelregal	Mansour Almazroui	Elvira I. Astafyeva
Yousf Abushalah	Ibrahim A. Almosallam	Ariana Luci Astorga
Sirisha Adamala	David Altadil	Barbara Atamaniuk
Jan Adamowski	Orhan Altan	Mustafa Mujdat Atanak
Alice Adenis	Nezih Altay	Rahman Atta-ur
Tjahyo Nugroho Adji	Luis Alva Valdivia	Pascal Audet
Muhammad Adnan	Hakimeh Amanipoor	Esben Auken
Qazi Adnan Ahmad	A. Amendola	Christine Authemayou
M. H. Afshar	John Amigun	Dmitry Avdeev
Amir AghaKouchak	Babak Amirataee	Anna Avdeeva
Afshin Aghayan	Angelo Amorosi	Mekonen Ayana
Babak Aghel	Amsalu Y. Anagaw	Naser Ayat
Zacharias Agioutantis	Emmanouil Anagnostou	Hossein Azizi
Mehdi Aharchaou	Marios Anagnostou	Alper Baba
Mobin-ud-Din Ahmad	Anca Radulescu	Golekar Rushikesh Baburao
Waqar Ahmad	Amalia Anderson	Jose Badal
Ali Ahmadalipour	Karin Andreassen	Marcela Elvira Baena Rivera
Hassan Ahmadi	Gina Andrei	Nooshin Bagha
Alireza Ahmadyfard	Jesse Andries	Majid Bagheri
Jamilu Bala Ahmed II	D. A. Angus	Dov Bahat
Valentin Aich	Marco Anzidei	Christopher Baiyegunhi
Mehdi Akhoondzadeh	George Apostolopoulos	Marianna Balasco

✉ Petra van Steenbergen
petra.vansteenbergen@springer.com

¹ Dordrecht, The Netherlands

Huang Baochun	Simone Bizzi	David Carpenter
Simone Barani	Viktória Blanka	Armando Carrillo-Vargas
Sandor Baranya	Lilian Blaser	Fernando Carvajal-Ramírez
Reza Barati	Leszek Blaszkiewicz	John P. Castagna
Gabriel Barberes	Jan Blecki	Silvia Castellaro
Udo Barckhausen	Fred Boadu	Andrea Castelletti
Adrian S. Barfod	Nikita Bobrov	Raúl R. Castro
Brandon Barkey	Heike Bock	Marcilio Castro de Matos
Eric Barrier	Ewa Bogdanowicz	Alessandro Cattapan
Roland Barthel	Jedrzej Bojanowski	Carlo Cauzzi
Rahim Barzegar	Bodo Bookhagen	Daniel Caviedes-Voullième
Jelena Basarić	David Boore	Alberto Cazzella
Stefano Basso	Angana Borah	Semet Celik
Paul Bates	Marco Borga	A. M. Chandler
George Bathrellos	José Fernando Borges	Vishal Chandole
Joelson Batista	Ronaldo Borja	Emmy Chang
Pedro V. G. Batista	Jannis Born	Liang-Cheng Chang
Rajan Batta	Pasquale Borrelli	Ping-Yu Chang
Michael L. Batzle	Maren Böse	Da-Yi Chen
D. Bau	Ada Vittoria Bosisio	Jin Chen
Stanisław Baudzis	Josef Boska	Ke Chen
Carlos Bautista-Capetillo	Gianluca Botter	Ling Chen
Paolo Bazzurro	Hamza Bouguerra	Shengbo Chen
Hylke Beck	Claire Bouligand	Shengchang Chen
L. Becker	Polyzois Bountzis	Yangkang Chen
Simon Beckouche	Housseyn Bouzeria	You-Cheng Chen
Jonathan Bedford	Heinrich Brasse	Yuanfang Chen
J. Bedi	Rebecca Briant	Ismail Chenini
Michael Behm	Alistair R. Brown	Piyaphong Chenrai
Linus Bengtsson	Holly M. Brown	Mohsen Cheraghi
Martin Beniston	Dominik Brunner	Dmitri B. Chklovskii
Djillali Benouar	Manuela Brunner	Hsiao Chi Chuang
Jens Berdermann	Pace Bruno	Chen Chieh-Hung
Francoise Bergerat	Axel Bruns	Francesco Chierici
Sten Bergstrom	Sander Bruun	Karem Chokmani
James G. Berryman	S. V. Bukharin	Richard Chopping
Falko Bethmann	Dalia Buresova	Konstantinos Chousianitis
Vittal Bhatt	Gerd Bürger	Jinder Chow
Shuvajit Bhattacharya	Roland Bürgmann	Jasti S Chowdary
Muhammad Tousif Bhatti	Vyacheslav Burov	Vasileios Christelis
Hussain Bux Bhutto	Michael Bush	Craig William Christensen
P. K. Bhuyan	Martina Bussettini	Maria Christodoulou
Huaxing Bi	P. Butlin	Hone-Jay Chu
Pier Francesco Biagi	Jimmy Byakatonda	Artur Cichowicz
Robert Bialik	Joanna Bzówka	Consuelo Cid
Vanessa Biando Ribeira	Tommaso Caloiero	Luigi Cimorelli
Ramón Bienes	A. Campisano	Begüm Civgin
Dieter Bilitza	Iza Canales García	Riccardo Civico
Ahmad Safuan Bin A Rashid	Ljiljana Cander	Pierluigi Claps
Dino Bindi	Jingjie Cao	Elizabeth S Cochran
Peter Bird	Philippe Caprioli	Richard E. LL. Collier
Arkoprovo Biswas	Roberto Carluccio	Andrea Cominola
Andrea Bizzarri	Chris Caroni	Christian Conoscenti

Rodolfo Console	Matteo Demurtas	Maurizio Fedi
R. Corthesy	Muhammet Deveci	Carozzi Fernanda
Giacomo Corti	Jacques Déverchère	Vito Ferro
Sara Corvaro	Roberto Devoti	Gianluca Fiandaca
Stephen Costabel	Subhasish Dey	Stephen Fick
Alex Costall	Amit Dhorde	F. Fiedrich
Carmine Covelli	Dirk Diederer	Julie S. Field
Corina Coviaga	Panayiotis Dimitriadis	Peter Fiener
Stephan Creelle	Radu Georgiu Dimitriu	Ilias Fikos
Gaetano Crispino	Savka Dineva	Renato Filjar
Alessandra Crosato	Pinbo Ding	Ruhul Firdaus
Russell Crosbie	Antonino D’Ippolito	Ms Fiza Sarwar
Brendan W. Crowell	Son Doan	Jan Fleckenstein
Marcin Cudny	Mihály Dobróka	Michael Floyd
Armando Cuéllar	Longjun Dong	Stéphane Follain
Gloria Curilem	Sheen Dong-Hoon	Elizabeth Follett
Joseph Curray	R. D. Dony	Sergey Fomel
Marta Cyz	Mahmut Drahor	Iluis Font
Rafał Czarny	Krzysztof Drozdowicz	Ioanna-Kleoniki Fontara
Zbigniew Czechowski	Changwen Du	J. Fontiela
Federico Da Col	Qingyun Duan	Clive Foss
Adisa Daberdini	Matthieu Dufresne	Dimitra Founda
Nicola D’Agostino	Stanica Dumitru	Paolo Franchin
Berit Dahl	Victor Dupuis	Anna Franczyk
Torsten Dahm	Ved Vyas Dwivedi	Gabriele Freni
Minglong Dai	Tomasz Dysarz	Vladimir Frid
Phil Dales	Beata Dziak-Jankowska	Horst Frischkorn
Irina Dallo	Y. Ebert	Vinay Kumar Gadi
Ali Danandeh Mehr	Mohsen Ebrahimi	Stefano Galelli
Tomasz Danek	Daniele Ehrlich	Ivan Galkin
Diaconu Daniel	Leo Eisner	Chiara Galletti
Andrea D’Aniello	Salamatu Abraham Ekpo	Danilo Galluzzo
Sandipan Das	John Elliott	Attila Galsa
Rachel Davidson	Sherif A. Abu El-Magd	Sreedhar Ganapuram
Ammar Dawood	H. M. El-Sayed	Jinghuai Gao
Shanker Daya	Ahmed El-Shafie	Jianjun Gao
Sikhhendra De	Mustafa Erdik	Huiji Gao
Inge de Graaf	Semih Ergintav	Yanhong Gao
Carmen De Jong	Tomasz Ernst	Pablo Garcia
Mariano Moreno de las Heras	Thebault Erwan	Amanda García Marín
Paolo Marco De Martino	Khalid Essa	Daniel Garcia-Castellanos
Riccardo De Ritiss	Mark Everett	Akhil Garg
Angelo De Santis	Yener Eyüboğlu	Evangelia Garini
Luca de Siena	Licia Faenza	Dariusz Gąsiorowski
Jerzy Dec	Kin Fai Ho	Paolo Gasperini
E. J. van Dedem	Maria Rosaria Falanga	Mohammed Gedefaw
M. Dehbozorgi	Giuseppe Falcone	Nina Gegenhuber
Nicholas Deichmann	Ke Fan	Violette Geissen
Manuel del Jesus	Ying Fan	Alexandra Gemtzi
B. Delouis	Chunhui Fang	Stefania Gentili
Mehmet Demirel	Xinding Fang	Roger Ghanem
Nihan Çetin Demirel	Mohammed Farfour	Aminuddin Ab Ghani
Tufan Demirel	Shahab Fazal	Mohammad Ali Ghannad

Amin Gharehbaghi	Ian Guymer	Christian Huggel
A. Ghasemi	Man Ha Quang	Allen Husker
Mohammad Ghayamghamian	John Bosco Habarulema	Yunjung Hyun
Lubna Ghazal	S. Hadiloo	Ismael M. Ibraheem
Abdul Halim B. Ghazali	Manouchehr Haghighi	Adam Idziak
Ali Gholami	Hojjat Haghshenas Lari	Iunio Iervolino
Mohammad Ghorbani	Bezalel C. Haimson	Olusegun Omoniyi Ige
Gopal Ghosh	Hamid Haitham	Takeshi Iinuma
Sayantana Ghosh	Afera Halefom	Nathan Intrator
Abolfazl Giglou	Hossein Hamidifar	Mansour Ioualalen
Ljubomir Gigović	Hugo Hammer	Mohsen Ahmadzadeh Irandoust
Stuart Gilder	Peng Han	Trevor Irons
Eric Gilleland	Songjun Han	Tsuneo Ishido
Agnieszka Gil-Swidarska	Paul Hardaker	Amir Ismail
Dorota Giritat	Bruce S. Hart	Turgay İşseven
Ruediger Glaser	K. Hasan	Ursula Iturrarán-Viveros
Milan Gocic	Youssef M. A. Hashash	Marian Ivan
Bettina Goertz-Allmann	Hossein Hashemi	Yazad Jabbar
Avto Gogichaishvili	Denis Hatzfeld	James Jackson
Tomisław Gołębiowski	M. Hayakawa	Forough Jafary
Ali Golkarian	Peng He	Alessandro Jaia
Gennady Goloshubin	Gabi Hegerl	Frantisek Janda
Abolfazl Golshan	Oliver Heidbach	Wojciech Jarmolowski
Francisco Gomez	Kosuke Heki	Jadwiga Jarzyna
Javier Gomez	Marcelo Henrique Leao-Santos	Ravidra Jaybhaye
David Gomez-Ortiz	Frank Herrmann	Maciej Jefimow
Xiangbo Gong	Khaled Hessami	Solomon Jekayinfa
Marjan Goodarzi	Richard Hey	Vergne Jérôme
Mustafa Goodarzi	C. Hibert	Madan K. Jha
Kazım Görgülü	Dannie Hidayat	WenHua Ji
Andrzej Górszczyk	Gudrun Hillebrand	Xiaofeng Jia
Ajanta Goswami	Gregor Hillers	Zeyu Jia
Yozo Goto	Ali M. Hinis	Han Jiajun
Teresa Grabowska	Ekkehard Holzbecher	Zhihong Jiang
Dariusz Graczyk	Catherine Homberg	Xinqi Jiao
Roy Grainger	Afshin Honarbakhsh	A. Y. Jillo
Veronica F. Grasso	Sunhak Hong	Hong Jin
Dariusz Grech	Yang Hong	Lai Jin
David R. Groholski	Young Hong Shin	Side Jin
Oksana Grynshyna-Poliuga	Zhang Hongming	Jorge Jódar
Marcin Grzesiak	Sun Hongyu	Ihab Jomaa
Chen Gu	Seyyed Behrouz Hosseyni	Ian Jones
Massimo Guerrero	Matthew D. Howland	Denis Jongmans
Jinyong Gui	Sigrún Hreinsdóttir	Geeta S. Joshi
Laura Gulia	Vagelis Hristidis	Laurence Jouniaux
Tamara L. Gulyaeva	Nai-Chi Hsiao	Waldemar Jozwiak
Thomas Günther	Ta-Chih Hsiao	Niklas Juhojuntti
Leicheng Guo	Bo Huang	Tadeusz Juliszewski
Yuhang Guo	Jianping Huang	Marek Kaczorowski
Ravi P. Gupta	Qinghua Huang	Diethelm Kaiser
Shivam Gupta	Shengzhi Huang	Koosha Kalhor
S. C. Gupta	Weilin Huang	I. I. Kalinnikov
Alexandra Guy	Zhen Huang	John Kalogiros

Thomas Kalscheuer	Sebastian Kowalczyk	Changhong Lin
Osamu Kamigaichi	Richard Krahenbuhl	Helen Lin
Masashi Kamogawa	Vitold Krajieski	Hongbo Lin
Ali Kanli	Adam Krajewski	Yu-Pin Lin
Utku Kânođlu	T. Kronrod	Larry Lines
Vassilis Karakostas	Lech Kryszinski	V. A. Liperovsky
Shawgar Karami	Wenhuan Kuang	Chaofeng Liu
Emilia Karamuz	Andrzej Kułak	Jann-Yenq Liu
Abbaspour Karim	Nagesh Kumar	Jing Liu
Subhankar Karmakar	Prashant Kumar	Lanbo Liu
Altin Karriqi	Rohini Kumar	Mian Liu
Dimitris Kaskaoutis	Zbigniew Kundzewicz	Naihao Liu
Elina Kasvi	Osamu Kuwano	Qiancheng Liu
Dimitris Katsanos	Serdar Kuyuk	Qingrui Liu
Richard W. Katz	H. Serdar Kuyuk	Shaoyong Liu
Rennie B. Kaunda	Jeff Kwang	Wei Liu
Kirsti Kauristi	Annie O. L. Kwok	Xiaobo Liu
Y. Kawada	Eun Hye Kwon	Xiaozhou Liu
J. R. Kayal	John W. Labadie	Yunfeng Liu
Hossein Kazemi	Borys Ladanivskyy	Anastasia Lobanova
Nasser Kazemi	A. V. Ladynin	Anna Łoboda
Saskia Keesstra	Carlo Giovanni Lai	Tobias Lochbühler
Elliott Kellner	Jinxing Lai	Rowena B. Lohman
Ahmed Kenawy	Rattan Lal	Monika Łój
Thomas Kenny	Giovanni Lanzano	Barbara Lolli
Perveiz Khalid	Stefano Lanzoni	Fernando Lopez-Caballero
Qudrat Khan	Christophe Larroque	Nastaran Lotfi
Manoj Khandelwal	Anna Laskownicka	Petroula Louka
E. Alam Kherani	Francisco Latosinski	Samuel Louvet
Malik Sikander Hayat Khiyal	Alexander Lavrov	Jun Lu
Stefan Kienzle	Jesse Lawrence	Yang Lu
Andrzej Kijko	D. Lawton	Eva Ludi
Nazan Kilic	Lazar Lazic	Steve Luis
Hung Soo Kim	H. Le	Jewell Lund
Sungwon Kim	Cheng-Haw Lee	Bin Luo
Peter Kinnell	Jae Min Lee	Jingrui Luo
Marta Kiraga	Saro Lee	Juhua Luo
A. Kiratzi	Maximilien Lehujeur	Song Luo
Kirill Kireyev	Matthieu Lepot	Yiqi Luo
Ozgur Kisi	Andrzej Lesniak	Lucia Luzi
Timea Kiss	Paweł Leśniak	Chao Ma
Bastian Klein	Hanna Levaniemi	Jianquan Ma
Zbigniew Klos	Bailing Li	Ming Ma
Krzysztof Kochanek	Elita Yunyue Li	Yueling Ma
Silvia Kohnova	Fangyu Li	Dave MacLeod
Kory Konsoer	Guilong Li	M. Macouin
Łukasz Kortas	Haipeng Li	Nafiz Maden
Jari Kortström	Jiaxuan Li	Rafael Magallanes-Quintanar
Konrad Kossacki	Jianzhu Li	Luis Fernando Magaña Solís
Alex Kostinski	Junlun Li	Viviana Maggioni
Vladimir Kostoglodov	Rui Li	Rezene Mahatsente
Andrzej Kotyrba	Ruopu Li	Sharafeldin Mahmoud
Svetlana Kovacikova	Zhanqing Li	Sahameddin Mahmoudi Kurdistani

Mojtaba Mahsuli	Mirka Mobilia	Vahid Nourani
Karen Mair	Victor Mocanu	Agata Novara
Mariusz Majdański	Jamshid Moghadasi	Izuchukwu Obiadi
Marija Maksin	Asoodeh Mojtaba	Mathias Obrebski
Luca Malagnini	Roberto S. Molina Garza	Jide Nosakare Ogunbo
Jacques Malavieille	Eugenio Molina Navarro	Philip Oguntunde
Arash Malekian	Juan Carlos Montalvo	Matthias Ohrnberger
Michal Malinowski	Lorena Montoya	Seda Okay
A. A. Malovichko	Seulgi Moon	Taikan Oki
Vivek Manekar	Demie Moore	Ilesanmi Oladapo
Alessandro Mangione	Max Moorkamp	Michailovich Oleg V.
Shattri Mansor	Shahin Moradi	Gerrit Olivier
B. P. Marchant	Isabelle Moretti	Nils Olsen
David Marchant	J. V. Morgan	Per-Ivar Olsson
Clezio Marcos De Nardin	Igor Morozov	Oluwole Omosuyi
Kurt J. Marfurt	Bill Morris	T. Ondoh
Krzysztof Markowicz	Jerzy Moscicki	Mahir Öner
Bertha Marquez Azua	Domiziano Mostacci	Sultan Ceren Öner
Ivan Dimitri Marroquin	Mehdi Mousavi	Victor N. Oparin
J. Marteau	Seyed Yaser Moussavi Alashloo	Emmanuel Opolot
J. D. Martínez-Vargas	Isaac Mugume	Modesto Ortiz Figueroa
Fatima Martin-Hernandez	Ehsan Muhammad Irfan	Mehdi Ostadhassan
Marta Martinkova	Babel Mukand	Marzena Osuch
Cristina Martín-Moreno	Birgit Müller	Gengxin Ou
Christian Massari	Asmita Murukumar	Dimitar Ouzounov
Demetrios Matsakis	Michael Musgrove	Elijah Oyeyemi
Elena Matta	Giovanni Musolino	A. Arda Ozacar
M. Mattei	Syed Md. Touhidul Mustafa	Özden Özdemir
Satya Maurya	L. P. Muthuwatta	Amalio F. Pacheco
Stefan Maus	Grzegorz Mutke	Francesca Pacor
Jessie Mayor	Syed Fahad Naeem	ellora padhi
Maria Maza	Norihiro Nakamura	John Paine
Andrzej Mazur	Jarek Napiorkowski	Debasish Pal
Timothy McClanahan	Balaji Narasimhan	Subodh Chandra Pal
Dan McEvoy	Nadine Nassif	Michele Palermo
Helen McFarlane	Jolanta Nastula	Athanasios D. Panagopoulos
Maurice J. Mchugh	F. Alejandro Nava Pichardo	Panos Panagos
Daniel E. McNamara	Ana D. Navarro-Ortega	D. Panagoulia
Brian J. O. L. Mcpherson	Adeel Nazeer	Biranchi Panda
Gerald Meehl	Marina Nenković-Riznić	R. Pandey
Nisar Memon	Mark New	Chris Paola
Banu Mena Cabrera	Van Giang Nguyen	Maria Nicolina Papa
Arlindo Meque	José Manuel Nicoalu Ibarra	Georgia Papacharalampous
Karina Meredith	Iacopo Nicolosi	Eleftheria Papadimitriou
Katrin Meusburger	Bogdan Niculescu	Juraj Parajka
Patrick Meyers	Jovana Nikolov	Duhee Park
Milosz Mezyk	Rewati Niraula	William Parton
Siyu Miao	Michi Nishioka	J. N. Patel
Mateusz Mikołajczak	Edwin Nissen	Guy Paxman
Rasoul Mirabbasi	Yaxiao Niu	Jana Pazdirkova
Vimal Mishra	Rosane Nobre	Song Peiping
Siddharth Misra	Pawel Nogas	José A. Peláez
Krzysztof Mizerski	Michael Nones	Louise Pellerin

Frederic Pellet	Sergei Pulinets	Adrian Rodriguez-Marek
Zhigang Peng	Jehangir Punthakey	Jerzy Rojek
Antonio Pepe	Hui Qian	Vincenzo Romano
Deborah Pereg	Changyuan Qin	Renata Romanowicz
Loredana Perrone	Tianling Qin	Sandoval-Florez Romulo
Claire Perry	Xiaojun Qiu	Wolfgang Rösler
Nils Perttu	Beatriz Quesada-Montano	Olga Rosowiecka
Laura Peruzza	Yoann Quesnel	Zachary E. Ross
Moczo Peter	Luis Quintanar	Matthias Roth
Andrew Pethick	Shahid Nadeem Qureshi	Hanna Rothkaehl
Carole Petit	Francesco Radaelli	Renata Rotondi
Gregor Petkovšek	Selvakumar Radhakrishnan	Pawel Rowinski
Andrea Petroselli	Prashanth Ragam	Anna Rozwadowska
Eduard Petrovsky	M. S. Rahman	Youyi Ruan
Huong Phan Thien	Omid Rahmati	Lucas Dominguez Ruben
Melodie Philippon	Rajkumar Raikar	Nils Ruther
Camilo Phillips	Carlo Rainieri	Christopher Ruybal
M. Picozzi	Agnieszka Rajwa-Kuligiewicz	Federico Sabina
Adam Pidlisecky	Hamidreza Ramazi	Mauricio Sacchi
R. Pielke	Leonardo Ramirez-Guzman	T. Sagiya
Jorge O. Pierini	Shilpesh Rana	Pratap Narayan Sahay
Jan Pietron	Trygve Randen	Netrananda Sahu
Elżbieta Pilecka	Eugene Rankey	Gaci Said
Zenon Pilecki	Ying Rao	Anup Saikia
Mark Pilkington	Sternfels Raphael	Pallabi Saikia
Santosh Pingale	Grzegorz Raś	Aduwati Sali
Mikolaj Piniewski	Dhananjay Ravat	Ali Sanati
Adam Piotrowski	Jan Reda	Luiz Carlos Sanches
M. A. PiRemli	Julie Regnier	Francisco Sánchez-Sesma
Vladilen Pisarenko	Julie Régnier	Wenjing Sang
Marta Pischiutta	Nora Reichert	Hamid Sanginabadi
Jacek Piskozub	Carl Reine	Langgeng Wahyu Santosa
Thomas Planes	Eduardo Reinoso Angulo	Miguel Santoyo
T. Plenefisch	Rajesh Rekapalli	Giulio Saracino
Boris Podobnik	Liliang Ren	Alexandr Saraev
Ina Pohle	Peizhen Ren	Coskun Sari
Orhan Polat	Abbas Rezaei	Sankar Sarkar
Olga Polechonska	Mohammad Rezaie-Balf	Nicholas Sarlis
Piero Poli	David Rhoades	Majid Sartaj
Silvia Pondrelli	Alessandra Ribodetti	Claudio Satriano
Felice Carlo Ponzio	Michael Riediker	Robert Sausen
Zbigniew Popek	Jaromír Říha	Silvia Scarpetta
Barbara Popielawska	Joakim Riml	Ralf Schaa
Ryan Porter	John Ristau	Michael Scheuerer
Sławomir Porzucek	Vladica Ristić	Erik Schiefer
Hamid Reza Pourghasemi	Coen J. Ritsema	Stanisław Schillak
Lubov Poustovalova	Monica Riva	Simon Schmidt
Biswajeet Pradhan	Fabian Rivera	Ingo Schnauder
Yellapu Siva Prasad	Enzo Rizzo	Roman Schreiber
Janire Prudencio	Gareth Roberts	Kai Schröter
Anna Przebindowska	Scott Robeson	Maike Schumacher
Stafano Pucci	Cordula Robinson	Jeffrey Schweitzer
David Pulido-Velázquez	Steven Roche	Laura Scognamiglio

Emmanuel Scordilis	Maxim Smirnov	Toyoizumi Taro
Christopher Scott	Fernando Soares Schlindwein	Matteo Taroni
Carlo Scotto	I. N. Sokolova	Mohammad Tatar
Uma Seeboonruang	Jerzy Sobotka	R. E. Tatevossian
Gopi Seemala	Elena Yurievna Sokolova	Claudia Tebaldi
Mrinal K Sen	Efthimios Sokos	Jacek Tejchman
Zekai Şen	Urszula Somorowska	Luciano Telesca
I. P. Senanayake	Luis Somoza	Donatella Termini
Ye-Won Seo	Maxime Souvignet	Asirat Teshome
M. Sepehr	Mike Spiliotis	Pankaj Thapa
Naya Servitzoglou	Shalivahan Srivastava	Nikolaos Theodoulidis
Anastasios G. Sextos	Olivier Stab	Wrona Thilo
Umut Sezen	Aaron Stanton	Brian Thomas
Amin Shaban	Ilias Stavrakas	Brian F. Thomas
Mojtaba Shadmani	Colby Steelman	Renie Thomas
Munawar Shah	Mchał Stefaniuk	Nan Tian
Tushaar Shah	Petra Štěpančíková	Larisa Tichtchenko
Kalil Shahbazi	Jonathan P. Stewart	Dieu Tien Bui
Shamsuddin Shahid	Simon Stewart	Timo Tiira
Guang-Zhou Shao	Stavroula Stolaki	Fred Tillman
Lei Shao	Stelios Stoulos	Elisa Tinti
Yaping Shao	F. O. Strasser	Cenk Tokar
Alan Shapiro	Cristine Stumpp	S. K. Tomar
Arabinda Sharma	Zhongbo Su	Michał Tomczak
Babita Sharma	K. P. Sudheer	François Tondeur
Alireza Shateri	Ettammal Suhas	Tomasz Topor
Ranjit Shaw	Jiajia Sun	Joan Miguel Torta
Bin She	Qiang Sun	Ivana Tosic
Yi Shen	Wenyi Sun	Vincenzo Totaro
Aleksey Y. Sheshukov	Nicola Surian	Roberta Tozzi
Peidong Shi	Adi Suryadi	Markos D. Tranos
Ying Shi	Rupert Sutherland	Jorg Trentmann
Zheming Shi	Norikazu Suzuki	Simon Tresch
Hiroshi Shinohara	Susan K. Swanson	Jui-Pin Tsai
Haan Shixing	Arsalan Syed	Alexandros Tsavdaris
David R. Shklyar	Norbert Szabo	Ioannis M. Tsodoulos
Saeed Shojaei	Dawid Szafranski	Kagan Tuncay
D. L. Shrestha	Artur Szkop	Endre Turai
Da Shuai	Jan Szolgay	Azra N. Tutuncu
James Shucksmith	Malgorzata Szwed	Serhat Tüzün
Marcin Siepak	Adam Szymkiewicz	Tomasz Tyminski
Fernando J. S. Silva Dias	T. J. P. M. Op t Root	Evangelos Tyrllis
Hengki Simanjuntak	Hossein Tabari	Andreas Tzanis
Slobodan Simonovic	Ali Rashid Tabrez	Trushit Upadhyay
Chandrani Singh	Isomura Takuya	Andrzej Urbaniec
Grzegorz Sinicyn	Ali Talebi	A. Uromeihy
Michael Sioutas	Morteza Talebian	Ruben Urraca
Andreas Skarlatoudis	Elcin Tan	Zbigniew Ustrnul
Agnieszka Skowron	Mou Leong Tan	Babak Vaheddoost
Louise Slater	Hiroyuki K. M. Tanaka	Rein Vaikmae
Michael Slawinski	Jizhou Tang	Katri Vaittinen
Evert Slob	Gulum Tanircan	Carlos Valdés González
Łukasz Słonka	Dan Tao	Jan Valenta

Raúl Valenzuela Wong	Lukasz Wiejaczka	Heesung Yoon
Le Van Anh Cuong	Simon D. P. Williams	Jiachun You
Henny A. J. Van Lanen	Peter Wintoft	Jinsheng You
Anne Van Loon	Piotr Witek	Qingyu You
R. D. Van Remortel	Axel Wittmann	Isma Younes
Cees Van Westen	Edwin Wnuk	E. G. Youngs
Andres Vargas-Luna	Kurt Wogau	Guo Yu
Emmanouil Varouchakis	Marek Wojdyla	Peng Yu
Arianna Varrani	Nam C. Woo	Yongcai Yu
Lampros Vasiliades	Andrew W. Wood	Sanyi Yuan
Ivana Vasiljevic	Anders Worman	Zainuddin M. Yuso
Massimiliano Vassallo	Daniel Wright	Hasan Zabihi
Saeed Vatankhah	Bangyu Wu	Hamidreza Zahabi
Arie Vatesia	Jianchao Wu	I. E. Zakharenkova
Sreeja Veettil	Rushan Wu	Ahmad Zamani
Ramesh Vellore	Wei Wu	Eddy Y. Zeng
Kalyana C. Veluvolu	Wensheng Wu	Hongyu Zhai
Sumit Verma	Yao Wu	Bo Zhang
D. J. Verschuur	Yih-Min Wu	Chao Zhang
Foteini Vervelidou	Zhonghai Wu	Chong Zhang
Davide Vettori	Zhongliang Wu	Hao Zhang
W. W. S. Vieira	Jobst Wurl	Rui Zhang
Daniele Viero	Gilead Wurman	Ruili Zhang
Olga Vigiak	C. Wylie Poag	Shaohua Zhang
Giulio Vignoli	Max Wyss	Wei Zhang
Francesco Visini	Andrzej Wyszogrodzki	Xiaoping Zhang
Elena Volpi	Bartłomiej Wyzga	Xiong Zhang
James Voogt	Stelios Xanthos	Yi-chun Zhang
Sergiy Vorogushyn	Liang Xiao	Yijie Zhang
M. T. Vu	Maohua Xiao	Zhen Zhang
Richard Walker	Tao Xu	Guangju Zhao
Clemens Walther	Youpeng Xu	Peiqiang Zhao
Yunxia Wan	Yungui Xu	Yan Zhao
Niko Wanders	Shibei Xue	Yang Zhao
Di Wang	R. N. Yadav	Xiaowei Zhao
Hao Wang	Sanjay Yadav	Yingcai Zheng
Hua Wang	Soheil Yahyapour	Xiefei Zhi
Hui Wang	Farzam Yaminifard	Jian Zhou
Kun Wang	Denghua Yan	Shiyong Zhou
Meng Wang	Jidong Yang	M. Zhu
Shifeng Wang	Qingjie Yang	Peimin Zhu
Tieyi Wang	Qinke Yang	Rixiang Zhu
Xiangchun Wang	Gang Yao	Senlin Zhu
Yuefeng Wang	Morgan Yarker	Marcin Zielinski
Arnaud Watlet	Zaher Yaseen	George Zittis
Kamila Wawrzyniak-Guz	Nergis Yasmin	Bogdan Zogala
Shimon Wdowinski	Eric Yee	Matthew Zook
Fabian Welc	Yi-Lung Yeh	Tomasz Zorski
Kuo-Liang Wen	Emrah Yenier	Joseph Zume
Yixin Wen	C. Yonezawa	Ramon Zuñiga
Friedemann Wenzel		

We apologize for any errors or inadvertent omissions.

Acta Geophysica

Volume 68
Number 1
2020

RESEARCH ARTICLES - SOLID EARTH SCIENCES

Seismic time–frequency spectrum analysis based on local polynomial Fourier transform
P. Qi · Y. Wang 1

Data integration for earthquake disaster using real-world data
C. Tian · G. Li 19

Full waveform inversion based on a local traveltimes correction and zero-mean cross-correlation-based misfit function
S. Dong · L. Han · Y. Hu · Y. Yin 29

Possibility of an earthquake prediction based on monitoring crustal deformation anomalies and thermal anomalies at the epicenter of earthquakes with oblique thrust faulting
Z. Alizadeh Zakaria · F. Farnood Ahmadi 51

Peak ground acceleration prediction by fuzzy logic modeling for Iranian plateau
B. Karimi Ghalehjough · R. Mahinroosta 75

RESEARCH ARTICLES - APPLIED GEOPHYSICS

Magnetotelluric sounding evidence of development of nappes in the Tuolai Sag, Yin-E Basin
Z. Penghui · F. Hui · Z. Xiaobo · Z. Yaoyang · X. Haihong · P. Yan · Y. Yongzhen 91

Assessing the effects of 1D assumption violation in vertical electrical sounding (VES) data processing and interpretation
S.R. Mashhadi · H. Ramazi · M. Asgarpour Shahreza 105

Novel wide-angle AVO attributes using rational function
Y. Xue · J. Xiang · X. Xu 123

RESEARCH ARTICLES - HYDROLOGY

The added value of spatially distributed meteorological data for simulating hydrological processes in a small Mediterranean catchment
A. Gara · K. Gader · S. Khelifi · M. Vandooster · D. Jendoubi · C. Bouvier 133

Estimation of groundwater potential using GIS modeling in Kohistan region Jamshoro district, Southern Indus basin, Sindh, Pakistan (a case study)
A. Memon · K. Ansari · A.G. Soomro · M.A. Jamali · B. Naeem · A. Ashraf 155

Novel hybrid approaches based on evolutionary strategy for streamflow forecasting in the Chellif River, Algeria
M. Zakhrouf · H. Bouchelkia · M. Stamboul · S. Kim 167

A new mathematical model to calculate the equilibrium scour depth around a pier
A.H. Gazi · M.S. Afzal 181

(Contents continued on inside back cover)