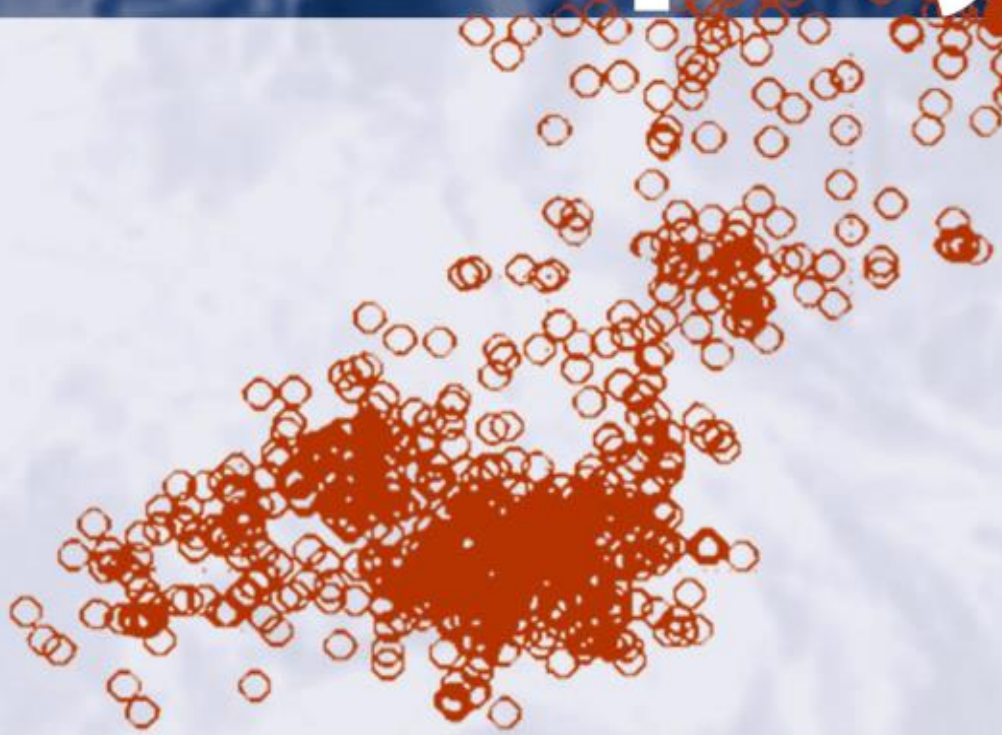
Volume 67 • Number 5 • October 2019

Acta Geophysica



PAN
POLISH ACADEMY OF SCIENCES

Institute of Geophysics
Polish Academy of Sciences

Springer

Acta Geophysica (2019) 67:1277–1295
https://doi.org/10.1007/s11600-019-00345-8

RESEARCH ARTICLE - SOLID EARTH SCIENCES

Interpretation of gravity anomaly to delineate thrust faults locations at the northeastern part of India and its adjacent areas using source edge detection technique, tilt derivative and Cos(θ) analysis

Gopal K. Ghosh[1]

Received: 2 May 2019 / Revised: 13 August 2019 / Accepted: 19 August 2019 / Published online: 28 August 2019

Abstract
The Northeast India and its adjacent areas converge among the three different plates, viz. Eurasia, India and Sunda plates. The tectonic interaction of Northeast India and underlying dynamics of the Himalayas as well as the Indo-Burma Ranges might cause the Assam Syntaxis. The area of study is located between latitude 23°–28°N and longitude 88°–96°E and situated in one of the most seismically active tectonic provinces in the world with seismic zone-V. This area had demonstrated several thrust faults activities and tectonic evident accomplishments during the recent past. The complicated geotectonic setups inspirits various smaller magnitude earthquakes, and the current seismicity shows seismic activities are still enduring in the Shillong Plateau, Arakan-Yoma fold belt, Bengal Basin, Naga Hills, Mikir Hills, Upper–Lower Brahmaputra Valley and Mismi Hills of Himalayan foothills. It is imperative to obtain wide-ranging learning tectonic configuration, thrust faults delineation for improved geoscientific study. Parts of the areas are extremely unreachable, and very limited thrust faults were marked by studying GIS map received from the various agencies and field geological study. During the past studies, most of the prominent lineaments/thrusts are marked; however, many active and hidden thrust faults are still unidentified. Seismic data can provide better information about the thrust faults locations, but due to small number of seismic data, the information is not adequate. In this paper, attempt has been made to study and reinterpret the available ground gravity data of northeastern parts of India for understanding thrust fault locations using various applications of gravity derivatives like analytical signal, horizontal gravity gradient, tilt derivative, horizontal tilt angle derivative and Cos(θ) analysis. Source edge detection technique has also been premeditated to categorize thrust fault locations. It is understandable that the low gravity is observed at Assam Valley which contributed sediment accumulations and higher gravity anomaly observed at Shillong Plateau and Bengal Basin containing denser formations. Bouguer gravity data is used after isostatic correction assuming Airy's isostasy root depth model and first-order trend removal using least square technique. The derived thrust fault locations from the present study are superimposed with the existing thrust-fault locations for correlation. Some additional thrust faults are narrated which are not previously mapped. It is also suggested that Brahmaputra Thrust, Dauki Fault, Naga Thrust, Disang Thrust and Kopili Fault have key responsibility for high seismicity and tectonic movement causing upliftment and depression that encouraged some anticlockwise rotation in the area.

Keywords Shillong plateau · Source edge detection (SED) · Total horizontal derivative (THDR) · Tilt derivative (TILT) · Horizontal tilt angle (TDX) · Cos(θ)

Introduction

The Northeast India and its surrounding areas converge among the three different plates, viz. Eurasia, India and Sunda plates (Fig. 1). Convergence of two continental plates causes subduction and collision. Such collision in the underthrusting lithosphere creates sever deformation on the upper part of the crust triggering a seismically active tectonic provinces. During the continental drifting process at different

✉ Gopal K. Ghosh
gk_ghosh@yahoo.com

[1] Oil India Limited, Duliajan, Assam, India

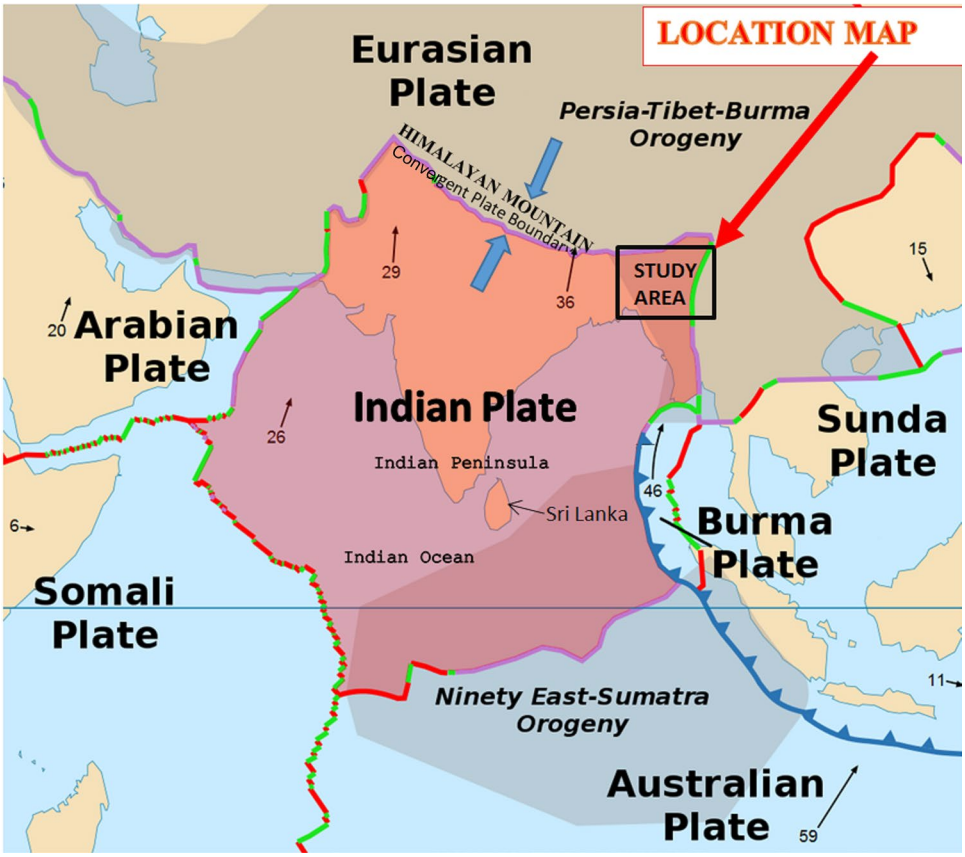


Fig. 1 Map shows various tectonic plates like Indian Plate, Eurasian Plate Sunda Plate, Australian Plate, Arabian Plate and Somali Plate surrounding the study area (after Wikipedia 2019)

span of time, 70 million years ago, 50 million years ago and India as on today, different positions of India have been shown in Fig. 2 (USGS Report 2015).

The tectonic interaction of Northeast India and underlying dynamics of the Himalayas as well as the Indo-Burma Ranges might cause the Assam Syntaxis. The study was carried out in the area between latitude 23°–28°N and longitude 88°–96°E (Fig. 3) which falls in the northeastern part of India and its adjacent areas. Study reveals that North–South compression of Eurasian Plate and Indian Plate (Fig. 2) creates plate convergence and seismotectonic activity in this area (USGS Report 2015). Angelier and Barua (2009) studied the focal mechanism of earthquakes of Indo-Burma Ranges and their kinematic implication in the Northeast. Tapponnier et al. (1982) stated that the spreading of Andaman Sea and the Sagaing Transform Fault motion in the Burma might be the reason for the formation of Assam Syntaxis. This northeast India has fallen in the seismic zone-V as the highest seismic zone (Fig. 4) (Bansal and Verma 2013).

Northeast India comprises various thrust faults which are oriented in different directions. The elevation in this area varies in the range of about 6 km (Fig. 3). Most elevated areas are undulating areas such as Shillong Plateau, northern part of the Himalayas, Indo-Burma Ranges and Mikir Hills. The depressions are distinguished at Bengal Basin, Assam Valley, Brahmaputra Valley and Molasse Basin. This depression can be better understood as a gravity difference and can be explained by mass deficiency. The aim of the present work is to understand the old gravity data and to interpret those using different interpretational approaches.

The Bouguer gravity map is prepared by National Geophysical Research Institute (NGRI) with the available data source from Survey of India (SOI) (Gulatee 1956), Burma Oil Company (BOC) (Evans and Crompton 1946), Oil and Natural Gas Commission (ONGC). NGRI has carried out additional gravity data observations in the Northeast India to fill up the data gap during the year 1972–1973 and prepared the map. Later on, these original gravity data were published by Verma and Mukhopadhay (1977). Ghosh et al. (2015) digitized the published gravity data from Verma and Mukhopadhay (1977) and re-gridded for further analysis (Fig. 6).

Gravity data (Verma and Mukhopadhay 1977) reveal that southern and southeastern parts have much data gap comparable to the other parts. The Bouguer gravity data has zonation in the northern part of the zone parallel to the Brahmaputra Thrust; discordant, NE–SW elongated gravity anomaly was observed in the NE part which is parallel to the Disang and Naga thrust. The N–S elongated gravity anomaly was observed in the southern part, and it is associated with Molasse Basin. The entire northeast India is characterized

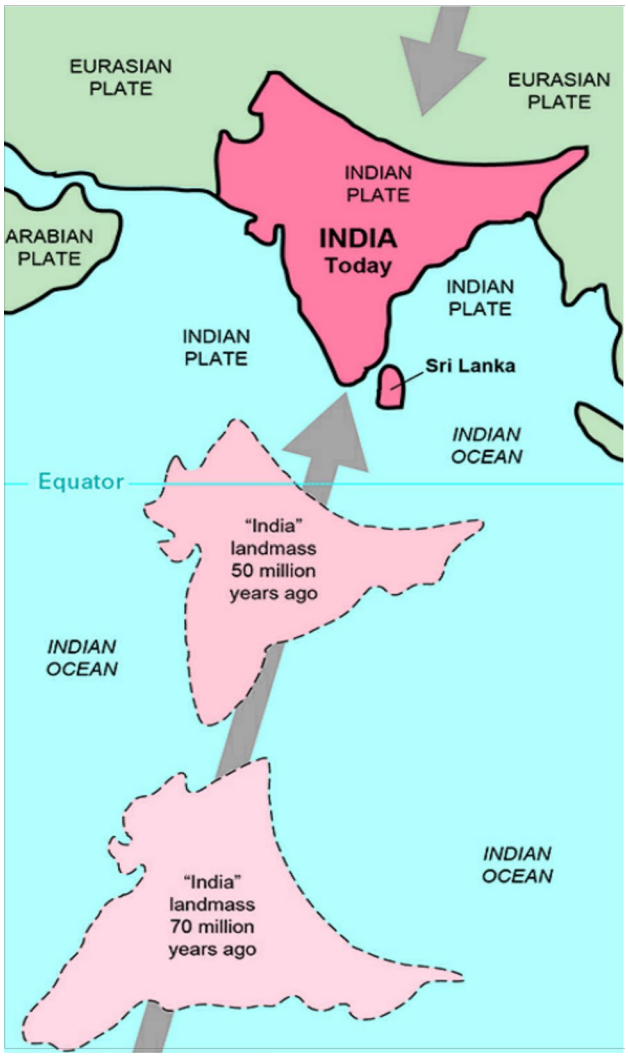


Fig. 2 Map shows different positions of Indian Plate (India landmass) before its collision with Eurasian Plate during the different span of time: 70 million years ago, 50 million years ago and India as on today. Solid lines indicate the present-day continents and dashed outlines the "India" landmass (visual reference only). Northern margins began to collide against the southward-moving Eurasian Plate about 40–50 million years ago when the "India" landmass was once situated well south of the Equator (after USGS report 2015)

by huge changes in Bouguer gravity value varying from +45 mGal at Shillong Plateau to −255 mGal at Assam Valley with a difference of 300 mGal. The higher gravity is observed at Bengal Basin (positive values), and the lower gravity values (negative values) are observed at north of Brahmaputra Thrust, Assam Valley and Indo-Burma Ranges and Molasse Basin. Northern part of the area suggests low gravity due to isostatic adjustment between Brahmaputra River and foot hills of the Himalayas.

To understand the mechanism of the tectonic activities of the area, various applications including analysis of analytical signal, total horizontal derivative (THDR), tilt derivative (TILT), horizontal tilt angle (TDX), Cos(θ) map, have been generated from gravity data. Various thrust-faults are identified with the help of different derivatives of gravity data. The know thrust faults like Dauki Fault, Brahmaputra Thrust, Kopili Fault, Naga Thrust, Kabaw Fault, Jorhat Fault, Oldham Fault and Main Central Thrust are superimposed on the new derived results and correlated. It is noted that the thrust faults derived previously are well correlated with the newly derived thrust faults by gravity delineation. Source edge detection (SED) technique is used based on source parameter imaging technique to demarcate the thrust faults and its dip directions. The above study suggests that the diversely directed forces are acting in this region causing upliftment and depression of subsurface, and this leads also to an anticlockwise rotation of Shillong Plateau and Mikir Hills (Singh et al. 2017).

It is also suggested that Brahmaputra Thrust, Dauki Fault, Naga Thrust, Disang Thrust and Kopili Fault are playing most important role in tectonic association. So, it is more substantial to identify the particulars of delineation of thrusts and faults structures which directly or indirectly impact seismic activity in the Northeast India and its adjacent areas. It is also studied that Shillong Plateau exhibits denser basement; however, Assam Valley demonstrates thicker sedimentary cover. It can be explained that horizontal density differences are responsible for the variation of gravity field.

Geological and tectonic setting

The Himalayan collision belt has formed due to the ongoing thrusting of Indian Plate toward the Eurasia Plate. Northeast India and its adjacent areas basically cover the northeastern Himalayas toward north, Indo-Burma Ranges toward east and Bengal Basin toward south. It is prudent that between these two plates' boundaries, parts of the Indian Plate including Assam Valley and Shillong–Mikir Plateau formed Assam Syntaxis. The eastern Himalayan collision belt is acting toward the east, whereas Burmese Plate is acting toward southeast direction. Southeastern parts of the Northeast India extended to Andaman–Sumatra region. These plates cause an intraplate distortion which is fairly complex. The northeast India and its surroundings exhibit a complex multifaceted tectonics setup producing high seismic activity zone-V (Fig. 4) (Bansal and Verma 2013).

The Shillong Plateau is still active due to counterclockwise movement of the Indian Plate against the Eurasian Plate (Harijan et al. 2003). The Himalayan upliftment activity might also be explained due to the tectonic resettlement causing gravity gliding to their shape and structures. The highest altitude of Shillong Plateau is more than 2000 m, and an average altitude in this area between Brahmaputra Thrust (BT) and the Dauki Fault (DF) is 1000 m above mean sea

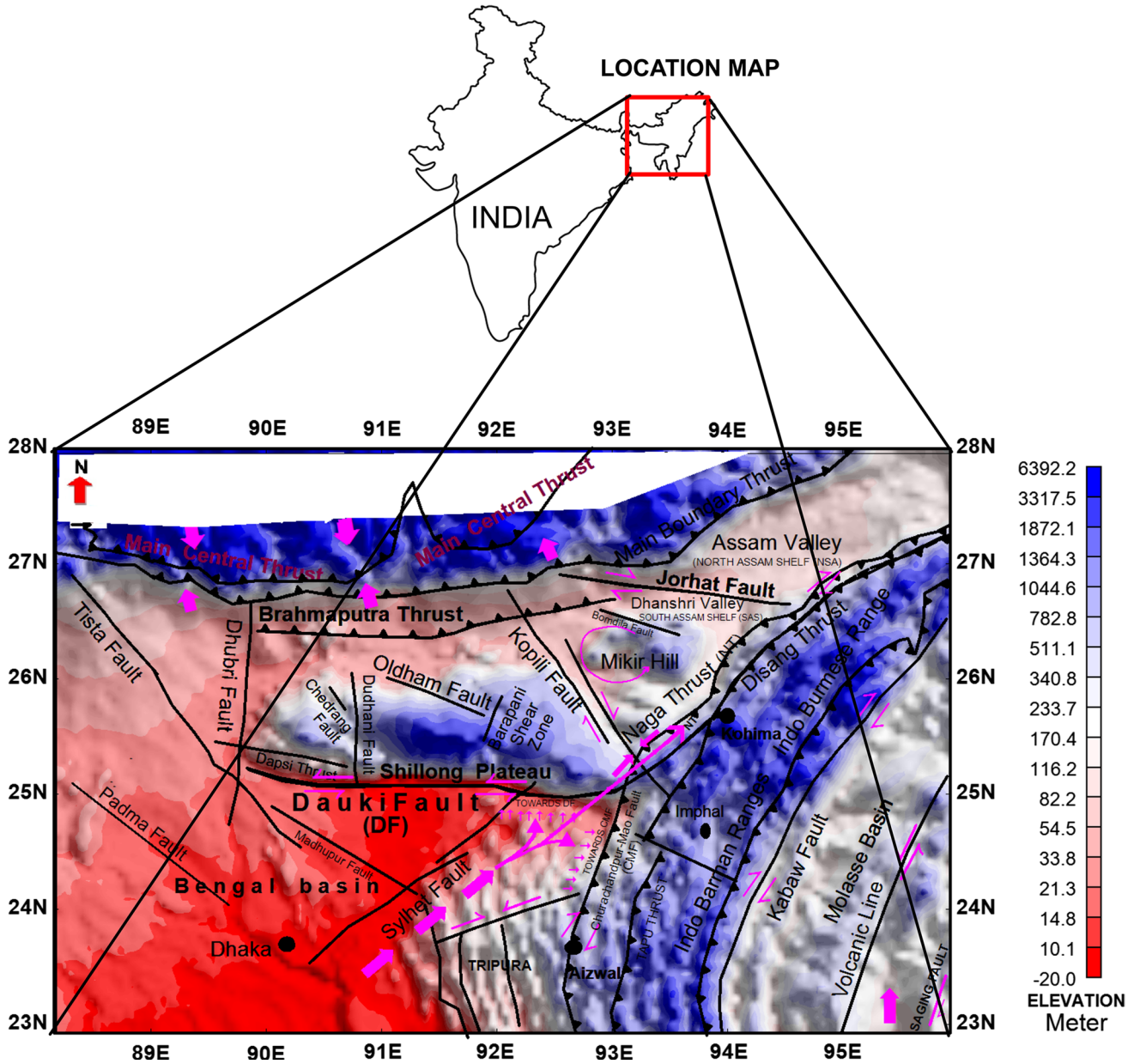


Fig. 3 Map shows the elevation of the study area. High elevations are noted at MBT, MCT, Shillong Plateau, Indo-Burma Ranges and Mikir Hills. Low elevation areas are at Bengal Basin and Assam Valley. The known (outcropping) thrust faults are marked by black lines. The major thrust faults are Main Boundary Thrust (MBT), Main Central Thrust (MCT), Brahmaputra Thrust (BT), Tista Fault (TF), Dudhani Fault (DhF), Dauki Fault (DF), Chedrang Fault (CF), Oldham Fault (OF), Kopili Fault (KF), Sylhet Fault (SF), Naga Thrust (NT), Disang Thrust (DT), Jorhat Fault (JF) and Kabaw Fault (KF)

level. Various thrust faults like Dapsi Thrust, Jorhat Fault, Kabaw Fault, Naga Thrust, Disang Thrust, Sylhet Fault, Tista Fault, Kopili Fault, Oldham Fault, Dhubri Fault are also playing important role in tectonic association and movement (Fig. 3). The past earthquakes (magnitude > 8.0 in Richter scale; 1950 and 1897) in the Northeast India had large impact in the area (Rao et al. 2006). The historical record shows that the great earthquake 8.7 magnitudes occurred along the north-dipping Dauki Fault (Oldham 1899). Based on the micro-earthquake data Kayal and De (1991) remarked about the location of seismicity at the Dapsi Thrust (Du T) as the supporting thrust of Dauki Fault. Moreover, the earthquakes are a consequence of a popup tectonic structure (Nayak et al. 2008) of the south-dipping Oldham Fault and north-dipping Dauki Fault as stated by Bilham and England (2001). This popup structures might be responsible for the anticlinal uplift with the doubly arranging fold system (McClay and Bonora 2001; Schellart and Nieuwland 2003).

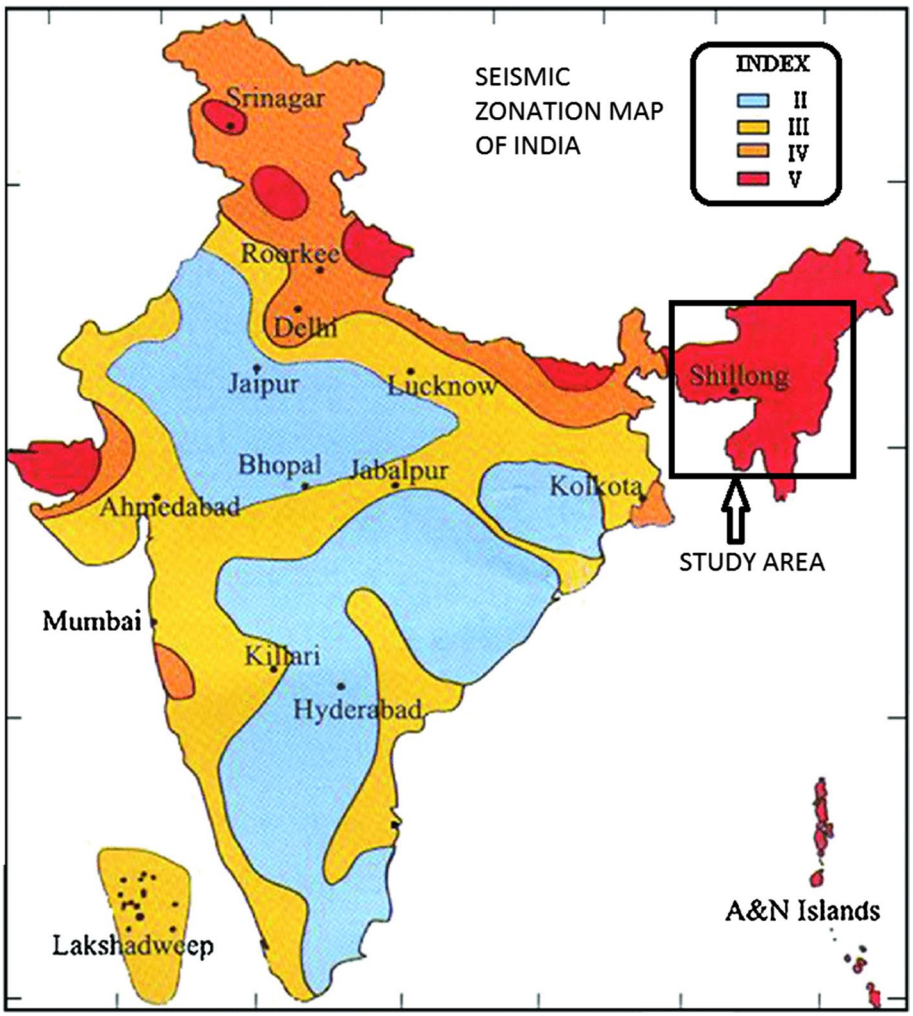


Fig. 4 Seismic zonation map of India (BIS-2002). The study area falls in the northeast Indian zone with highest seismic zone-V (after Bansal and Verma 2013)

Chandra (1984) stated about the existing of Shillong Plateau signifying a convergence arc in between Himalayan arc and the Burmese arc. Seno and Rehman (2011) have suggested that the tectonic disturbances in this area are caused by the Indian Plate and Eurasian Plate collision at the Himalayan zone caused by subduction beneath the Burmese Plate. It has been stated that the various thrust faults within the Shillong Plateau are oriented in the different directions due to the tectonic disturbances caused by the Indo-Burma subduction zone and the Himalayan collision zone (Rajesekhar and Mishra 2008; Kayal 2001). The Shillong massive plateau comprises Archean gneissic complex (Archean rock), metasedimentary Shillong Group rocks (1530–1550 Ma), igneous rocks, porphyritic granites and ultramafic alkali–carbonate complexes (Evans 1964; Mitra 1998; Mishra and Sen 2001; Devi and Sarma 2010). Cretaceous–Tertiary sediments are present in the southern part of the Shillong Plateau (Biswas and Grasemann 2005). Further to the south of the Dauki Fault, the thickness of sedimentary cover is larger in Sylhet varying from 13 to 18 km dating from the Tertiary to recent time (Biswas and Grasemann 2005; Evans 1964; Alam et al. 2003; Ghosh et al. 2015). In Assam Valley, various types of rocks are present at different ages like Upper Precambrian–Lower Proterozoic, Gondwana and tertiaries which constituted major portion of the sedimentary rocks. The thicknesses of sedimentary formation in this region is more than 6 km and decreases continuously toward east. Oldham Fault 110 km long dips at angle approximately 57° toward S–SW direction and submerges 9–45 km beneath the Shillong Plateau. The Oldham Fault might be "Cryptic" reverse fault extended to the northern boundary of the plateau (Bilham and England 2001). The Brahmaputra Thrust (BT) is situated in the northern end of the Oldham Fault (Rajendran et al. 2004; Kayal et al. 2006). It is noted that the Mikir Massif (Mikir Hill) was supposedly a fragmented portion of the Shillong Plateau caused by the major Kopili Fault and experienced large earthquake in 1869 (Mw = 7.38, Oldham 1883; Nandy 2001; Ambraseys and Douglas 2004) and 1943 (Mw = 7.42, Nandy and Dasgupta 1991). The simplified geological map of the study area covering northeast India, Bangladesh, China and Myanmar is shown in Fig. 5 (after Robinson et al. 2014; Awasthi et al. 2014).

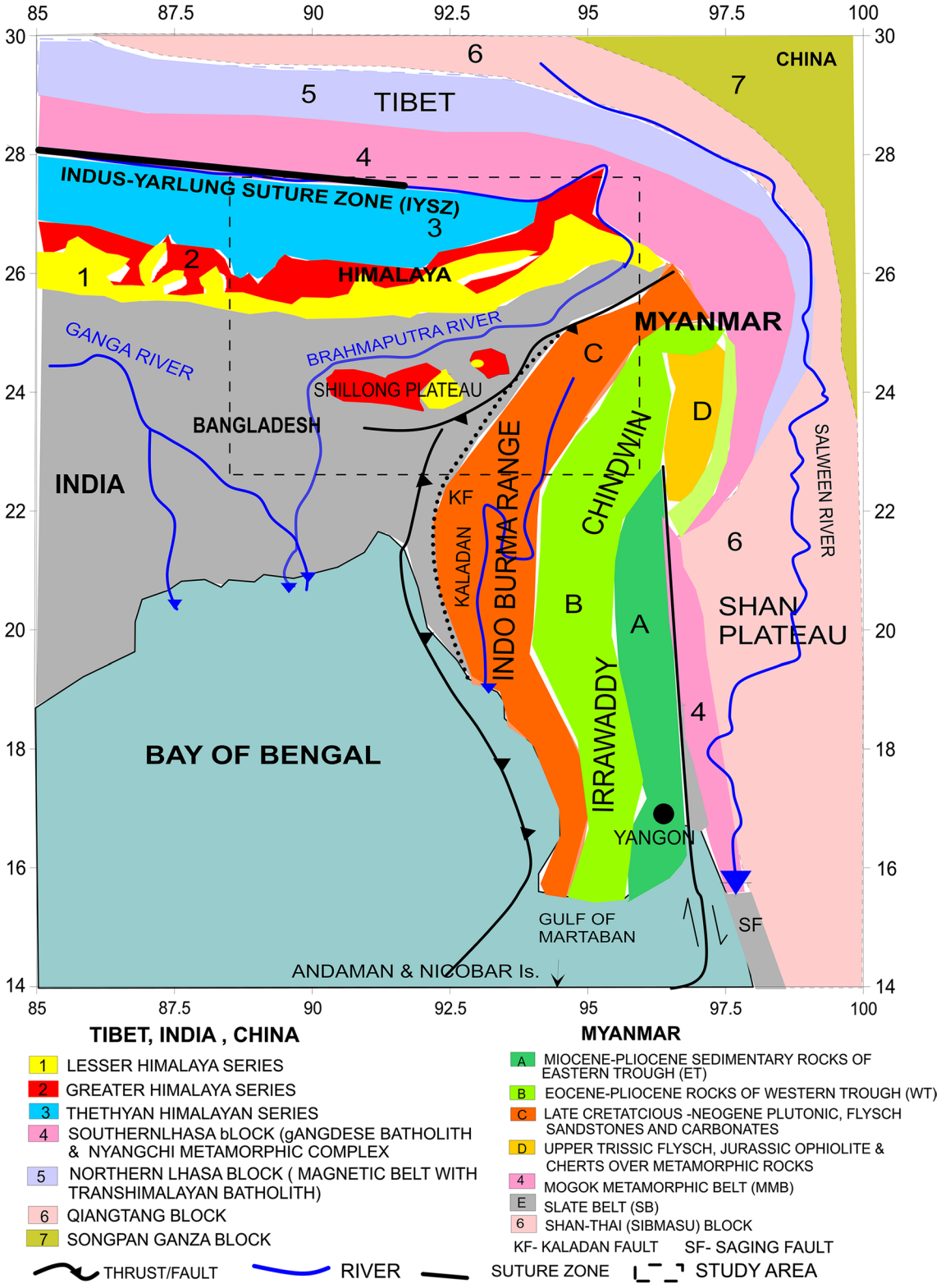


Fig. 5 Simplified geological map of the Northeast India and its surroundings indicating various locations of boundaries, thrust faults, rivers, suture zone, etc. (figure modified after Robinson et al. 2014; Awasthi et al. 2014)

Gravity anomaly

Original Bouguer gravity data by Verma and Mukhopadhay (1977) have been digitized and re-gridded by Ghosh et al. (2015) (Fig. 6) to carry out the interpretation work. Average elevation is changes of more than 1000 m above from the Shillong Plateau covering Brahmaputra Thrust (BT) in the north to 500 m in the Dauki Fault (DF) in the south. However, maximum gravity value of +45 mGal is noted at Shillong Plateau where elevation is 1700 m. The lowest gravity value of −255 mGal is observed at Assam Valley. The northern part of the study area near Main Boundary Thrust (MBT) and Main Central Thrust (MCT) has as low gravity field with higher elevation. This Bouguer gravity field has lineation oriented in W–E direction parallel to the Brahmaputra Thrust. The low gravity field is also observed at the eastern part of the investigated area, whereas orientation of isolines turns from S to ENE direction. This low gravity is observed at Molasse Basin, Kabaw Fault and at the Indo-Burma Ranges. The extreme southeastern part of the area that is toward the east of the Molasse Basin again shows comparatively higher gravity values. The moderate high gravity value is noted at the Bengal Basin (Fig. 6). The sedimentary thickness is varying by more than 6 km increasing from westward to eastward direction in the northeastern part of India (Verma and Mukhopadhay 1977). These strong positive high gravity anomaly values at Shillong Plateau suggest that there could be shallower basement rock or higher-density sedimentary rocks can underlay the shallower ones there.

It can also be predicted that the thickness of sedimentary rocks will be lesser below the Bengal Basin. The upper mantle rock might have the relatively higher density. So, the higher gravity anomaly might be the cause of excess mass which needs to be corrected. Based on the Bouguer anomaly, Qureshy (1971) and Mukhopadhyay (1974) generate a conclusion with thicker denser crust underlain the plateau. However, Woollard (1962) suggested that below 250 km depth, isostatic compensation need not be required because isostasy characterizes lithosphere, and below this depth the equilibrium is closer to hydrostatic one. A comparative study of the crustal depth at the Shillong Plateau, Upper Assam Valley and the Bengal Basin are presented by Worzel and Shurbet (1955) using models for Shillong Plateau, for Upper Assam Valley and for Bengal Basin. The authors suggested different density variations in Shillong Plateau, Assam Valley and Bengal Basin. It is summarized that the thick sedimentary

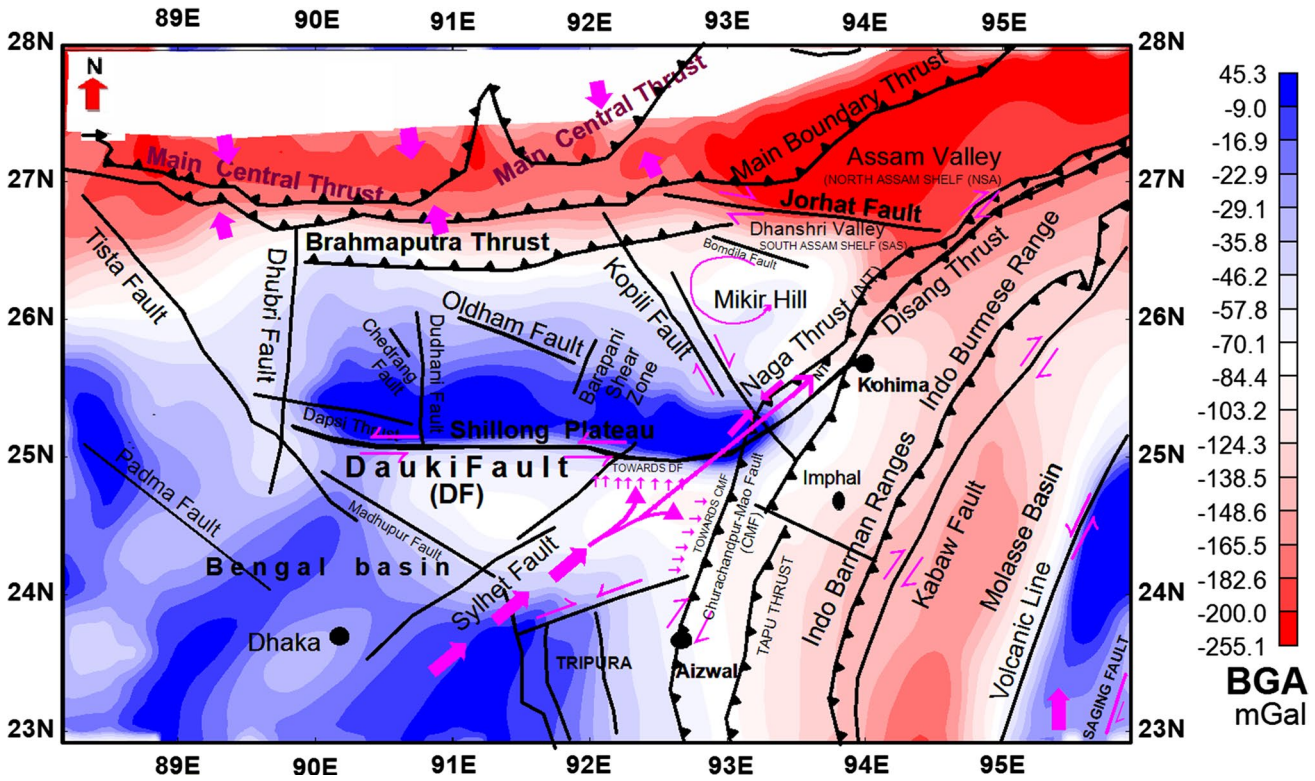


Fig. 6 Bouguer gravity anomaly map of the study area (after Verma and Mukhopadhay 1977). The gravity anomaly varies from −255 to 45 mGal. Low gravity value observed at the northern Himalayas, Main Boundary Thrust (MBT) and Main Central Thrust (MCT) zones, Assam Valley and Molasse Basin

cover of the crust at Bengal Basin is 33 km, i.e., 13-km-thick sedimentary cover and 20-km-thick crystalline basement. For Upper Assam Valley, the crustal thickness is 33.75 km, i.e., 4.1-km sedimentary formation and 29.65-km crystalline basement. In the case of Shillong Plateau, sedimentary formation is 1 km thick and the crystalline basement would have 46.3 km thickness. Tandon (1954) commented on the last result that crustal thickness 46.3 km and average density 2960×10^3 kg/m^3 modeled based on seismological study which is rather nor typical for standard normal crust.

It is observed that the gravity variation is visibly influenced by isostatic equilibrium. The mass deficiency is suggested at northern part of Brahmaputra Thrust and Assam Valley and similarly extra masses are suggested at Shillong Plateau. The isostatic regional correction is applied to Bouguer gravity anomaly in this area using Airy's isostasy model where average 39 km crust thickness is considered after studying various research opinions although suggested depth to Moho values are varying from 32 to 46 km (Kumar et al. 2004; Nayek et al. 2008; Bora and Baruah 2012; Rajesekhar and Mishra 2008; Mitra et al. 2005; Borah et al. 2016). The general observations are that the Bouguer anomaly trend at Molasse Basin is oriented in NE–E to SW–W direction parallel to Indo-Burma Ranges. The gravity value is moderate, whereas north of Brahmaputra Thrust has low gravity with low-density sediments, and Bengal Basin contains higher gravity with higher-density sediments. It is observed that at Shillong Plateau and Mikir Hills, considerable basements rocks (Lower Precambrian) are exposed.

It is stated that elevation variation along with gravity variation is very high, and hence, gravity data need to be corrected for expected isostatic compensation to topographic loads, to moderate mid-wavelength the influence of the lower-lithospheric sources. The elevation map (topo grid, Fig. 3) has been used to calculate root depth (m) (Fig. 7) necessary to balance the topographic load (according to isostasy principle) for each topo grid point (Fig. 3). Isostatic regional gravity anomaly (Fig. 8) has been calculated using Bouguer density of 2670×10^3 kg/m^3 (Evans and Crompton 1946), and Moho density contrast is 335×10^3 kg/m^3 assuming Airy's isostasy model calculated root depth (Fig. 7) at each grid point. Isostatic residual field (Fig. 9) is calculated by subtracting isostatic regional gravity anomaly grid (Fig. 8) from the Bouguer gravity anomaly (Fig. 6). It is noted that isostatic residual gravity anomaly has a huge trend (− 176 to 123 mGal), and it is corrected further using first-order correction with the help of least square technique. After the removal of the trend-corrected isostatic residual gravity anomaly varies from − 132 to 139 mGal in the area shown in Fig. 10. Trend-corrected isostatic residual gravity anomaly further is applied to calculate the first vertical gravity gradient and the other two orthogonal horizontal gradient

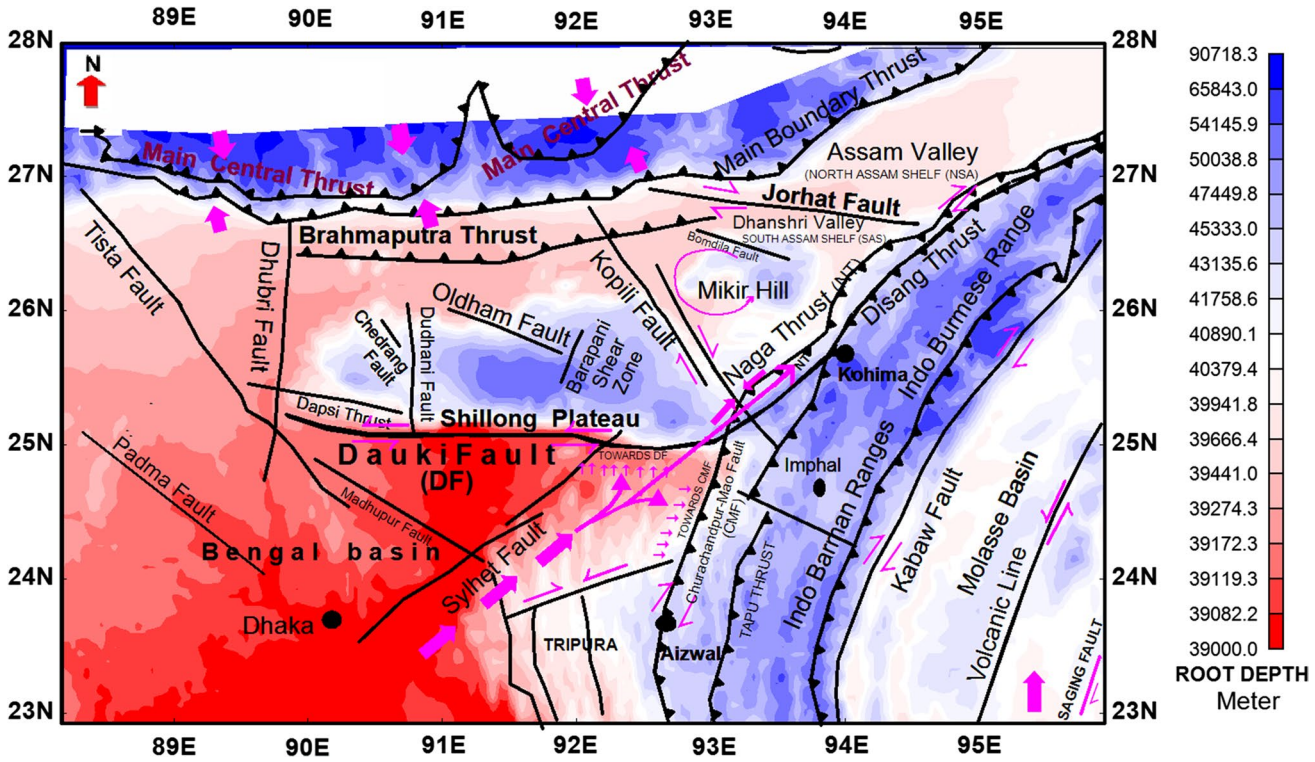


Fig. 7 Depth of balanced isostatic columns calculated assuming Airy's isostasy model at each grid point with the use of topo map. The assumed crustal density is 2670×10^3 kg/m^3, and Moho density contrast is 335×10^3 kg/m^3

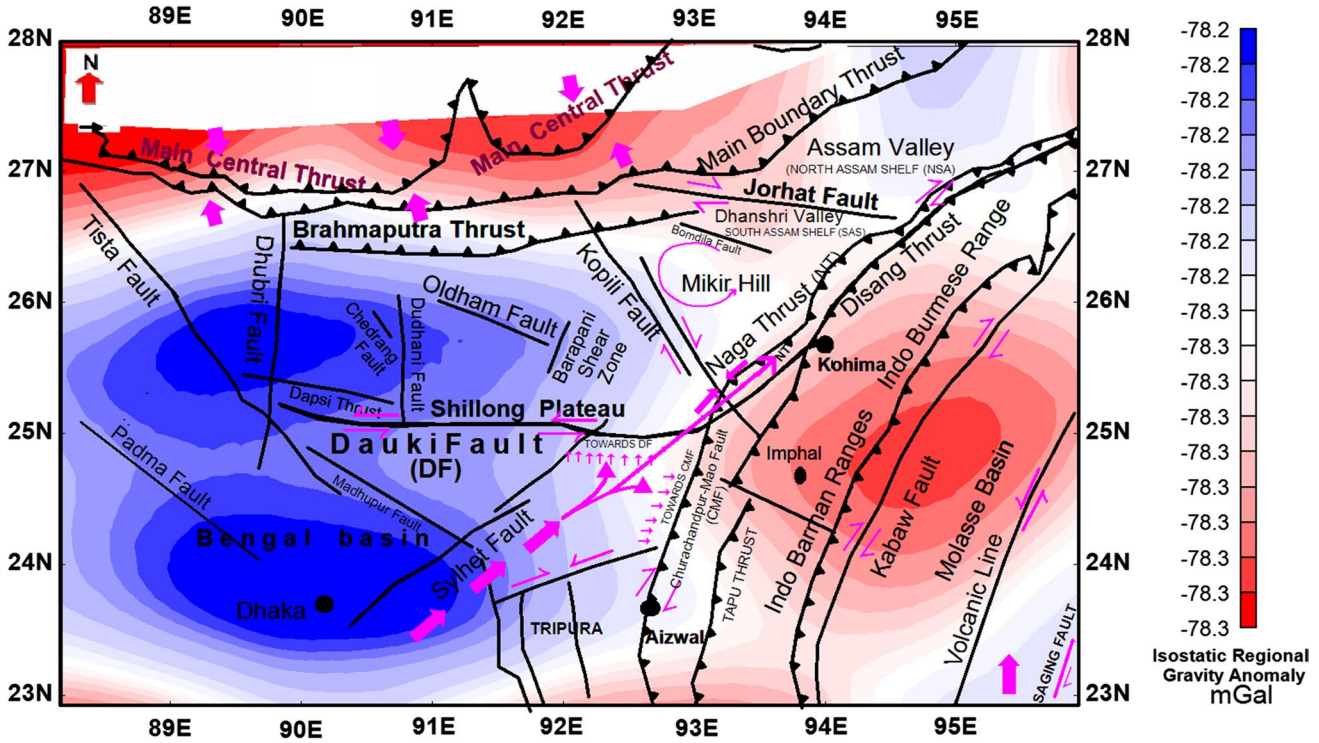


Fig. 8 Isostatic regional gravity anomaly response (mGal) calculated from Airy's root depth and using topography grid map

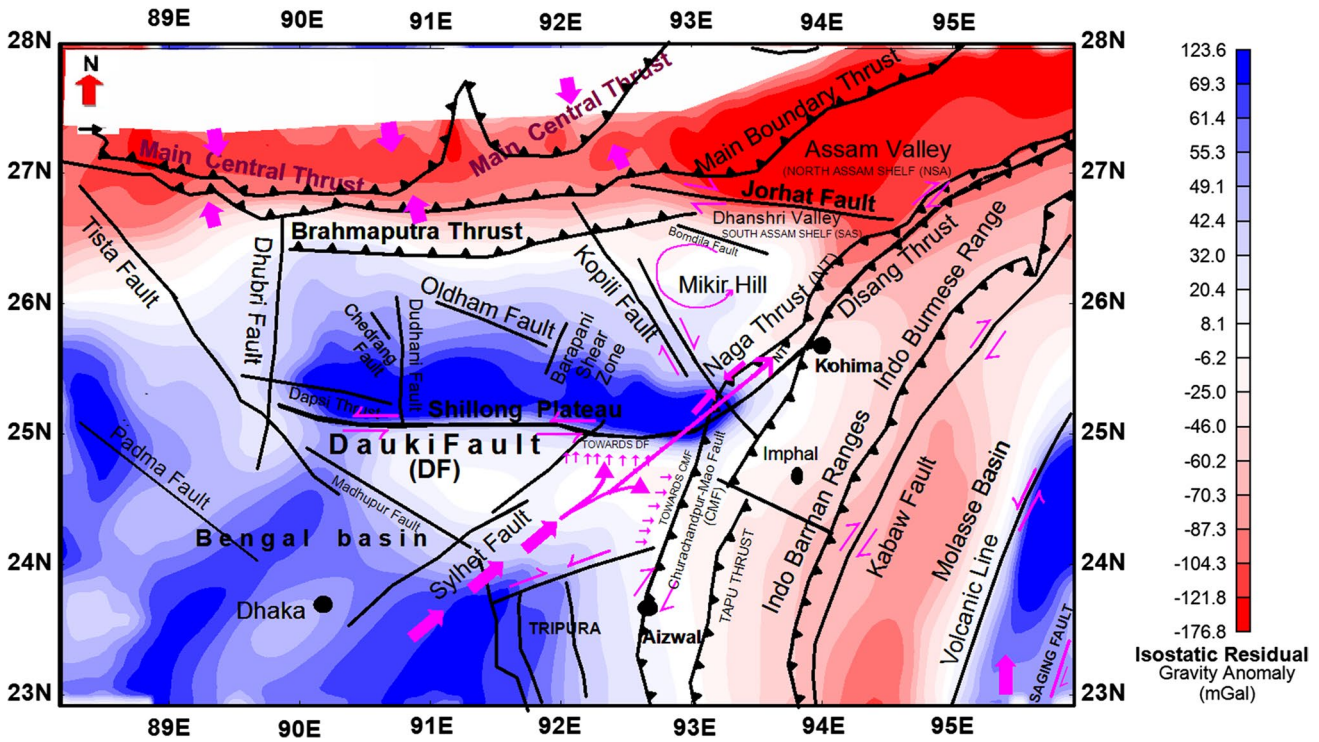


Fig. 9 Isostatic residual gravity anomaly derived after subtraction of isostatic regional gravity anomaly from the Bouguer gravity anomaly. Isostatic residual gravity anomaly suggests central high gravity value situated at Shillong Plateau. Gravity lows are located at MBT, MCT, Assam Valley and Molasse Basin

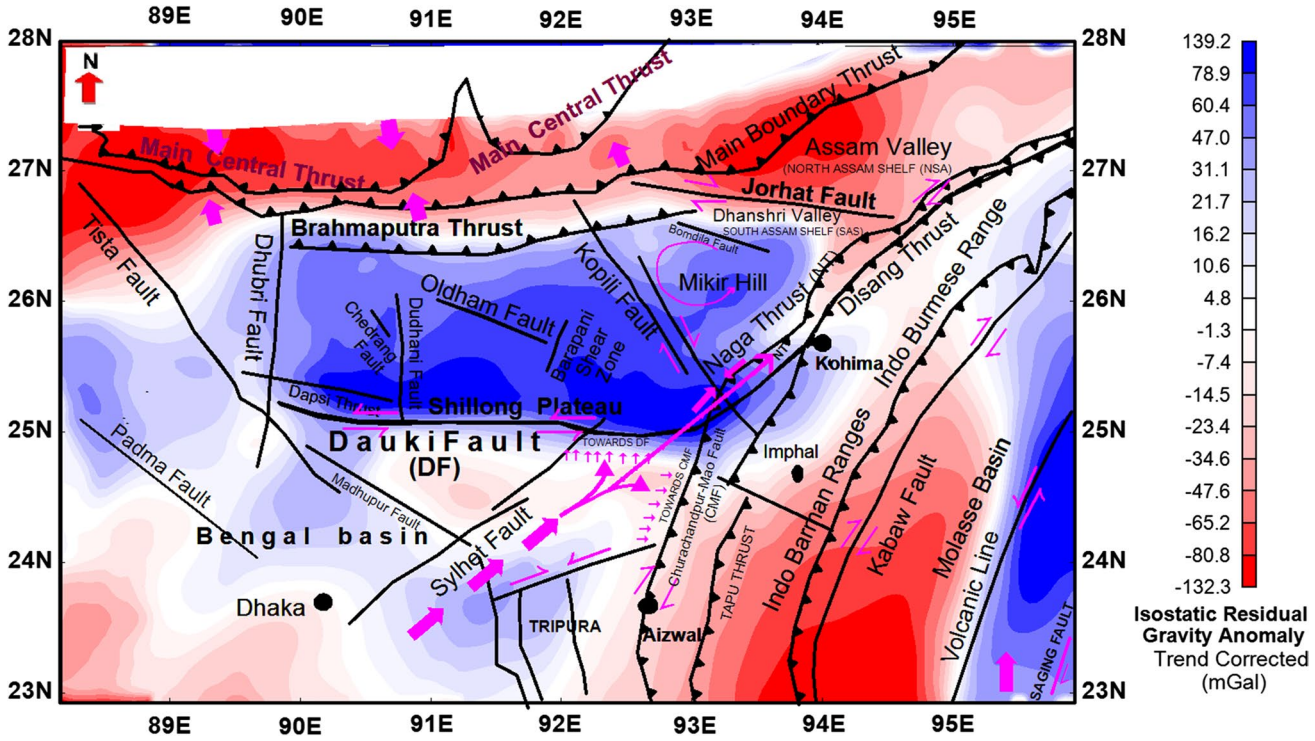


Fig. 10 The corrected isostatic residual gravity anomaly calculated after trend removal. The corrected residual anomaly varies in the range from − 132 to 139 mGal

components in x and y directions. Finally, using these vertical and horizontal derivatives, other gravity gradients are calculated. The details of these calculations are explained in separate chapter.

Methodology

The various techniques for gravity data interpretation are discussed here briefly. THDR, tilt, Tdx and Cos(θ) maps have been discussed to delineate thrust and fault boundaries.

Total horizontal derivative (THDR)

Cordell and Grauch (1985) and Grauch et al. (2001) defined the total horizontal derivative (THDR) as mathematically linked to the horizontal derivatives as shown in Eq. (1).

$$\mathrm{THDR} = \left(\frac{\partial G}{\partial x}\right)^2 + \left(\frac{\partial G}{\partial Y}\right)^2 \quad (1)$$

where G is the corrected residual gravity field anomaly (vertical component), $\frac{\partial G}{\partial x}$ and $\frac{\partial G}{\partial y}$ are two orthogonal horizontal derivatives of the gravity anomaly. The main characteristic of this filter is better mapping of shallower structures and relatively neglecting deeper ones. Numerous researchers (Ferreira et al. 2013; Wang et al. 2009) have used THDR for the interpretation of potential field data. This technique has the ability to detect the edges in complex geological setup and enhance the sharp response at the edges.

Source edge detection (SED)

Source edge detection (SED) method uses different derivatives of gridded gravity data, for determination of the geological boundaries position. Horizontal derivatives can be derived perpendicular to the strike direction of elongated source bodies using least square methodology (Thurston and Brown 1994; Thurston and Smith 1997; Cordell and Grauch 1982). The source edge detection technique is more useful for both near-surface and deep-seated structures. This technique provides the strikes direction and direct structural trend of the elongated source bodies.

Tilt derivative

Tilt of G gradients was first analyzed by Miller and Singh (1994) for identifying a potential field source. Later on, Verduzco et al. (2004) generalized this for gridded and profile dataset. Mathematically, tilt can be expressed as the inverse tangent of the ratio of vertical to total horizontal derivative of G as expressed in Eq. (2)

$$\text{Tilt} = \tan^{-1}\left(\frac{\text{VDR}}{\text{THDR}}\right) \quad (2)$$

where tilt varies in the range $-\pi/2$ to $+\pi/2$. VDR is the vertical derivative, and THDR is the Total horizontal derivative. Tilt derivatives indicate zero values over the source edges and help to trace the outline edges. The similar type of work has been carried out by the various workers (Salem et al. 2007, 2008, 2010; Ferreira et al. 2013; Lahti and Karinen 2010; Fairhead et al. 2011; Ghosh 2016a).

Horizontal tilt angle (TDX)

The horizontal tilt angle is the normalization of the amplitude of the total horizontal derivative by the vertical derivative. Cooper and Cowan (2006) introduced TDX using THDR and absolute value of VDR as shown in the mathematical expression (3).

$$\text{TDX} = \tan^{-1}\left(\frac{\text{THDR}}{|\text{VDR}|}\right) \quad (3)$$

TDX is varying in the range $-\pi/2$ to $+\pi/2$ similar to tilt derivative method. Both the methods TDX and TDR show a contrast variation along the boundaries; however, TDX analysis shows more contrast along the boundaries. Different workers have carried out TDX derivative analysis on potential filed data for source edge detection (Fairhead and Williams 2006; Coraggio et al. 2012; Phillips 2000).

Analytical signal amplitude

The analytical signal (ASA) can be expressed for a 3D structure of the vertical gravity gradient as stated by Marson and Klingele (1993) is as follows:

$$\text{ASA} = \sqrt{(\partial G/\partial x)^2 + (\partial G/\partial y)^2 + (\partial G/\partial z)^2} \quad (4)$$

where $\partial G/\partial z$ is the vertical derivative of the gravity field and $\partial G/\partial x$ and $\partial G/\partial y$ are the two horizontal derivatives. ASA is more useful for delineating edges of gravity sources because ASA shows peak over gravity sources.

Cos(θ) map

Wijns et al. (2005) have used the application of Cos(θ) map which is the ratio of THDR to normalized analytical signal for the detection of edge of the causative bodies using potential field data as shown in Eq. (4).

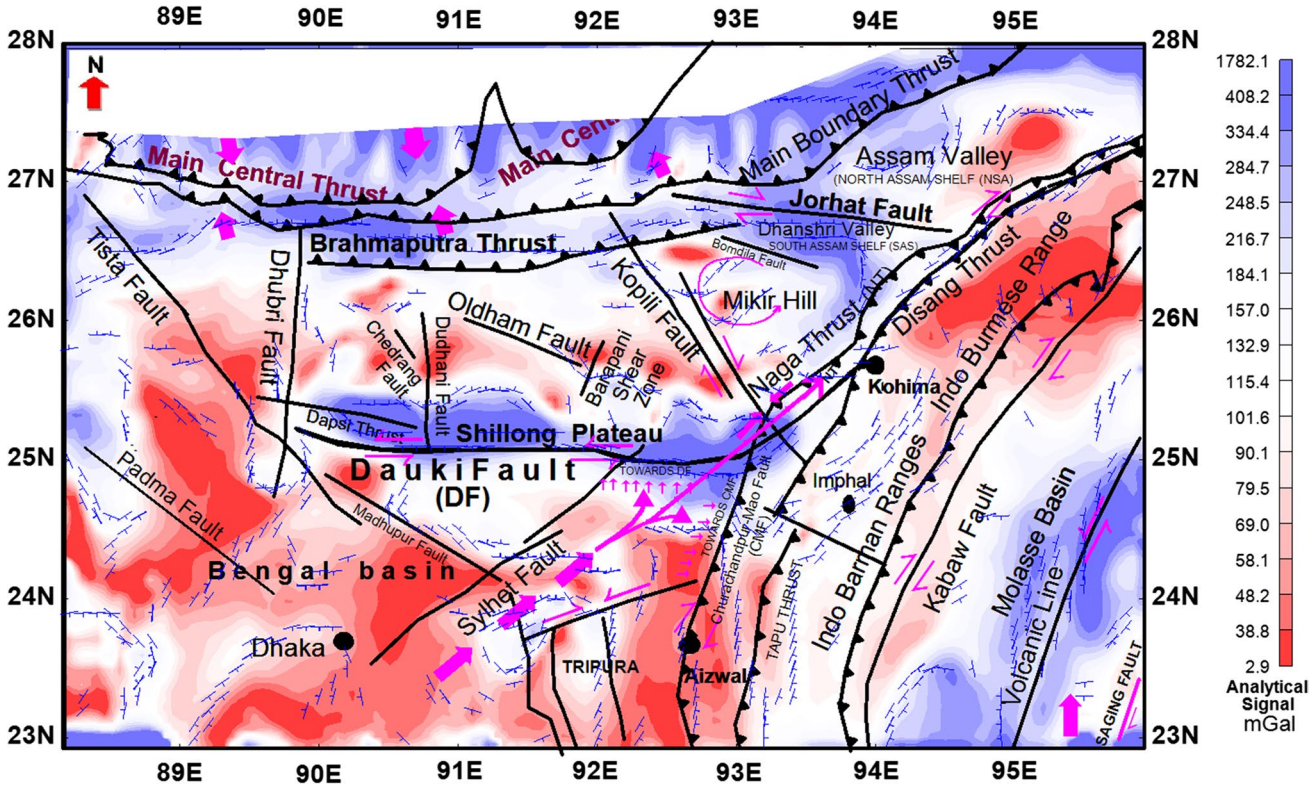


Fig. 11 The map of analytical signal (AS, i.e., the total gradient of the residual field vertical component) of gravity anomaly outlining the contrasts of anomalous density distribution. High value of AS is observed at Shillong Plateau, near MBT/MCT, Jorhat fault and at Molasse Basin and volcanic line. Low values of AS signals are observed at Bengal Basin, Assam Valley and Indo-Burma Ranges

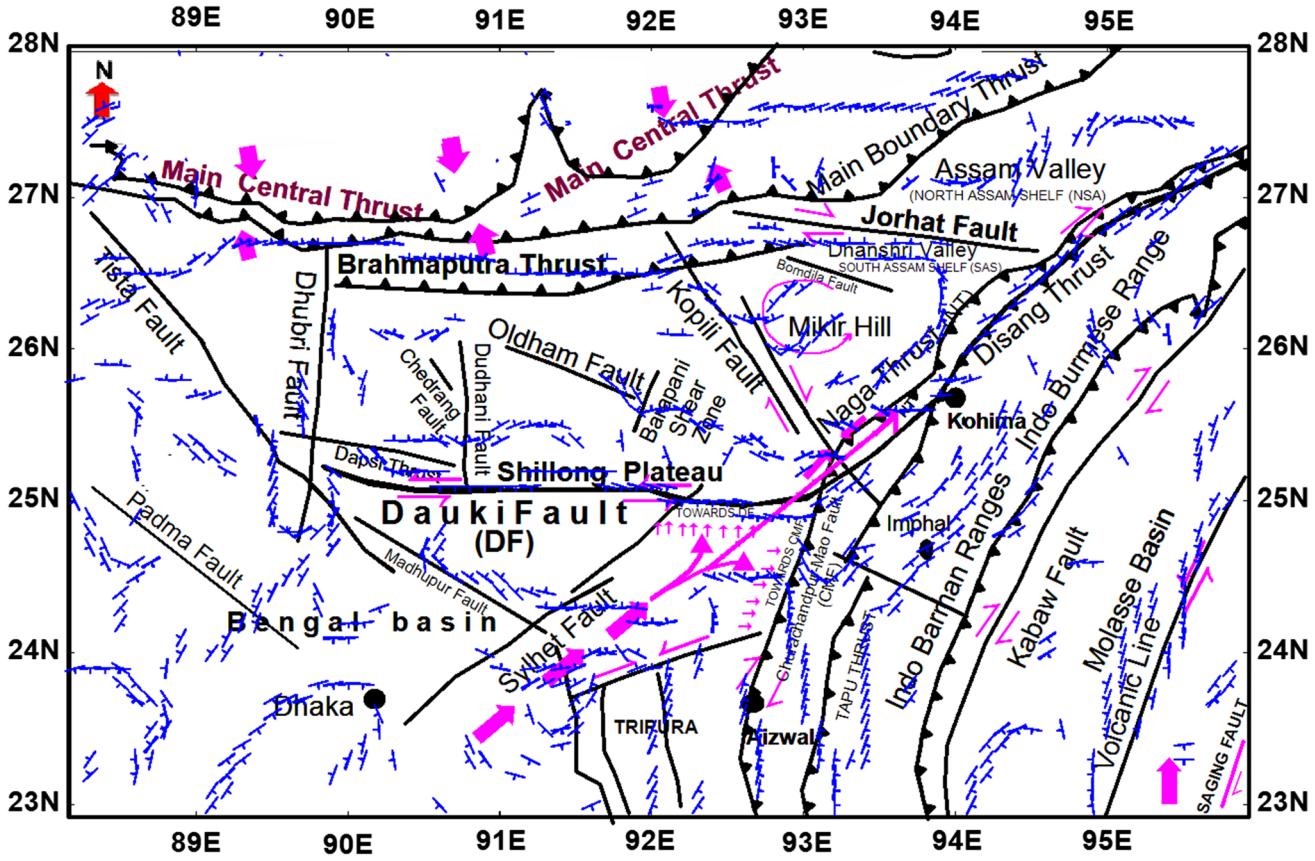


Fig. 12 Location of thrust fault determined using source edge detection technique (SED) based on source parameter imaging. ⊥ locations are marked by blue line with dip and strike indications. These gravity-derived locations are superimposed on the known outcropping. The gravity analysis identified many hidden thrust-fault in this area, which also were not marked earlier

$$\mathrm{Cos}(\theta) = \frac{\mathrm{THDR}}{|A|} \tag{5}$$

where

$$\mathrm{ASA} = \sqrt{(\partial G/\partial x)^2 + (\partial G/\partial y)^2 + (\partial G/\partial z)^2} \tag{6}$$

where $|A|$ indicates the amplitude of the 3D analytical signal amplitude (ASA). Similar types of work have been carried out by the various potential researchers for better source edge location and identification of potential field data boundaries (Fairhead and Williams 2006; Fairhead et al. 2007; Reid 2007).

Gravity data interpretation

Original Bouguer gravity data (Verma and Mukhopadhay 1977) were digitized and re-gridded by Ghosh et al. (2015). To interpret the gravity data, separation into regional and residual parts was performed. Conventional gravity data interpretation using Bouguer gravity anomaly (Fig. 6) or isostatic residual gravity data interpretation (trend corrected) (Fig. 10) is unable to mark the thrust and fault location properly. Previously, 2.5-dimensional modeling studies along with the few profiles across the Brahmaputra Thrust and Dauki Fault in the N–S direction had been carried out by Ghosh et al. (2015) to understand the thickness of basement depth across the Brahmaputra Thrust and Dauki Fault. However, 2.5-dimensional data interpretation using limited profile data interpretation may not provide better results compared to the three-dimensional modeling. The results obtained by using analytical signal amplitude (ASA), total horizontal derivative (THDR), tilt derivative (TILT), horizontal tilt angle (TDX) and Cos(θ) analysis for delineating the thrust fault are more appropriate for interpreting thrust fault interpretation. SED technique is used to demarcate better geological boundary in strike direction along with dip direction. Similar type of studies has been used in SED technique provided fruitful result (Ghosh 2016a, b). The demarcation of identified thrust faults of the area had been marked by black color (Fig. 6) by the potential workers (Verma and Mukhopadhay 1977; Bhattacharya et al. 2008; Baruah and

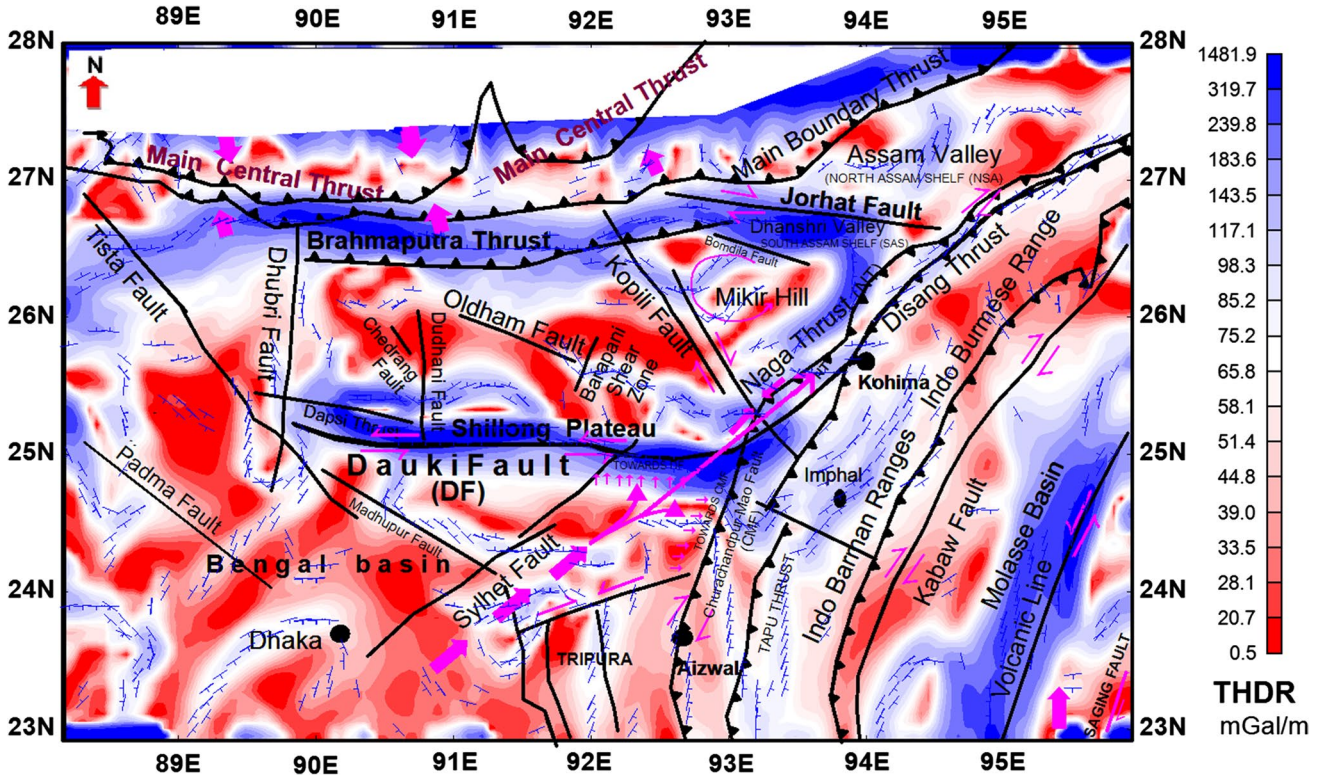


Fig. 13 Total horizontal derivative (THDR) map suggesting the edge boundaries. The ridges of the THDR map indicate approximate location of the major edges. The pattern of the color contrast in the THDR map follows the thrust fault pattern as marked. Location of thrust fault determined using source edge detection technique (SED) is superimposed for comparison

Hazarika 2008; Yadav et al. 2009). The gravity-interpreted thrust fault locations are superimposed to the known thrust fault marked by black color for comparative study. Some additional thrust faults (marked by blue color, Fig. 12) were also identified using SED techniques in this area which is not marked so far. It is studied that the derived thrust faults are following the trend of the known thrust fault derived by the previous workers. Pink arrow (Fig. 12) signifies various contractions' force (along Brahmaputra Thrust, MBT and MCT) at the northern Himalayas. Diverting thrust is acting in between Dauki Fault and Churachandpur-Mao Fault (CMF) and further extending toward Naga Thrust (NT). Various strike slip faults are marked by arrow (Sagaing Fault, Kopili Fault, Kabaw Fault, Naga Thrust, Dauki Fault, Jorhat Fault, etc.).

Analytical signal amplitude map (ASA) (Fig. 11) suggests the lineament pattern in the form of wide zone but cannot provide the detailed thrust fault information; however, isolines of zero tilt value precisely follow the thrust faults lines. Nevertheless, in many parts of the area, additional new source edges are notices. Various smaller and hidden faults might be possible and need further attention. However, the present derived result shows good correlation with the existing thrust fault location.

Source edge detection (SED) technique has been used for source parameter imaging and for determination thrust fault locations (Fig. 12). The dip and strike positions are calculated in each grid point from the gravity signature either from positive or negative anomaly. SED map is indicated with the symbol "⊥" indicating dip and strike of the source bodies. The vertical line indicates the dip direction upright on the strike direction (Fig. 12). The long axis (horizontal line) indicates strike direction of the edges (contact). The map covers all the strike directions varying from $0° \leq \text{strike} \leq 90°$, $90° \leq \text{strike} \leq 180°$, $180° \leq \text{strike} \leq 270°$ to $270° \leq \text{strike} \leq 360°$ and is measured in the clockwise direction ranging from 0° to 360°. The blue color indicates the derived thrust fault along with the strike and dip directions. Previously identified thrust fault locations are marked by black color (Fig. 3). It is noted that derived thrust fault locations are matching with known thrust fault (Fig. 12).

It is noted that Bouguer gravity anomaly, residual gravity anomaly and other derivatives are unable to detect thrust fault locations except providing the pattern of thrust faults. The total horizontal derivative (THDR) map (Fig. 13) is recommended for the identification of the source edge boundaries. The total horizontal derivative shows positive

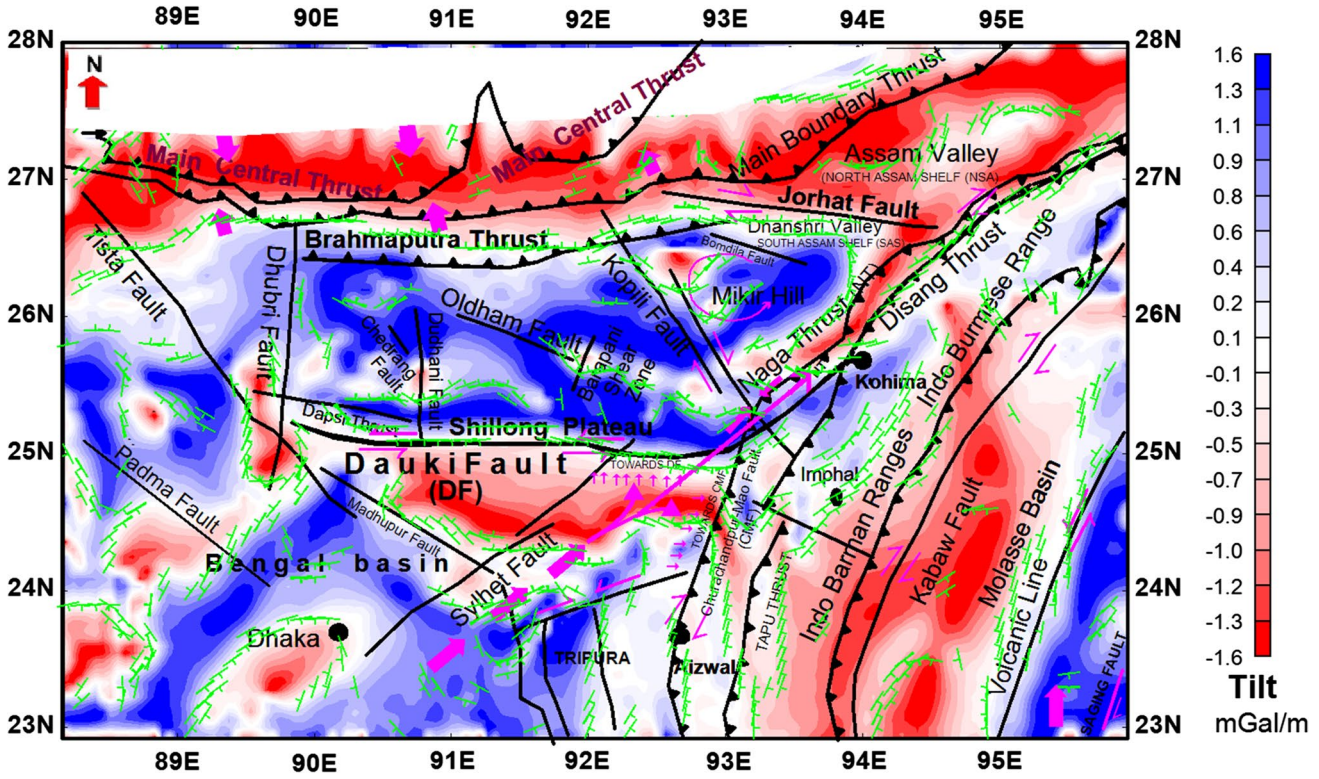


Fig. 14 Tilt derivative (tilt angle, i.e., the tilt of the vector of anomaly G) map suggesting the edge boundaries shown by green symbols. High and low value colors follow the lineament zone, and the zero isolines of tilt angle significantly correlate with the major fault lines of the area. Location of thrust fault determined using source edge detection technique (SED) is superimposed for comparison

values (marked in blue color) indicating the edges of the source bodies (Fig. 13). The thrust fault locations derived from the gravity data show good correlation with geological observations marked by the earlier authors (Verma and Mukhopadhay 1977; Bhattacharya et al. 2008; Baruah and Hazarika 2008; Yadav et al. 2009). However, in many parts of the area, additional edges noticed might correspond to the smaller or hidden faults which need further study.

Tilt derivative (TDR) has been calculated using VDR and THDR, and its value varies in the range from -1.6 to $+1.6$ ($-\pi/2$ to $+\pi/2$ rad) (Fig. 14). The high values are indicated by blue color and low are shown by red color in the tilt derivative map. The thrust fault locations were indicated by green color, and they lie in between the blue and red colors. The color zone pattern (strike of isolines) of TDR map is accordant with the known (outcropped) thrust and faults and gravity-derived thrust faults determined by the use of SED technique (Fig. 12).

Similarly, the horizontal tilt angle (TDX) map is calculated using THDR and VDR. The derived thrust faults from SED maps are superimposed on the TDX map (Fig. 15). These TDX values are also varying in the range of -1.6 to $+1.6$ ($-\pi/2$ to $+\pi/2$ rad). TDX results represent the lineament pattern and more sharply expressed compared to the TDR over the edges body. The maximum value is indicated in black color and minimum value in orange color as shown in the color scale. The thrust fault locations are proposed in between black and orange colors which is shown by superimposing the SED location (green color, Fig. 15). thrust fault locations using TDX derivative show a good correlation with the known thrust fault locations.

The Cos(θ) map is calculated using total horizontal derivative and the analytical signal. The color scale bar shows that Cos(θ) values are varying from 0 to $+1$. The derived thrust and faults from SED techniques (green color) are also superimposed on the Cos(θ) map and show good correlation (Fig. 16). Gravity data interpretation suggests that there are various minor thrust faults identified which may not have been previously marked.

The gravity field pattern looks similar to that in the derivative maps, viz. THDR, TDR, TDX and Cos(θ). However, the results derived from Cos(θ) give more prolific solution for identifying thrust fault locations. The above-mentioned methods are very useful for estimating thrust faults in geologically and tectonically complex areas. The Cos(θ) map results are well projected and indicated additional thrust fault locations.

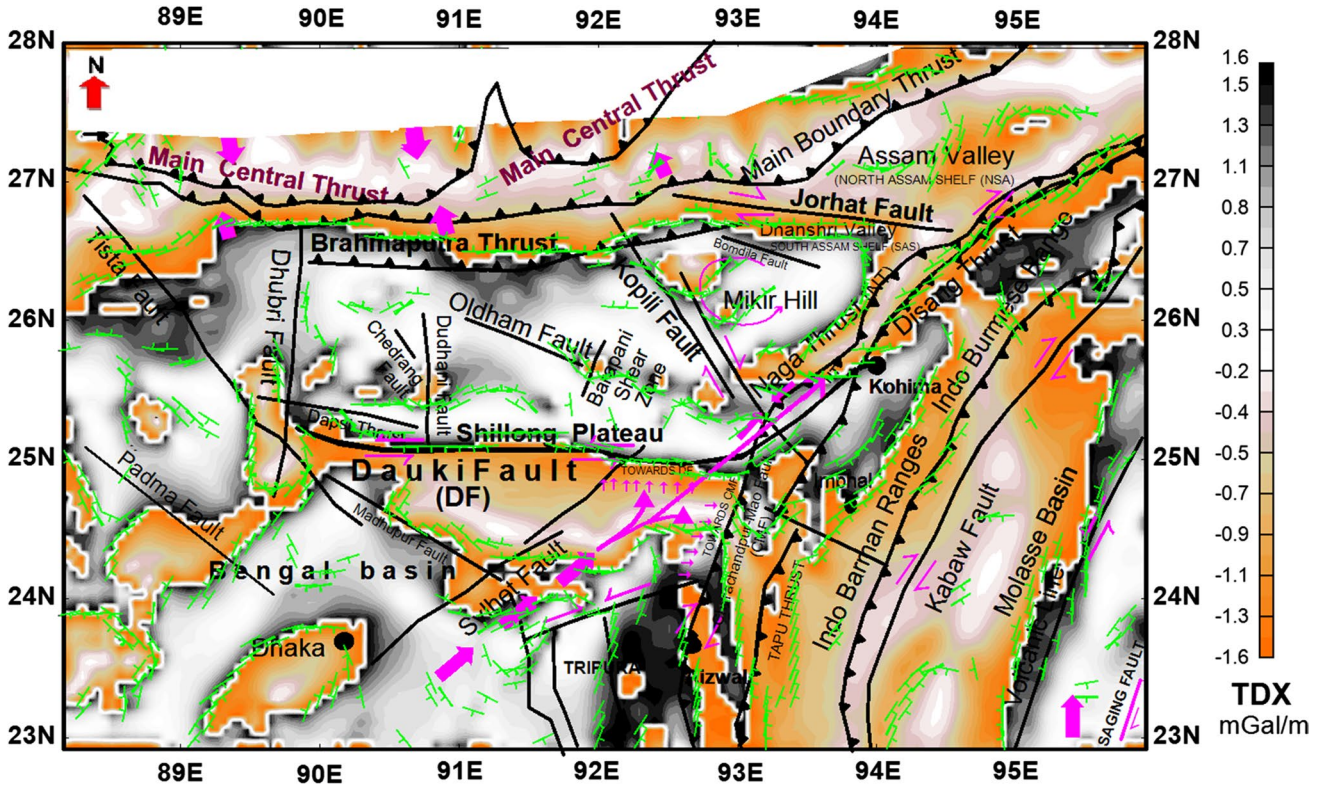


Fig. 15 Horizontal tilt angle (TDX) map also provides the lineament pattern, but it is more sharply expressed compared to the tilt derivative. The thrust faults are indicated by contact between the orange and black colors, and white simply corresponds to zero isoline of the tilt. Location of thrust fault determined using source edge detection technique (SED) is superimposed for comparison

Summary and conclusions

The aim of this paper is to carry out gravity data interpretation to delineate thrust fault locations using latest techniques like THDR derivatives, tilt derivative (TDR), horizontal tilt angle (TDX), Cos(θ) map and SED technique. These interpreted maps correlate well with the earlier derived thrust faults locations. However, among these interpretation, tilt (TDR), especially its zero isoline, and TDX distribution indicate better geological lineaments/structures in the northeastern part of India (Figs. 14 and 15). In addition to this, many hitherto unknown thrust faults were identified by the use of source edge detection technique (Fig. 12) and dip and strike directions were also determined. The location of known outcropped faults which were described in earlier extensive research work was compared to the present gravity-derived results marked in black color (Figs. 12, 13, 14, 15 and 16). There are numerous thrust faults marked in the eastern part of the area, between Naga thrust and volcanic lines. Numerous thrust faults are marked at Kohima, Imphal and Aizawl areas and oriented in S–NNE direction. There are some hidden thrust faults also marked at Molasse Basin separated from the Kabaw Fault and the volcanic lines (Fig. 12). Some unseen buried thrust faults are also detected at the southern part of the Dauki Fault. Similarly, more or less hidden thrust and faults are also perceived at MBT, MCT and Brahmaputra Thrust and oriented in E–W direction. Generally, the source edge detection technique is very useful in mapping thrust faults and dips.

The tectonic activity in this area suggested that the Shillong Plateau is the center of the various activities occurring around it. It may be noted that MBT, MCT and BT are oriented in the E–W direction in the northward part of the Shillong Plateau. The Naga thrust, Disang Thrust, Indo-Burma Ranges, Kabaw Fault and volcanic line are oriented from SW to NNE direction and situated in the eastern part of the Shillong Plateau. The Dauki Fault is oriented in E–W direction and extended westward of the Shillong Plateau. Southward of Dauki Fault/Shillong Plateau some faults are oriented in SW–NNE direction. A Dhubri Fault is oriented in N–S direction, and Tista Fault is oriented in the NW–SE direction, and these faults are situated in the western part of Shillong Plateau. Shillong Plateau is assumed to be divided by various faults; these are Dhubri Fault, Chedrang Fault, Dudhani Fault, Oldham Fault, Kopili Fault (Fig. 3). Jorhat Fault is one of the fault boundaries which separate the extended part of the Shillong Plateau that is Mikir Hills and Assam Valley.

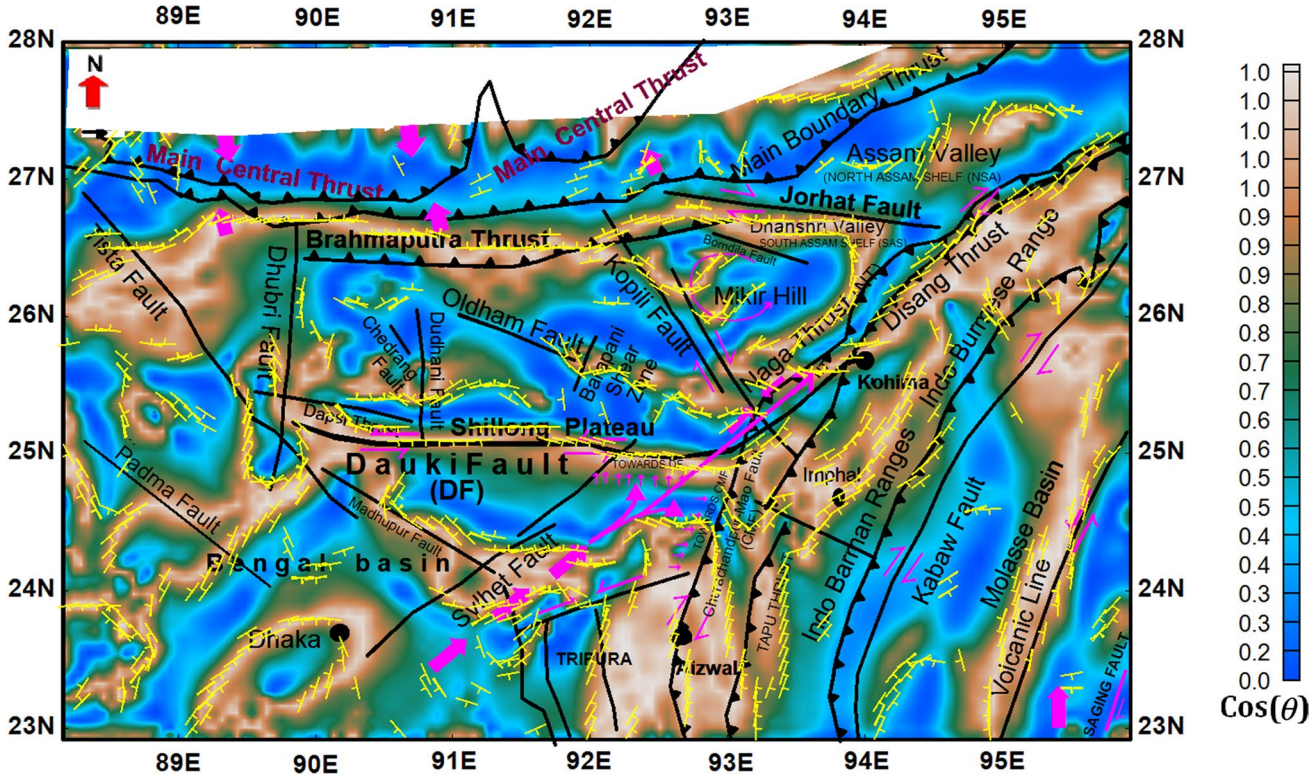


Fig. 16 Cos(θ) map indicates good thrust fault boundaries more prominently and well correlated with the previously identified thrust fault. The Cos(θ) value varies from 0 to 1 where angle (θ) is varying from 0 to $\pi/2$ rad. The highest Cos(θ) values are indicating thrust fault as marked by whitish-gray color. Location of thrust fault determined using source edge detection technique (SED) is superimposed on this map, and it shows good correlation between old thrust fault location and newly derived thrust fault location by Cos(θ) analysis

It is suggested that NE directional force through Bengal Basin extended into two components of force: One is acting toward Dauki Fault (velocity 8–9 mm/year) and another is acting toward Churachandpur-Mao Fault (CMF) (velocity 14 mm/year) and further extended to Naga Thrust (NT) (Panda et al. 2018). Gravity data suggested that Kopili Fault is separated from Shillong Plateau and Mikir Hills. The orientation of Kopili Fault is marked by arrow and as shown in (Fig. 12) (Singh et al. 2017). The Disang thrust and Dauki Fault orientation are also marked in Fig. 12 (Singh et al. 2017). Barman et al. (2016) suggested that some contraction forces are acting between Mikir Hill and Shillong Plateau along the Kopili Fault. Singh et al. (2017) suggested that Mikir Hills are facing an anticlockwise rotational force due to Kopili and Bomdila faults (Fig. 12). A major East–West Lineament (Jorhat Fault) divides Assam shelf into two parts, viz. North Assam Shelf (NAS) and South Assam Shelf (SAS). NAS is generally known as Brahmaputra Valley, and SAS is known as Dhansiri Valley (Singh et al. 2011) (marked by lower gravity value). It is observed that Shillong Plateau has greater gravity value compared to the MBT/MCT (toward north), Assam Valley (NE part), Molasse Basin (SE part) and also has comparatively higher gravity value from Bengal Basin (southward) and Tista and Padma faults (Western part).

It seems that some SW to NE directional force (at Bengal Basin across the Sylhet Fault) are acting toward Shillong Plateau. SE to NW directional forces are acting (originated from Myanmar) toward the Indo-Burma Ranges and further extended to Shillong Plateau. NE–E to SW–W directional force is acting toward Naga Thrust at Assam Valley extending to Jorhat fault and Mikir Hill. NE to SW directional forces are acting from the Main Central Thrust (MCT) toward Brahmaputra Thrust and extending toward Shillong Plateau. NW to SE directional forces are acting along the Tista Fault extending toward Shillong Plateau. The different forces are acting as shown by the arrow marks on the area which causes development of various thrust faults/lineaments and makes this area seismically more active causing changes in gravity field. Geologically it is suggested that there are numerous forces acting toward Shillong Plateau/Indo-Burma Ranges (assuming as center location) subsequently generating an anticlockwise rotational force in this area and Mikir Hills (Fig. 17) (Singh et al. 2017). There may be possibilities that due to these numerous activities, thrust faults are active which causes upliftment and depression in the area.

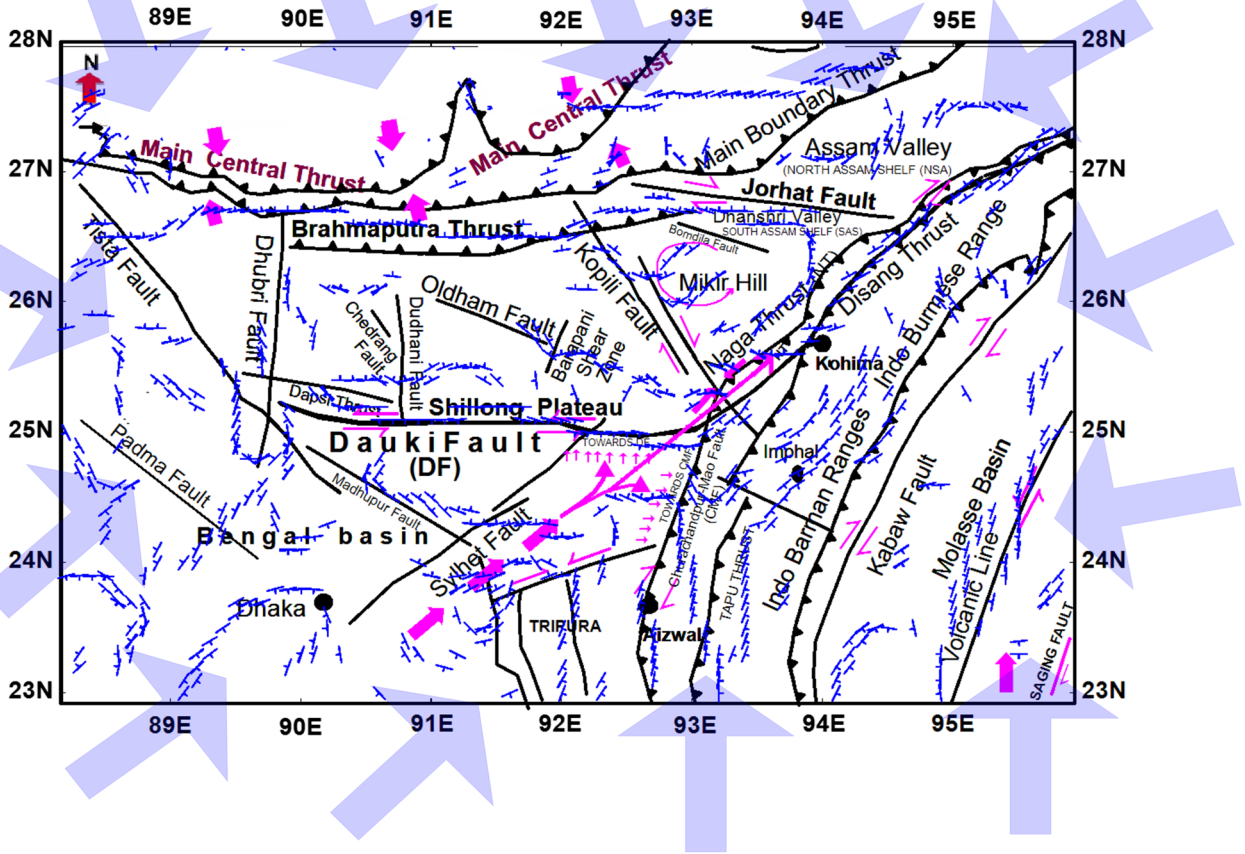


Fig. 17 Various thrust fault locations which are previously identified are shown in black color and newly derived thrust faults are shown in blue color with strike and dip directions. The different forces are acting as shown in by arrow marks on the area which causes development of various thrust faults/lineaments and makes this area seismically more active. Geologically, it is suggested that there are numerous forces acting toward Shillong Plateau/Indo-Burma Ranges (assuming as center location) subsequently generating an anticlockwise rotational force in this area

References

Alam M, Alam MM, Currey JR, Chowdhury MLR, Gani MR (2003) An overview of the sedimentary geology of the Bengal Basin relation to the regional tectonic framework and basin-fill history. Sediment Geol 155:179–208

Ambraseys N, Douglas J (2004) Magnitude calibration of north Indian N. earthquakes. Geophys J Int 159:165–206

Angelier J, Barua S (2009) Seismotectonics in Northeast India: a stress analysis of focal mechanism solutions of earthquakes and its kinematic implications. Geophys J Int 178(1):303–326. https://doi.org/10.1111/j.1365-246X.2009.04107.x

Awasthi N, Ray JS, Singh AK, Band ST, Rai VK (2014) Provenance of the late quaternary sediments in the Andaman Sea: implications for monsoon variability and ocean circulation. Geochem Geophys Geosyst 15:3890–3906. https://doi.org/10.1002/2014GC005462

Bansal BK, Verma M (2013) Science and technology based earthquake risk reduction strategies: the Indian scenario. Acta Geophys 61(4):808–821. https://doi.org/10.2478/s11600-013-0105-5

Barman P, Ray JD, Kumar A, Chowdhury JD, Mahanta K (2016) Estimation of present-day inter-seismic deformation in Kopili fault zone of north-east India using GPS measurements. Geomat Nat Hazards Risk 7:586–599. https://doi.org/10.1080/19475705.2014.983187

Baruah S, Hazarika D (2008) A GIS based tectonic map of north-eastern India. Curr Sci 95(2):176–177

Bhattacharya PM, Mukhopadhyay S, Majumadar RK, Kayal JR (2008) 3-D seismic structure of the northeast India region and its implications for local and regional tectonics. J Asian Earth Sci 33(1–2):25–41. https://doi.org/10.1016/j.jseaes.2007.10.020

Bilham R, England P (2001) Plateau 'pop up' in the great, 1897, Assam earthquake. Nature 410:806–809

Biswas S, Grasemann B (2005) Quantitative morphotectonics of the southern Shillong Plateau (Bangladesh/India). Aust J Earth Sci 97:82–93

Bora DK, Baruah S (2012) Mapping the crustal thickness in Shillong–Mikir Hills Plateau and its adjoining region of northeastern India using Moho reflected waves. J Asian Earth Sci 48:83–92

Borah K, Bora DK, Goyal A, Kumar R (2016) Crustal structure beneath northeast India inferred from receiver function

modeling. Phys Earth Planet Interiors 258:15–27. https://doi.org/10.1016/j.pepi.2016.07.005
Chandra U (1984) Seismicity, earthquake mechanisms, and tectonics of Burma, 200–280N. Geophys J R Astron Soc 40:367–381
Cooper GRJ, Cowan DR (2006) Enhancing potential field data using filters based on the local phase. Comput Geosci 32:1585–1591
Coraggio F, Bernardelli P, Gabbriellini G (2012) Structural reconstruction using potential field data in hydrocarbon exploration. In: SEG Las Vegas 2012 annual meeting. https://doi.org/10.1190/segam2012-0983.1
Cordell L, Grauch VJS (1982) Mapping basement magnetization zones from aeromagnetic data in the San Juan Basin; New Mexico. In: Presented at the 52nd ann. internat. mtg., sot. explor. geophys. Dallas; abstracts and biographies. pp 246–247
Cordell L, Grauch VJS (1985) Mapping basement magnetization zones from aeromagnetic data in the San Juan Basin, New Mexico. In: Hinze WJ (ed) The utility of regional gravity and magnetic anomaly maps. Society of Exploration Geophysicists, Tulsa, pp 181–197. https://doi.org/10.1190/1.0931830346.ch16
Devi NR, Sarma KP (2010) Strain analysis and stratigraphic status of Nongkhya, Sumer and Mawmaram conglomerates of Shillong basin, Meghalaya, India. J Earth Syst Sci 119:161–174
Evans P (1964) The tectonic framework of Assam. J Geol Soc India 5:80–96
Evans P, Crompton W (1946) Geological factors in gravity interpretation by evidence from India and Burma. Q J Geol Soc Lond 102:211–249
Fairhead JD, Williams SE (2006) Evaluating normalized magnetic derivatives for structural mapping. SEG 2006 new orleans extended abstract
Fairhead D, Williams S, Salem AB (2007) Structural mapping from high resolution aeromagnetic data in Namibia using normalized derivatives. In: EGM 2007 international workshop, Italy, April 15–18
Fairhead JD, Salem A, Cascone L, Hammill M, Masterton S, Samson E (2011) New developments of the magnetic tilt-depth method to improve structural mapping of sedimentary basins. Geophys Prospect 59:1072–1086. https://doi.org/10.1111/gpr.2011.59.issue-6
Ferreira FJF, de Castro LG, Bongiolo ABS, de Souza J, Romeiro MAT (2013) Enhancement of the total horizontal gradient of magnetic anomalies using tilt derivatives: Part II—application to real data. In: 81st Annual international meeting. SEG, Expanded Abstracts, pp 887–891
Ghosh GK (2016a) Interpretation of gravity data using 3D euler deconvolution, tilt angle, horizontal tilt angle and source edge approximation of North-West Himalaya. J Acta Geophysica 64(4):1112–1138. https://doi.org/10.1515/acgeo-2016-0042. ISSN: 1895-6572, Print
Ghosh GK (2016b) Magnetic data interpretation for the source-edge locations in parts of the tectonically active transition zone of the Narmada-Son Lineament in Central India. Pure Appl Geophys 173(2):555–571, 1420–9136. https://doi.org/10.1007/s00024-015-1082-1. ISSN 0033-4553, Print
Ghosh GK, Dasgupta R, Reddy BJ, Singh SN (2015) Gravity data interpretation across the Brahmaputra Thrust and Dauki Fault in the north-Eastern India. J Geophys 7(1):31–38. ISSN: 2230-9497, print
Grauch VJS, Hudson MN, Minor SA (2001) Aeromagnetic expression of faults that offset basin fill, Albuquerque basin, New Mexico. Geophysics 66:707–720
Gulatee BL (1956) Gravity data in India. Suv India Tech Pap 10:l–95
Harijan N, Sen AK, Sarkar S, Das JD, Kanungo DP (2003) Geomorphotectonics around the Sung valley carbonatite complex, Shillong plateau, NE India: remote sensing and GIS approach. J Geol Soc India 62(1):103–109
Kayal JR (2001) Microearthquake activity in some parts of the Himalaya and the tectonic model. Tectonophysics 339:331–351
Kayal JR, De R (1991) Microseismicity and tectonics in northeast India. Bull Seismol Soc Am 77:1718–1727
Kayal JR, Arefiev SS, Barua S, Hazarika D, Gogoi N, Kumar A, Chowdhury SN, Kalita S (2006) Shillong plateau earthquakes in northeast India region: complex tectonic model. Curr Sci 91:109–114
Kumar MR, Raju PS, Devi UE, Saul J, Ramesh DS (2004) Crustal structure variations in northeast India from converted phases. Geophys Res Lett 31:17605. https://doi.org/10.1029/2004gl020576
Lahti I, Karinen T (2010) Tilt derivative multiscale edges of magnetic data. Lead Edge 29:24–29. https://doi.org/10.1190/1.3284049
Marson I, Klingele EE (1993) Advantages of using the vertical gradient of gravity for 3-D interpretation. Geophysics 58(11):1588–1595
Mcclay KR, Bonora M (2001) Analog models of restraining stopovers in strike–slip fault systems. Am Assoc Pet Geol Bull 85:233–260
Miller HG, Singh V (1994) Potential field tilt—a new concept for location of potential field sources. J Appl Geophys 32:213–217
Mishra UK, Sen S (2001) Dinosaur bones from Meghalaya. Curr Sci 80:1053–1056
Mitra SK (1998) Structure, sulphide mineralization and age of the Shillong group of rocks, Meghalaya. In: Krishnan SK (ed) Centenary commemorative national seminar. 1–2 November, 1998, Calcutta, pp 118–119 (Abstract)
Mitra S, Pristley K, Bhattacharya A, Gaur VK (2005) Crustal structure and earthquake focal depths beneath northeastern India and southern Tibet. Geophys J Int 160:227–248
Mukhopadhyay M (1974) Relationship of gravity and seismicity to tectonics in Northeastern India. Unpubl. Ph.D. thesis, Indian School of Mines, Dhanbad
Nandy DR (2001) Geodynamics of Northeastern India and the adjoining region. ACB Publication, Calcutta, p 209
Nandy DR, Dasgupta S (1991) Seismotectonic domains of north-eastern India and adjacent areas. Phys Chem Earth 1:371–384
Nayak GK, Rao VK, Rambabu HV, Kayal JR (2008) Pop-up tectonics of the Shillong plateau in the great 1897 earthquake (Ms 8.7): insights from the gravity in conjunction with the recent seismological results. Tectonics 27:1–8. https://doi.org/10.1029/2006TC002027
Oldham T (1883) The Cachar earthquake of 10 January 1869. The Geological Survey of India Memoir, pp 1–88
Oldham RD (1899) Report of the great earthquake of 12th June. Mem Geol Surv India 46:257–276
Panda D, Kundu B, Santosh M (2018) Oblique convergence and strain partitioning in the outer deformation front of NE Himalaya. Sci Rep 8(1):10564. https://doi.org/10.1038/s41598-018-28774-3
Phillips JD (2000) Locating magnetic contacts: a comparison of the horizontal gradient, analytic signal, and local wavenumber methods. SEG expanded abstracts
Qureshy MN (1971) Relation of gravity to elevation and rejuvenation of blocks in India. J Geophys Res 76:545–557
Rajendran CP, Rajendran K, Daurah BP, Earnest A (2004) Interpreting the style of faulting and paleoseismicity associated with the Shillong, northeast India, and earthquake: implication for regional tectonism. Tectonics 23, TC4009. https://doi.org/10.1029/2003tc001605
Rajesekhar RP, Mishra DC (2008) Crustal structure of Bengal Basin and Shillong Plateau: extension of Eastern Ghat and Saatpura Mobile Belts to Himalayan fronts and seismotectonics. Gondwana Res. https://doi.org/10.1016/j.gr.2007.10.009
Rao NP, Kumar P, Tsukuda T, Ramesh DS (2006) The devastating Muzaffarbad earthquake of 8 October 2005: new insights into Himalayan seismicity and tectonics. Gondwana Res 9:365–378

Reid A (2007) Semi-automated methods of potential field interpretation—innovations, and recent and future developments. In: EGM 2007 international workshop. Capri, Italy, April 15–18, 2007

Robinson RAJ, Brezina CA, Parrish RR, Horstwood MSA, Oo NW, Bird MI, Thein M, Walters ASG, Oliver JH, Zaw K (2014) Large rivers and orogens: the evolution of the Yarlung Tsangpo-Irrawaddy system and the eastern Himalayan syntaxis. Gondwana Res 26:112–121

Salem A, Williams S, Fairhead JD, Ravat D, Smith R (2007) Tilt-depth method. A simple depth estimation method using first-order magnetic derivatives: SEG The Leading Edge 26(12):1502–1505

Salem A, Williams S, Fairhead JD, Smith R, Ravat DJ (2008) Interpretation of magnetic data using tilt-angle derivatives. Geophysics 73:L1–L10

Salem A, Williams S, Samson E, Fairhead D, Ravat D, Blakely RJ (2010) Sedimentary basins reconnaissance using the magnetic tilt-depth method. Explor Geophys 41:198–209

Schellart WP, Nieuwland DA (2003) 3D evolution of a pop-up structure above a double basement strike–slip fault: some insights from analogue modelling. In: Nieuwland DA (ed) New insights into structural interpretation and modelling. Geological Society Special Publications, London, pp 169–179. https://doi.org/10.1144/GSL.SP.2003.212.01.11

Seno T, Rehman HU (2011) When and why the continental crust is subducted: examples of Hindu Kush and Burma. Gondwana Res 19:327–333

Singh RK, Bhaumik P, Akhtari MDS, Singh HJ, Mayor S, Asthana M (2011) Gondwana sediments and their hydrocarbon prospectivity in Dhansiri Valley, Assam and Assam Arakan Basin—India. GEO-India, Greater Noida, New Delhi, India, January 12–14

Singh AP, Purnachandra NR, Ravi Kumar M, Hsieh MC, Zhao L (2017) Role of the Kopili fault in deformation tectonics of the Indo-Burmese arc inferred from the rupture process of the 3 January 2016 Mw 6.7 Imphal Earthquake. Bull Seismol Soc Am 107(2):1041–1047. https://doi.org/10.1785/0120160276

Tandon AN (1954) A study of Assam earthquake of August 1950 and its aftershocks. Indian J Meteorol Geophys 5:95–137

Tapponnier P, Peltzer G, Le Dain AY, Armijo R, Cobbold P (1982) Propagating extrusion tectonics in Asia: new insights from simple experiments with plasticine. Geology 10:611–616

Thurston JB, Brown RJ (1994) Automated source-edge location with a new variable pass-band horizontal-gradient operator. Geophysics 59:546–554

Thurston JB, Smith RS (1997) Automatic conversion of magnetic data to depth, dip, and susceptibility contrast using the SPI (TM) method. Geophysics 62:807–813

USGS Report (2015) The Himalayas: two continents collide. https://pubs.usgs.gov/publications/text/himalaya.html

Verduzco B, Fairhead JD, Green CM, Mackenzie C (2004) New insights into magnetic derivatives. Lead Edge 22:116–119. https://doi.org/10.1190/1.1651454

Verma RK, Mukhopadhay M (1977) An analysis of the gravity field in north-eastern India. Tectonophysics 42:283–317

Wang W, Pan Y, Qiu Z (2009) A new edge recognition technology based on the normalized vertical derivative of the total horizontal derivative for potential field data. Appl Geophys 6(3):226–233. https://doi.org/10.1007/s11770-009-0026-x

Wijns C, Perez C, Kowalczyk P (2005) Theta map: edge detection in magnetic data. Geophysics 70:L39–L43

Wikipedia (2019) Indian Plate (after https://en.wikipedia.org/wiki/Indian_Plate)

Woollard GP (1962) Relation of gravity anomalies to surface elevation, crustal structure and geology. Univ Wise Geophys Polar Res Gen Rep 62–9:292

Worzel JL, Shurbet GL (1955) Gravity interpretations from standard oceanic and continental crustal sections. Geol Soc Am Spec Pap 62:87–100

Yadav RBS, Bormann P, Rastogi BK, Das MC, Chopra S (2009) A homogeneous and complete earthquake catalog for Northeast India and the adjoining region. Seismol Res Lett 80(4):609–627. https://doi.org/10.1785/gssrl.80.4.609

Acta Geophysica (2019) 67:1297–1306
https://doi.org/10.1007/s11600-019-00319-w

RESEARCH ARTICLE - APPLIED GEOPHYSICS

Assessment of signal processing methods for geomagnetic precursor of the 2012 M6.9 Visayas, Philippines earthquake

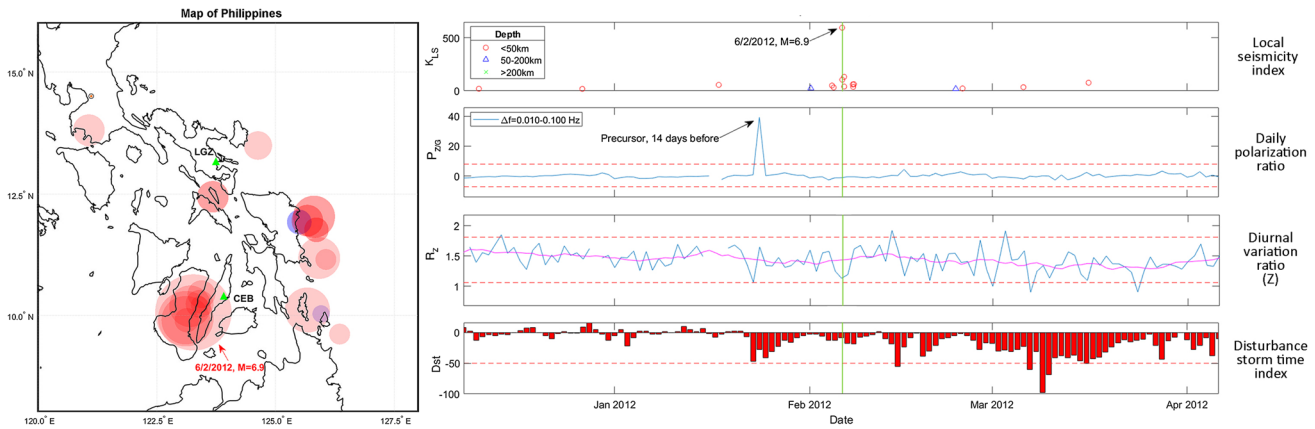
Khairul Adib Yusof[1,2] · Nurul Shazana Abdul Hamid[1,3] · Mardina Abdullah[1,4] · Suaidi Ahadi[5] · Akimasa Yoshikawa[6]

Received: 22 July 2018 / Accepted: 25 June 2019 / Published online: 2 July 2019

Abstract

This study assessed two signal processing methods on geomagnetic data to detect precursory signals appearing before the M6.9 Visayas, Philippines earthquake on 6 February 2012. It aimed to compare the polarization ratio analysis method with the diurnal variation ratio method in terms of reliability and effectiveness. The geomagnetic data were obtained from the MAGDAS magnetometer network for Cebu (CEB) and Legazpi (LGZ) stations which served as the primary and remote stations, respectively. The polarization ratio analysis method was performed on the primary station data to obtain power spectral density in an ultra-low-frequency range before finding the ratio of vertical to total horizontal component. Meanwhile, the diurnal variation ratio method was used to calculate the difference between the daily maximum and minimum values. Then, the ratio of the daily differences of the primary station to the remote station for each individual component was calculated. The disturbance storm time index which describes global geomagnetic activity originating from the Sun was utilized to verify that any observed geomagnetic fluctuations were not caused by solar-terrestrial effect. A precursory anomaly was found using the polarization ratio analysis method which appeared 2 weeks before the earthquake. It is concluded that this method might be effective and reliable in detecting geomagnetic anomalies preceding upcoming earthquakes. In contrast, although the diurnal variation ratio method did show perceivable fluctuations, the running averages were not statistically significant to be considered a precursor. The discrepancy between the analytical results of the two methods may be attributed to the detectability of the earthquake being studied which had a relatively low magnitude. Hence, future studies which utilize more earthquake events need to be conducted to reach a definitive conclusion.

Graphic abstract



Keywords Earthquake precursor · Geomagnetic field · The Philippines · Polarization ratio analysis · Diurnal variation ratio

✉ Nurul Shazana Abdul Hamid
shazana.ukm@gmail.com

Extended author information available on the last page of the article

Introduction

Earthquakes are undoubtedly one of the deadliest natural disasters known to humankind. This creates the urgency for the development of short-term earthquake forecasting which has a time scale of about a week to a month before the event. Short-term earthquake forecasting is currently unachievable by solely using the seismometer due to its limited sensitivity in detecting the generation of microfractures before an earthquake (Hayakawa 2015). This, therefore, imposes the urgency for a non-seismological approach, and one of the most well-established ways is via geomagnetic disturbance observation (Hayakawa 2015).

Previous works have shown that disturbances in the geomagnetic field provide potential precursor to earthquakes, especially in an ultra-low-frequency (ULF) range (Febriani et al. 2014; Stanica et al. 2018). These works started with the finding by Fraser-Smith et al. (1990) where it was reported that the amplitude of geomagnetic noise at ULF was enhanced 3 h before the 1989 M7.1 Loma Prieta earthquake. They suspected that the noise might be related to the quake. Hayakawa et al. (1996) reconfirmed the existence of anomalous variation in the vertical component of the geomagnetic field 1 month before the 1995 Guam earthquake. Since then, studies in geomagnetic earthquake forecasting have been increasing exponentially.

There are five proposed mechanisms which cause anomalous variation in geomagnetic field as a result of seismic activity including (1) induction of electrical current caused by underground conductivity variation (Sorokin and Pokhotelov 2010), (2) displacements of boundaries between high and low conductive crustal blocks (Dudkin et al. 2003), (3) electrokinetic effect (Fedorov et al. 2001), (4) piezoelectric or piezomagnetic effects (Dudkin et al. 2010), and (5) microfracture electrification (Molchanov and Hayakawa 1998).

While electromagnetic emission propagates indefinitely, in the case of lithospheric emission, attenuation reduces the distance at which the emission can travel and be detected on the ground. This is due to the low-pass filter function of the lithosphere which severely damps the high-frequency component (Prattes et al. 2011). Based on the relationship given by the skin depth effect, for emissions originating from earthquake hypocenters with depth not greater than ~ 160 km, signals at 0.01 Hz can be generally detected by ground geomagnetic stations located close to the epicenters (Prattes et al. 2011). Since ULF has a great skin depth characteristic, it experiences low attenuation and deeper penetration into the crust and this is where the ULF range has proven to be useful (Chauhan et al. 2012). ULF (0.01–0.10 Hz, or any value within this range) emission studied in the form of ratio of vertical component to total horizontal component is commonly referred as polarization ratio analysis (Currie and Waters 2014; Prattes et al. 2011).

Even though the polarization ratio analysis is the primary method of detecting earthquake precursors in previous studies, the requirement for 1-s sampling rate data is sometimes a constraint. This is because high-resolution geomagnetic data are not widely available and are distributed by data providers compared to 1-min data. As an alternative, the diurnal variation ratio method which only requires low-resolution data that has been employed by researchers such as Chen et al. (2010) and Xu et al. (2013) in their studies can be considered. However, the drawback of this method is the need for at least two datasets from different stations. Additionally, both stations must be separated at a reasonably far distance to ensure earthquake effects on the primary station have minimal geomagnetic impact on the far, remote station.

Based on the available literature on this subject, it is found that earthquake forecasting studies in the Southeast Asia region are not as abundant as in other parts of the world despite the high rate of earthquake occurrences. One of the significant earthquakes to have occurred in this century in this region is the 2012 Visayas, Philippines earthquake. The disaster happened on 6 February 2012 at 03:49 UTC off the coast of Negros Oriental, Philippines (9.97°N, 123.14°E), with a magnitude of M6.9. The hypocentral depth was 10 km which caused tsunamis, landslides and structural destructions as well as more than 50 fatalities (Aurelio et al. 2017).

In this study, we assessed two signal processing methods on geomagnetic data for earthquake precursory signal detection. The polarization ratio analysis, which has been demonstrated by prior studies to be effective especially in the ULF range, was our first method. In order to overcome the limitation of high-resolution data unavailability and promote greater flexibility, this study also implemented a second method called the diurnal variation method. We studied the 2012 M6.9 Visayas, Philippines earthquake as the main event and other small-magnitude earthquakes; these will be described in detail in the Instrumentation and data section. The effectiveness and reliability of these methods were compared in addition to their distinguishing features.

Instrumentation and data

Earthquake data

The period of observation in this study covered 60 days before until 60 days after the main shock which is from 8 December 2011 until 6 April 2012. The studied earthquake events included all $M > 5.0$ earthquakes which struck within 300 km from both Cebu (CEB) and Legazpi (LGZ) magnetometer stations. Earthquake data were obtained from the European-Mediterranean

Seismological Centre (EMSC) online earthquake catalog (www.emsc-csem.org) and are listed inTable 1.

Characteristics such as magnitude and distance are very important to quantize the impact of each earthquake toward local geomagnetic field. In order to achieve this, local seismicity index (K_{LS}) introduced by Molchanov and Hayakawa (2008) was utilized where the following relationship is assumed to be valid:

$$K_{LS} = \frac{10^{0.75M}}{R + 100} \quad (1)$$

where M is the earthquake magnitude and R is the epicentral distance. This index is used as a form of representation for each individual earthquake in the Results and discussion. It is assumed that the earthquakes with the highest K_{LS} value in an earthquake swarm are more likely to be the main source of any precursor detected.

Hayakawa et al. (2007) hypothesized that the following condition needs to be satisfied for the ULF signal emitted from an earthquake epicenter to be detectable by a magnetometer station, based on the distance between the two locations (epicentral distance), R, and the earthquake magnitude, M:

$$0.025R \leq M - 4.5 \quad (2)$$

$$R \leq \frac{M - 4.5}{0.025}$$

Table 1 List of earthquake details in the vicinity of CEB station. In this table, R refers to the distance of the epicenter from CEB station, while R_{max} is the maximum detectable distance calculated by using Eq. (3)

Date and time (UT)	Coordinates	Earthquake magnitude	R (km)	R_{max} (km)
10/12/2011 09:51	11.16°N, 126.06°E	5.1	251.1	24
26/12/2011 20:34	09.61°N, 126.35°E	5.1	279.9	24
17/01/2012 12:55	11.19°N, 125.92°E	5.7	238.2	48
01/02/2012 04:30	11.93°N, 125.50°E	5.2	246.1	28
04/02/2012 13:09	12.04°N, 125.81°E	5.7	279.0	48
04/02/2012 19:29	11.90°N, 125.67°E	5.4	261.0	36
06/02/2012 03:49	**10.06°N, 123.27°E**	**6.9**	**77.6**	**96**
06/02/2012 04:20	10.19°N, 123.20°E	5.7	79.9	48
06/02/2012 10:10	09.93°N, 123.11°E	6.1	99.8	64
06/02/2012 11:33	09.90°N, 123.11°E	5.9	101.4	56
06/02/2012 11:40	09.90°N, 123.13°E	5.2	99.5	28
07/02/2012 20:37	10.24°N, 123.40°E	5.3	57.4	32
07/02/2012 21:59	10.22°N, 123.40°E	5.0	57.9	20
08/02/2012 00:41	10.44°N, 123.44°E	5.3	52.2	32
24/02/2012 05:42	10.03°N, 125.96°E	5.0	227.3	20
25/02/2012 07:01	11.77°N, 125.86°E	5.2	264.3	28
05/03/2012 23:06	12.45°N, 123.68°E	5.4	233.7	36
16/03/2012 07:58	10.11°N, 125.67°E	5.8	194.6	52

The bold row corresponds to the main earthquake

$$R_{max} = \frac{M - 4.5}{0.025} \quad (3)$$

By rearranging the variables in Eq. (2), we obtained Eq. (3) where R_{max} refers to the maximum detectable distance.

From Eq. (3), it is expected that no precursory signal is to be detected further than 20 km for our weakest studied earthquake with $M = 5.0$. Meanwhile, for the main earthquake with $M = 6.9$, the R_{max} was estimated to be 96 km, which is written in bold in Table 1. Thus, the decision to include earthquakes within 300 km was rather an overestimation to widen the forecasting possibility. Figure 1 illustrates the map of the Philippines with all studied earthquakes shown in circles. Circle radii are proportional to earthquake magnitudes, calculated based on relationship shown in Eq. (3). It was hypothesized that any earthquake in which its circle coincides with either magnetometer station would produce detectable ULF emission at the station.

Geomagnetic index

In order to eliminate the effects of global geomagnetic activity and avoid false precursor, this study utilized a global geomagnetic index, namely the disturbance storm time (Dst) index recorded during the same period of observation. It is used to monitor solar wind activities of which any daily value that is lower than − 50 nT indicates an occurrence of geomagnetic storm (Hamid et al. 2009). Anomalies in geomagnetic field during disturbed days should be ignored as it is due to geomagnetic environmental factors (Hasbi et al. 2011).

Geomagnetic field data

The geomagnetic field data were acquired from the Magnetic Data Acquisition System (MAGDAS) magnetometer network. MAGDAS magnetometers are ring-core-type fluxgate that is capable of measuring small-amplitude geomagnetic fluctuations. There are three sensors which measure three geomagnetic field components, i.e., north–south (H), east–west (D) and vertical (Z) components. The data were sampled at 16-Hz which were then arithmetically averaged on-board into 1-Hz sampling frequency (Bello et al. 2017). In this study, we obtained 1-Hz data which were collected during the period of observation from CEB and LGZ stations (Table 2). Based on the distance from the epicenter of the M6.9 earthquake, CEB served as the primary station and LGZ acted as the remote station. Raw data for the whole period of observation are displayed in Fig. 2a–c, in addition to Fig. 2d–f which zoom in on data for 23 January 2012

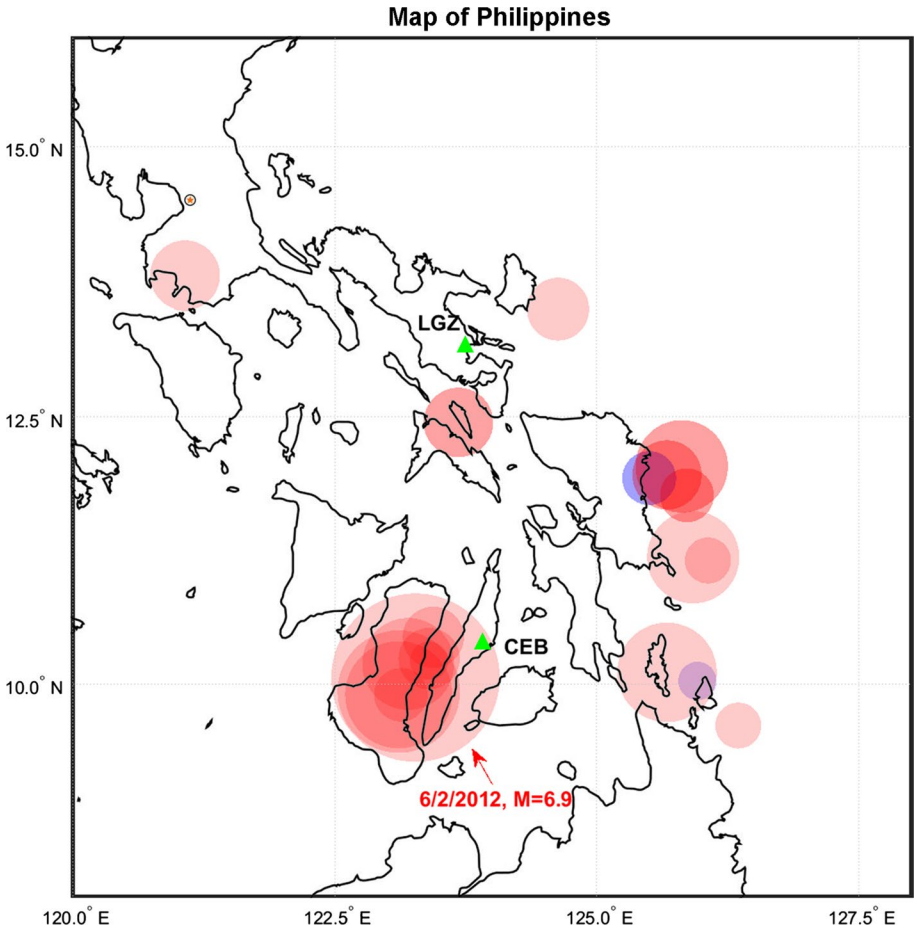


Fig. 1 The location of the magnetometer stations (shown by green triangles) and all included earthquakes. The distance between the two stations is 311 km. Earthquake depths are color-coded; red for shallow (< 50 km) and blue for intermediate earthquakes (50–200 km). Red arrow shows the main earthquake event

Table 2 Details of CEB and LGZ stations and their respective distances from the M6.9 earthquake

Station name (code)	Coordinates	Distance from M6.9 earthquake (km)
Cebu (CEB)	10.36°N, 123.90°E	77.6
Legazpi (LGZ)	13.01°N, 123.74°E	347.2

since an anomaly (when the polarization ratio analysis was applied) appeared on that date.

Methodology on signal processing

Polarization ratio analysis

The following steps were applied to each individual geomagnetic component (H, D and Z) from CEB station. We removed extreme high and low value noises as well as outliers that existed in our raw data by applying median filter. Previous studies suggest that local nighttime data contain less man-made noise (Potirakis et al. 2017; Prattes et al. 2011); thus, only those within the time period of 22:00–02:00 LT were extracted and further processed. Four hours of data were obtained for each day which then underwent the Welch's processing method (the reader is referred to Stranneby and Walker (2004) for mathematical description). This method divides the daily data into smaller segments and then performs the fast Fourier transform (FFT) analysis to each segment. The FFT converts time series data into the frequency domain in the form of power spectrum density (PSD). The Hamming function was used for windowing and the number of FFT bin was $2048/2 + 1$. Neighboring segments (eight segments for each day) were not overlapped and the length of the segment was 1800, which represented 30-min of data. All eight segments were then averaged to obtain one PSD spectrum for each day. The values in frequency of between 0.01 and 0.10 Hz were picked from the spectrum and then averaged over this frequency range to acquire the daily mean value, μ_{day}. To exclude monthly trend, normalization process was performed. Monthly means (μ_{month}) and monthly standard deviations (σ_{month}) were calculated from

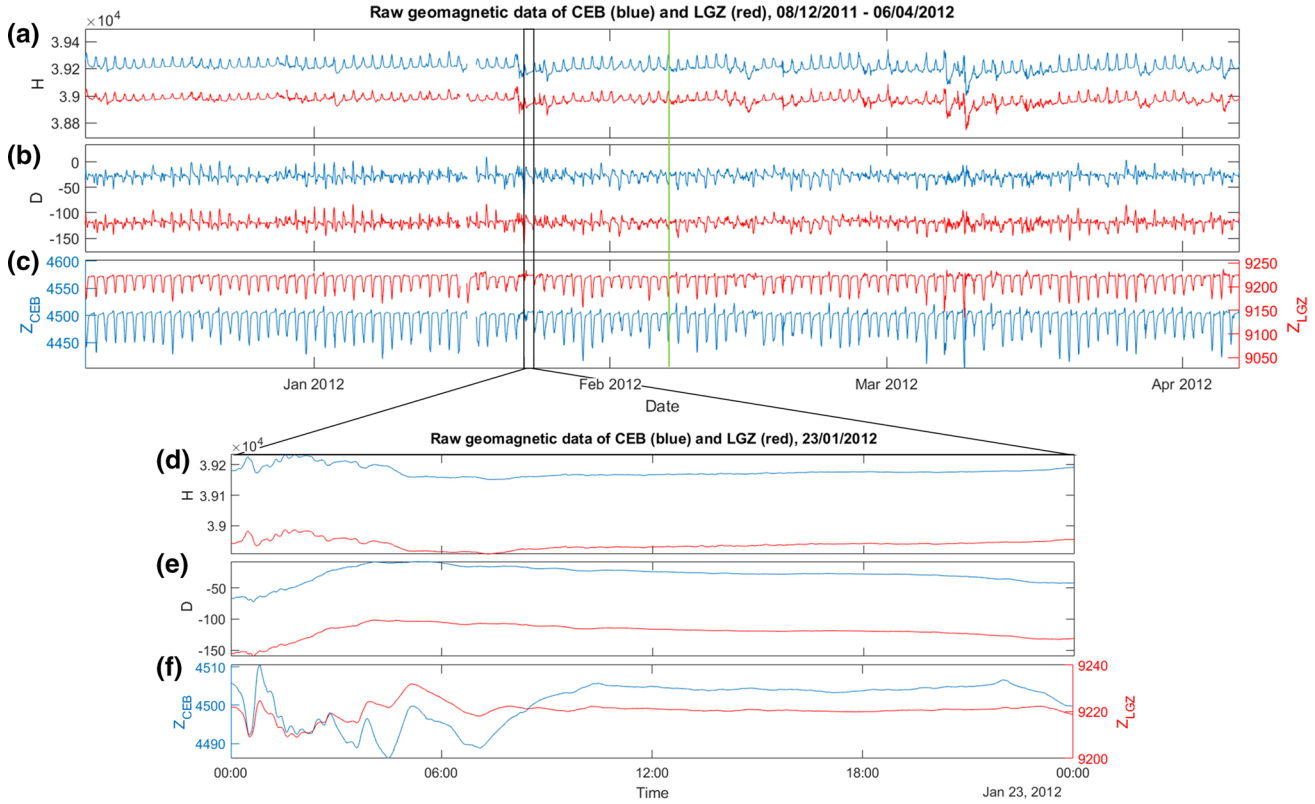


Fig. 2 Raw geomagnetic data from 8 December 2011 to 6 April 2012 (the period of observation) for both CEB (blue line) and LGZ (red line) stations. **a**–**c** show H, D and Z components, respectively. The green vertical line marks the M6.9 earthquake event. Panels **d**–**f** magnify the data on 23 January 2012

μ_{day} values in their respective month: e.g., 8–31 December for December, 1–31 January for January and so on. Then, using these calculated values, normalized daily value, $N_{\text{day},i}$ was calculated by using this formula:

$$N_{\text{day},i} = \frac{\mu_{\text{day},i} - \mu_{\text{month},i}}{\sigma_{\text{month},i}} \tag{4}$$

where i refers to H, D, Z, which implies that each geomagnetic field component was calculated separately. Finally, daily polarization ratio, $P_{Z/G}$ (G indicates the total horizontal component, where $G = \sqrt{H^2 + D^2}$) was calculated using the following expression:

$$P_{Z/G} = \frac{N_{\text{day},Z}}{\sqrt{N^2_{\text{day},H} + N^2_{\text{day},D}}} \tag{5}$$

This ratio is known to exhibit anomalous increment when seismogenic (i.e., associated with earthquakes) emission is present (Chavez et al. 2011). For any given day, it is identified as anomalous when the $P_{Z/G}$ value exceeds $\mu_P \pm 2\sigma_P$ where the deflection is assumed to be statistically significant and not random, as used by previous studies (Febriani et al. 2014; Prattes et al. 2011). Parameters μ_P and σ_P in the expression are the mean and the standard deviation of $P_{Z/G}$, respectively, which were computed from the whole period of observation.

Diurnal variation ratio

The following steps were applied to each geomagnetic component for both stations, namely CEB and LGZ, individually. The noises and outliers' removal process was carried out similar to the previous method. Data were then downsampled from 1-s to 1-min sampling period to demonstrate the possibility of using low-resolution data for this method. Then, the diurnal variation range, $\Delta X_{j,i}$, was calculated by subtracting the minimum value from the maximum daily value:

$$\Delta X_{j,i} = \max(X_{j,i}) - \min(X_{j,i}) \tag{6}$$

where i refers to H, D and Z components and j refers to CEB and LGZ stations. It is important to correctly identify the range for daily data in the local time zone frame as the original data are recorded in universal time (UT). The ratio

of daily variation of the primary (CEB) to the remote (LGZ) station, R_i, was calculated for each component:

$$R_i = \frac{\Delta X_{\mathrm{CEB},i}}{\Delta X_{\mathrm{LGZ},i}} \tag{7}$$

Direct observation of this ratio reveals multiple probable random fluctuations. Therefore, long-term trend of the ratio was obtained, similar to the study by Xu et al. (2013). In their study, 10-day running mean was used; however, in our study, we used 5-day one instead since our period of observation is shorter. For anomaly identification, Xu et al. (2013) used $\mu_{R,i} \pm k\sigma_{R,i}$ ($k=3$) threshold where any daily R_i value exceeding this threshold is considered an anomaly. Parameters $\mu_{R,i}$ and $\sigma_{R,i}$ in the expression are the mean and the standard deviation of R_i, calculated from the whole period of observation. In this paper, $k=2$ was used instead since the earthquake focused in this study has a much lower magnitude than the one in Xu et al. (2013) where $M=9.0$; thus, for our study, a lower threshold is more suitable.

Results and discussion

Polarization ratio analysis

In this section, the results of the processed geomagnetic data using the polarization ratio analysis method are presented. Figure 3 illustrates the daily polarization ratio, $P_{Z/G}$ in the middle panel together with K_{LS} index in the top and Dst in the bottom panels. Temporal evolution of $P_{Z/G}$ (Fig. 3b) was consistently within $\mu_P \pm 2\sigma_P$ threshold throughout the period of observation, except for a clear anomalous peak which appeared 14 days before the main earthquake, which was on 23 January 2012. Raw geomagnetic data as shown in Fig. 2d–f (blue lines) indicate a sudden drop in H and D components; meanwhile, Z component displays an absence of diurnal variation on the date.

Considerably high Dst (Fig. 3c) around the occurrence of possible precursor is acknowledged; however, its value which was above −50 nT suggests uncorrelation between $P_{Z/G}$ and global geomagnetic activity. To further support this claim, geomagnetic storms which began on 7 March 2012 did not cause any increment in $P_{Z/G}$ value.

To closely inspect the apparent anomaly, we observed normalized daily values of Z ($N_{\mathrm{day},Z}$) and $1/G$ ($N_{\mathrm{day},1/G}$) during the same time period; their temporal evolutions are illustrated in Fig. 4a and b, respectively. The green dashed box (in Fig. 4a and b) highlights the period when the anomalous peak appeared, which is shown in Fig. 3b. As shown in Fig. 4a and b, positive value of $N_{\mathrm{day},Z}$ and high value of $N_{\mathrm{day},1/G}$ contribute to the peak. We also note that another high value of $N_{\mathrm{day},1/G}$ to the left of the green box did not produce another peak in $P_{Z/G}$ (Fig. 3b) since $N_{\mathrm{day},Z}$ value during this period is negative, thus canceling the effect of $N_{\mathrm{day},1/G}$.

The choice of frequency range, while is consistently within 0.01–0.10 Hz in most earlier studies, is rather arbitrary in terms of specific frequency range being used by them (Yusof et al. 2019). For example, Schekotov et al. (2013) used 0.03–0.05 Hz, Ida et al. (2008) used ~0.01 Hz, while Prattes et al. (2011) selected 0.010–0.015 Hz for their analyses. Therefore, we also included $P_{Z/G}$ when 0.010–0.015 Hz (low narrowband in Fig. 5c) and when 0.09–0.10 Hz (high narrowband in Fig. 5d) were used to compare with our first plot (wideband). The comparison is depicted in Fig. 5.

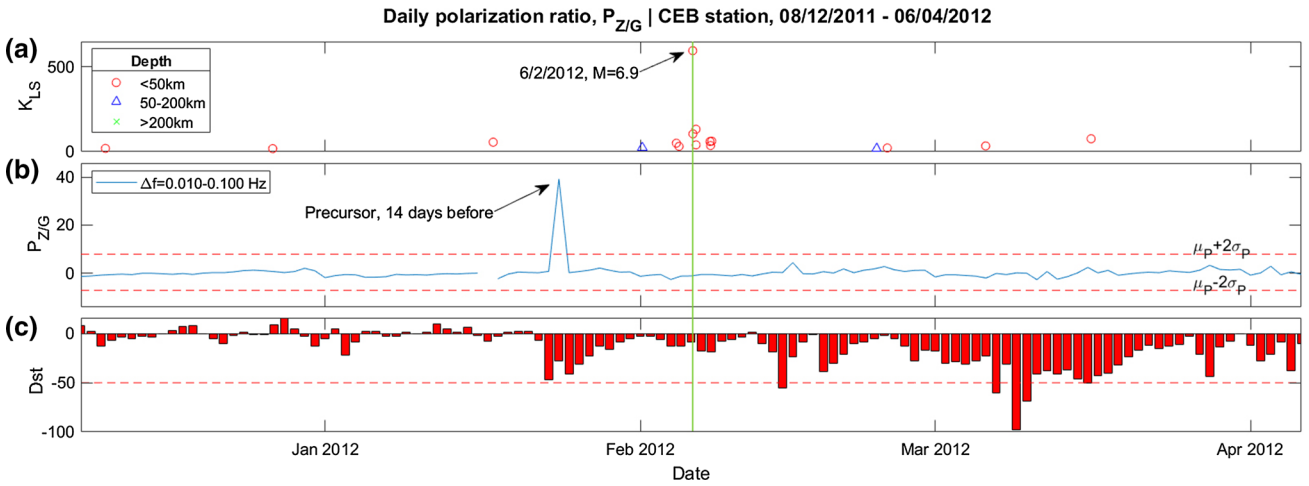


Fig. 3 Temporal evolution from 8 December 2011 to 6 April 2012 of **a** K_{LS} index for all earthquakes near CEB station with their depths specified using the symbol shapes, **b** daily polarization ratio, $P_{Z/G}$ and the $\mu_P \pm 2\sigma_P$ value in dashed red lines for $\Delta f = 0.01$–0.1 Hz, and **c** Dst index with its threshold considered active in dashed lines. The green vertical line marks the M6.9 earthquake event

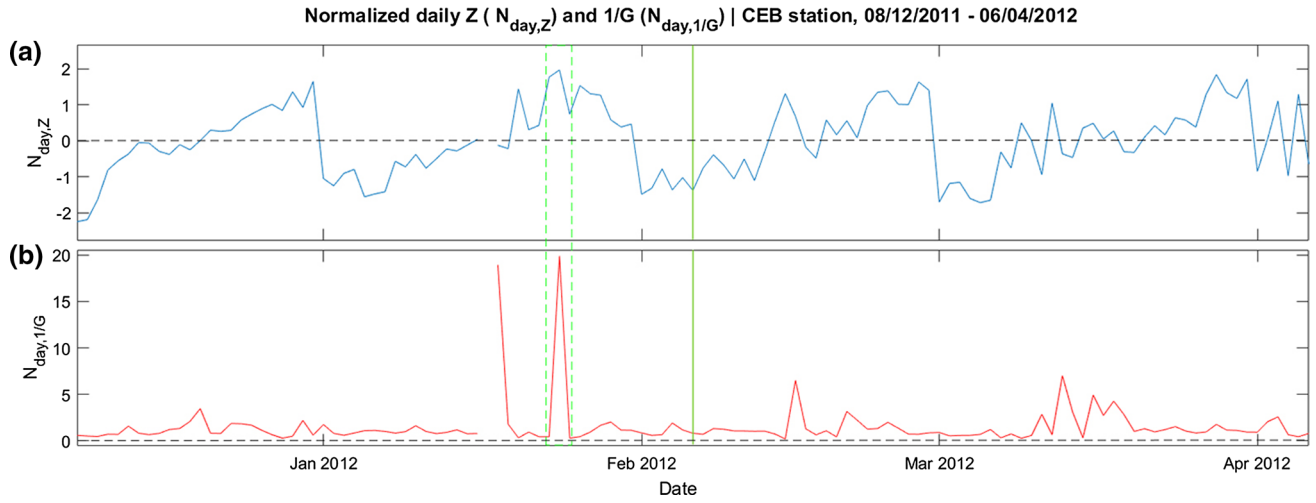


Fig. 4 Temporal evolution of **a** $N_{\mathrm{day,Z}}$ and **b** $N_{\mathrm{day,1/G}}$. The dashed green box highlights the period when the peak in Fig. 3b appeared. The green vertical line marks the M6.9 earthquake event

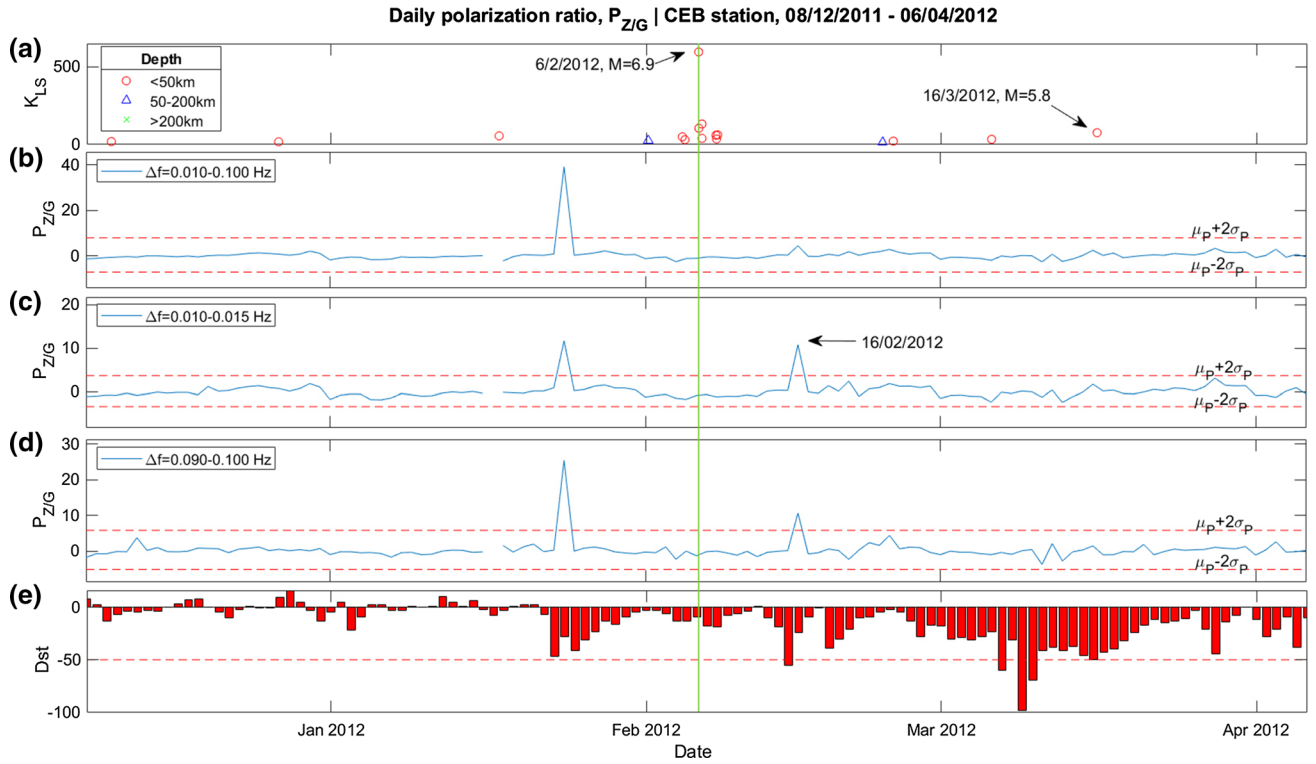


Fig. 5 Temporal evolution of **a** K_{LS} index for all earthquakes near CEB station, **b–d** daily polarization ratio, $P_{Z/G}$ for $\Delta f = 0.01$–0.1 Hz, $\Delta f = 0.010$–0.015 Hz and $\Delta f = 0.09$–0.10 Hz, respectively, with their respective $\mu_P \pm 2\sigma_P$ values in dashed red lines, and **e** Dst index

Based on Fig. 5, we can say that the wideband (Fig. 5b, similar to Fig. 3b) produced the most obvious peak before the main earthquake but at the same time suppressed an apparent second peak which occurred on 16 February 2012. We need to mention that there was another earthquake on 16 March 2012, exactly 1 month after the second peak. While this earthquake had a notably large magnitude of $M = 5.8$, its epicentral distance of 194.6 km did not convince us that it would produce a detectable precursor because it was beyond its R_{max} of 52 km.

This second peak had an amplitude comparable to the first one when the low narrowband (Fig. 5c) was used. Meanwhile, while it had a smaller amplitude, it still exceeded the threshold value when the high narrowband (Fig. 5d)

was analyzed. The inference from this observation can be explained by the skin depth effect relationship which states that the frequency of emission is inversely proportional to the square of depth (Prattes et al. 2008). Hence, it is implied that the choice of frequency range used has great influences on the depth of seismogenic emission sources which are going to be detected on the ground. By including a wideband, i.e., 0.01–0.10 Hz, we eliminate biased observation that prefers certain depth to provide us with thorough understanding of lithospheric activities underground. Thus, we can safely say that for this study, wideband is the most optimum to produce a reliable precursor which has an amplitude that is consistent with the earthquake magnitude and local seismicity index. From our finding, we also suggest that in order for an anomalous peak to be considered as a precursor, its appearance and amplitude should be persistent regardless of the frequency range.

Diurnal variation ratio

In this section, the results of the processed geomagnetic data using the diurnal variation ratio method are presented. Figure 6 illustrates CEB/LGZ diurnal variation ratios, R_i (blue lines) for each component ($i = H, D, Z$) and their respective moving averages (magenta lines) which are presented in Fig. 6b–d, with similar topmost (Fig. 6a) and bottommost (Fig. 6e) panels, as shown in Figs. 3 and 5. Moving average of R_H exhibited a drop on 16 January 2012, which is marked by an arrow. Meanwhile, moving average of R_D increased steadily until 21 February 2012 when it dropped, also marked by an arrow. Moving average of R_Z mostly showed a consistent trend along the time period. Despite the observed fluctuations, none of the moving averages exceeded the predefined $\mu_{R,i} \pm 2\sigma_{R,i}$ thresholds. Hence, they were not statistically significant and could not be considered a precursor. The results obtained from this method might be a result of the relatively low-magnitude main earthquake being studied, which is M6.9. As a comparison, Xu et al. (2013) who used the same method on the 2011 M9.0 Tohoku earthquake did observe a sharp peak in vertical (Z) diurnal variation about 2 months before the earthquake.

Conclusion

Two signal processing methods were performed on geomagnetic data obtained from the MAGDAS Cebu (CEB) and Legazpi (LGZ) stations 60 days before and 60 days after the

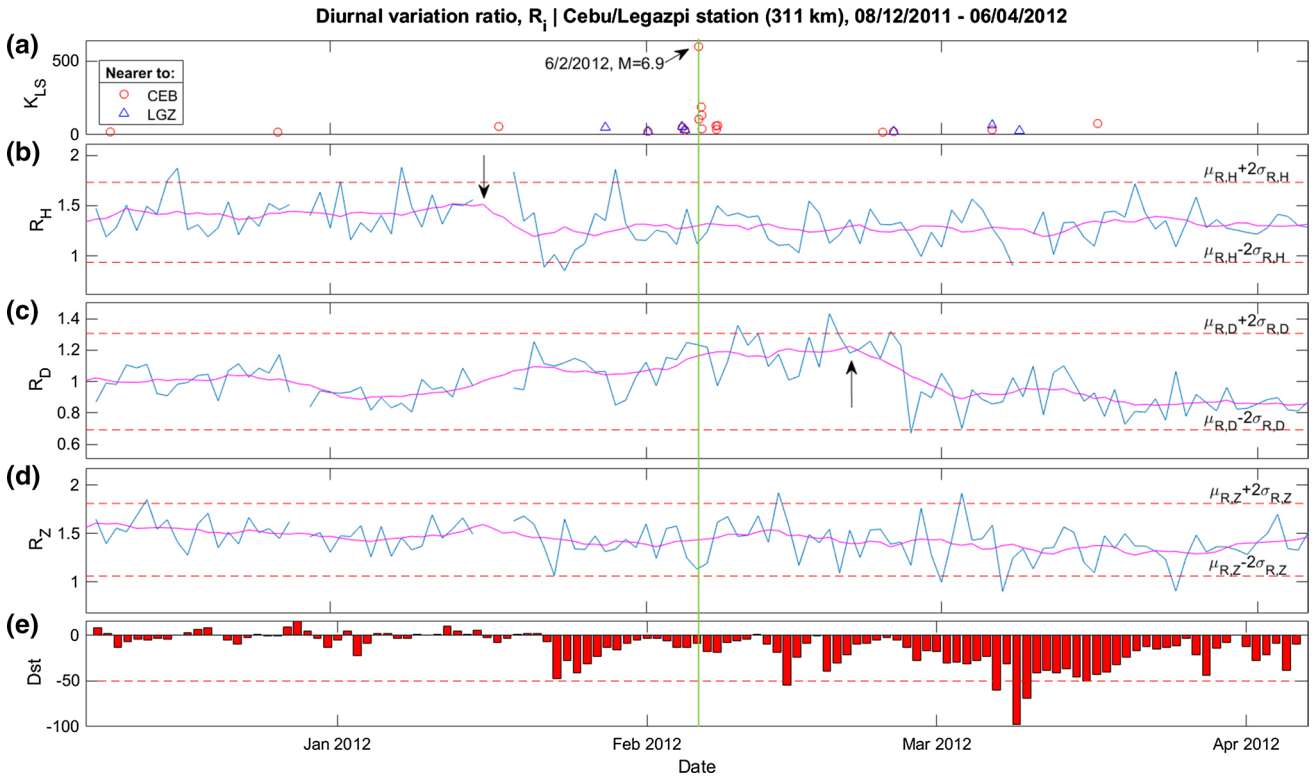


Fig. 6 Temporal evolution of **a** K_{LS} index for all earthquakes nearer to CEB and LGZ stations, indicated by circle and triangle symbols, respectively, **b–d** diurnal variation ratio of CEB/LGZ for all three components (R_H, R_D and R_Z) in blue lines, the corresponding 5-day moving average indicated in magenta lines and their respective $\mu_R \pm 2\sigma_R$ values in dashed red lines, and **e** Dst index. The green vertical line marks the M6.9 earthquake event

M6.9 Visayas, Philippines earthquake on 6 February 2012. Both methods were assessed and compared in term of effectiveness and reliability in detecting precursor of upcoming earthquakes.

The polarization ratio analysis method has successfully detected an anomaly preceding the main earthquake. The incoming earthquake was preceded by the geomagnetic anomaly 14 days before the event which would enable necessary mitigation actions to be taken. Besides, the anomaly appeared prior to the earthquake without other anomalous signature appearing in any other period even during the occurrences of geomagnetic storm. This indicates that the anomaly is reliable and uncorrelated with global geomagnetic activity. Furthermore, this method gives insight into the depth of the upcoming earthquakes by producing distinguished patterns from different frequency ranges. Since shallow earthquakes will typically cause greater destructions, forecasting the earthquake depth would enable the government agencies to determine the danger level they are encountering.

In contrast, the use of the diurnal variation ratio method in this study did not produce any unambiguous precursor prior to the main earthquake, even though fluctuations were observed in the *H* and *D* components. The relatively low-magnitude earthquake being studied might be the reason for the discrepancy between the results obtained from the two methods. Therefore, future studies which utilize multiple earthquakes with varying properties need to be conducted to confirm this discrepancy.

Acknowledgements The authors are thankful to International Center for Space Weather Science and Education, Kyushu University, for providing us with MAGDAS geomagnetic data.

Authors' contributions KAY and NSAH obtained and analyzed geomagnetic data. AY provided geomagnetic data for analysis. KAY prepared the manuscript, and all authors read and approved the final manuscript.

Funding The work of Akimasa Yoshikawa is supported by MEXT/JSPS KAKENHI Grant 15H05815.

Availability of data and materials Dst data were obtained from and always available at the GSFC/SPDF OMNIWeb interface at https://omniweb.gsfc.nasa.gov.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Aurelio MA, Dianala JDB, Taguibao KJL, Pastoriza LR, Reyes K, Sarande R, Lucero A (2017) Seismotectonics of the 6 February 2012 Mw 6.7 Negros earthquake, central Philippines. J Asian Earth Sci 142:93–108. https://doi.org/10.1016/j.jseaes.2016.12.018

Bello SA, Abdullah M, Hamid NSA, Yoshikawa A, Olawepo AO (2017) Variations of B0 and B1 with the solar quiet Sq-current system and comparison with IRI-2012 model at Ilorin. Adv Space Res 60:307–316. https://doi.org/10.1016/j.asr.2017.02.003

Chauhan V, Singh OP, Pandey U, Singh B, Arrora BR, Rawat G, Pathan BM, Sinha AK, Sharma AK, Patil AV (2012) A search for precursor of earthquakes from multi-station ULF observation and TEC measurements in India. Indian J Radio Space 41:543–556

Chavez O, Pérez-Enríquez R, Cruz-Abeyro JA, Millan-Almaraz JR, Kotsarenko A, Rojas E (2011) Detection of electromagnetic anomalies of three earthquakes in Mexico with an improved statistical method. Nat Hazards Earth Syst Sci 11:2021–2027. https://doi.org/10.5194/nhess-11-2021-2011

Chen C-H, Liu J-Y, Lin P-Y, Yen H-Y, Hattori K, Liang W-T, Chen Y-I, Yeh Y-H, Zeng X (2010) Pre-seismic geomagnetic anomaly and earthquake location. Tectonophysics 489:240–247. https://doi.org/10.1016/j.tecto.2010.04.018

Currie JL, Waters CL (2014) On the use of geomagnetic indices and ULF waves for earthquake precursor signatures. J Geophys Res Space Phys 119:992–1003. https://doi.org/10.1002/2013JA019530

Dudkin F, de Santis A, Korepanov V (2003) Active EM sounding for early warning of earthquakes and volcanic eruptions. Phys Earth Planet Inter 139:187–195. https://doi.org/10.1016/S0031-9201(03)00157-2

Dudkin F, Rawat G, Arora BR, Korepanov V, Leontyeva O, Sharma AK (2010) Application of polarization ellipse technique for analysis of ULF magnetic fields from two distant stations in Koyna-Warna seismoactive region, West India. Nat Hazards Earth Syst Sci 10:1513–1522. https://doi.org/10.5194/nhess-10-1513-2010

Febriani F, Han P, Yoshino C, Hattori K, Nurdiyanto B, Effendi N, Maulana I, Gaffar E (2014) Ultra low frequency (ULF) electromagnetic anomalies associated with large earthquakes in Java Island, Indonesia by using wavelet transform and detrended fluctuation analysis. Nat Hazards Earth Syst Sci 14:789–798. https://doi.org/10.5194/nhess-14-789-2014

Fedorov E, Pilipenko V, Uyeda S (2001) Electric and magnetic fields generated by electrokinetic processes in a conductive crust. Phys Chem Earth Part C 26:793–799. https://doi.org/10.1016/S1464-1917(01)95027-5

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

Hamid NSA, Gopir G, Ismail M, Misran N, Hasbi AM, Usang MD, Yumoto K (2009) The Hurst exponents of the geomagnetic horizontal component during quiet and active periods. In: International conference on space science and communication 2009. IEEE, pp 186–190

Hasbi AM, Mohd Ali MA, Misran N (2011) Ionospheric variations before some large earthquakes over Sumatra. Nat Hazards Earth Syst Sci 11:597–611. https://doi.org/10.5194/nhess-11-597-2011

Hayakawa M (2015) Earthquake prediction with radio techniques. Wiley, Singapore

Hayakawa M, Kawate R, Molchanov OA, Yumoto K (1996) Results of ultra-low-frequency magnetic field measurements during the Guam earthquake of 8 August 1993. Geophys Res Lett 23:241–244. https://doi.org/10.1029/95GL02863

Hayakawa M, Hattori K, Ohta K (2007) Monitoring of ULF geomagnetic variations associated with earthquakes. Sensors 7:1108–1122. https://doi.org/10.3390/s7071108

Ida Y, Yang D, Li Q, Sun H, Hayakawa M (2008) Detection of ULF electromagnetic emissions as a precursor to an earthquake in China with an improved polarization analysis. Nat Hazards Earth Syst Sci 8:775–777. https://doi.org/10.5194/nhess-8-775-2008

Molchanov OA, Hayakawa M (1998) On the generation mechanism of ULF seismogenic electromagnetic emissions. Phys Earth Planet Inter 105:201–210. https://doi.org/10.1016/S0031-9201(97)00091-5

Molchanov OA, Hayakawa M (2008) Seismo electromagnetics and related phenomena: history and latest results. Terrapub, Tokyo

Potirakis SM, Hayakawa M, Schekotov A (2017) Fractal analysis of the ground-recorded ULF magnetic fields prior to the 11 March 2011 Tohoku earthquake ($M_w = 9$): discriminating possible earthquake precursors from space-sourced disturbances. Nat Hazards 85:59–86. https://doi.org/10.1007/s11069-016-2558-8

Prattes G, Schwingenschuh K, Eichelberger HU, Magnes W, Boudjada M, Stachel M, Vellante M, Wesztergom V, Nenovski P (2008) Multi-point ground-based ULF magnetic field observations in Europe during seismic active periods in 2004 and 2005. Nat Hazards Earth Syst Sci 8:501–507. https://doi.org/10.5194/nhess-8-501-2008

Prattes G, Schwingenschuh K, Eichelberger HU, Magnes W, Boudjada M, Stachel M, Vellante M, Villante U, Wesztergom V, Nenovski P (2011) Ultra Low Frequency (ULF) European multi station magnetic field analysis before and during the 2009 earthquake at L'Aquila regarding regional geotechnical information. Nat Hazards Earth Syst Sci 11:1959–1968. https://doi.org/10.5194/nhess-11-1959-2011

Schekotov A, Fedorov E, Hobara Y, Hayakawa M (2013) ULF magnetic field depression as a possible precursor to the 2011/3.11 Japan earthquake. J Atmos Elect 33:41–51. https://doi.org/10.1541/jae.33.41

Sorokin VM, Pokhotelov OA (2010) Generation of ULF geomagnetic pulsations during early stage of earthquake preparation. J Atmos Solar Terr Phys 72:763–766. https://doi.org/10.1016/j.jastp.2010.03.021

Stanica DA, Stanica D, Błęcki J, Ernst T, Jóźwiak W, Słomiński J (2018) Pre-seismic geomagnetic and ionosphere signatures related to the M_w5.7 earthquake occurred in Vrancea zone on September 24, 2016. Acta Geophys 66:167–177. https://doi.org/10.1007/s11600-018-0115-4

Stranneby D, Walker W (2004) Digital signal processing and applications, 2nd edn. Newnes, Oxford

Xu G, Han P, Huang Q, Hattori K, Febriani F, Yamaguchi H (2013) Anomalous behaviors of geomagnetic diurnal variations prior to the 2011 off the Pacific coast of Tohoku earthquake (M_w9.0). J Asian Earth Sci 77:59–65. https://doi.org/10.1016/j.jseaes.2013.08.011

Yusof KA, Abdullah M, Hamid NSA, Ahadi S (2019) On effective ULF frequency ranges for geomagnetic earthquake precursor. J Phys Conf Ser 1152:12033. https://doi.org/10.1088/1742-6596/1152/1/012033

Affiliations

Khairul Adib Yusof[1,2] · Nurul Shazana Abdul Hamid[1,3] · Mardina Abdullah[1,4] · Suaidi Ahadi[5] · Akimasa Yoshikawa[6]

Khairul Adib Yusof
mkhairuladibmyusof@gmail.com

Mardina Abdullah
mardina@ukm.edu.my

Suaidi Ahadi
suaidi.ahadi@gmail.com

Akimasa Yoshikawa
yoshi@geo.kyushu-u.ac.jp

1 Space Science Centre, Institute of Climate Change, Universiti Kebangsaan Malaysia, Selangor, Malaysia

2 Department of Physics, Faculty of Science, Universiti Putra Malaysia, Selangor, Malaysia

3 School of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia

4 Centre of Advanced Electronic and Communication Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Selangor, Malaysia

5 Department of Geophysics, Indonesian Agency for Meteorology, Climatology and Geophysics, Jakarta, Indonesia

6 International Center for Space Weather Science and Education, Kyushu University, Fukuoka 812-8581, Japan

Acta Geophysica (2019) 67:1307–1318
https://doi.org/10.1007/s11600-019-00321-2

RESEARCH ARTICLE - APPLIED GEOPHYSICS

Study on forward and inversion modeling of array laterolog logging in a horizontal/highly deviated well

Peng Zhu[1] · Zhiqiang Li[2] · Ming Chen[3] · Yixin Dong[4]

Received: 2 November 2018 / Accepted: 27 June 2019 / Published online: 10 July 2019

Abstract
Electric field synthesis was carried out using the multi-field superposition method according to the working principle of the array laterolog electrode system. The field distribution of each subfield was simulated with the 3D finite element method, and the laterolog response of the array was obtained using the linear superposition principle of electric field. The detection depth and thin layer response at different angles of the array laterolog were analyzed. The forward response calculation shows that the radial detection depth of the array laterolog is smaller than the deep laterolog detection depth. When the inclination angle of the well is less than 15°, the logging response of the array laterolog is less affected by the well inclination, and the well inclination correction need not be performed. The logging response values of highly deviated wells with inclination angles exceeding 60° and horizontal wells are quite different from those of vertical wells; thus, well deviation correction must be performed. To improve the stability of array laterolog logging inversion using the accurate forward response, a Newton–singular value decomposition method based on particle swarm optimization is proposed to realize inversion of array laterolog logging, and the stability and reliability of logging inversion are greatly improved. Thus, application of the theoretical model and actual data processing and analysis show that the proposed method can effectively and accurately eliminate the influence of a complex logging environment and obtain real formation parameters.

Keywords Newton–SVD method · Particle swarm · Finite element method · Inversion · Array laterolog

Introduction

Horizontal/highly deviated well technology is one of the most important technologies for oil and gas exploration and development, and has thus become a major strategic measure for oil companies. The formation of horizontal/highly deviated wells is characterized by a lack of rotational axis symmetry, and thus, it is not possible to simplify the meridional plane problem. The original mixing method and conventional analytical methods are no longer suitable for 3D strata. The 3D finite element method, however, can effectively obtain the array laterolog logging response of the array in 3D strata (Liu et al. 1994, 1997; Li et al. 1996, Li 1998; Davydycheva et al. 2003). Logging response characteristics of the array laterolog differ considerably between horizontal/highly deviated wells and vertical wells. Therefore, numerical simulation of the array laterolog can help recognize the detection characteristics of the instrument and provide a theoretical basis for its research and development. Moreover, it can help analysts improve their understanding of logging curves. This paper uses the 3D finite element method to calculate the logging response of array laterologs and analyzes their logging response characteristics for horizontal/highly deviated wells.

The inversion technique is generally used to eliminate the influence of the environment on the logging response of array laterologs (Zhu et al. 2015). For electrical logging inversions, the formation parameters are generally obtained in terms of resistivity (e.g., invasion radius, invasion zone resistivity, and true formation resistivity). Charts were used to perform inversions in the early days of logging

✉ Peng Zhu
zhupeng@cdut.edu.cn

[1] College of Energy Resources, Chengdu University of Technology, Chengdu 610059, People's Republic of China

[2] China Radio Wave Research Institute, Xinxiang 453003, Henan, People's Republic of China

[3] CNOOC (China) Co., Ltd. Zhanjiang Branch, Zhanjiang 524000, Guangdong, People's Republic of China

[4] Institute of Sedimentary Geology, Chengdu University of Technology, Chengdu 610059, People's Republic of China

development. They caused the following inconveniences. First, only a limited number of parameters could be considered. Second, because the charts were made using the simplified model, they could only be processed step-by-step in terms of a single factor, and the accuracy of the result was doubtful. With the development of computer technology, it became possible to carry out large-scale calculations. Currently, the main method employed for inversion of array laterolog logging linearizes the nonlinear problem using the Taylor expansion model, and then solves it using singular value decomposition, the generalized inverse method, and so on. However, due to the local optimum, the inversion would easily achieve a local solution, and thus, the global optimal solution cannot be obtained accurately.

In recent years, people have begun to use intelligent methods to conduct logging inversion. Datta-Gupta et al. (1995) used the simulated annealing algorithm to determine the random infiltration region of porous media based on the inversion model. Ding et al. (2002) proposed using neural networks to invert dual laterolog resistivity. Tan and wang (2005) used genetic algorithms to calculate geophysical inverse problems, whereas Chen et al. (2007) applied the particle swarm method to study the inversion of dual laterolog resistivity.

Particle swarm optimization (PSO) (Chauhan et al. 2013; Zhang et al. 2015) is an optimization technique based on community; it is a kind of community intelligent optimization algorithm. Particle swarms correct individual action strategies through information sharing between community and individual experiences and finally find the optimal solution of the inversion. The initial stage of PSO has good convergence, but the algorithm begins to show inertness in the latter stage of the calculation, leading to slow convergence and poor precision. The Newton algorithm is characterized by fast convergence and high calculation accuracy, but it is greatly affected by the initial value. If the initial value is not well selected, the convergence of the algorithm is negatively affected. This paper proposes a Newton–singular value decomposition (SVD) method based on PSO to overcome two problems: slow convergence in the later period of calculation in the PSO method and the fact that the Newton method is strongly affected by the initial value.

Thus, we intend to apply the Newton–SVD method to realize inversion of array laterolog logging and improve its stability using the accurate forward response.

Tool and methods

Array laterolog electrode structure

Schlumberger introduced a high-resolution new array laterolog logging tool in 1998, with a main current transmitting electrode A_0 in the middle, six pairs of symmetrical electrodes on each side (A_1, A_2, A_3, A_4, A_5, and A_6), and two pairs of supervising electrodes (M_1 and M_2) between the electrodes A_0 and A_1. Consider that A_0 emits a current and the other shield electrodes are loop electrodes. This gives rise to R_{A0}. One pair of shield electrodes is added to carry the current from A_0 to both sides, and R_{A1}, R_{A2}, R_{A3}, R_{A4}, and R_{A5} are obtained with different detection depths. The resistivity curves of six different detection depths were obtained from the measurement of the array laterolog, and the shallowest detection mainly reflected the influence of the resistivity of the mud cake (Fig. 1).

Forward model

Direct current logging uses low-frequency alternating current as the power supply. Due to its low frequency, the array laterolog logging response can be reduced to a direct current field calculation. The electric field of the array lateral logging can be represented by the differential equation seen below (Li 1980; Sibbit and Faivre 1985; Clavier 1991; Zhang 1996; Zhu et al. 2005; Li et al. 2010):

$$\nabla \cdot (\sigma \nabla U) = 0 \tag{1}$$

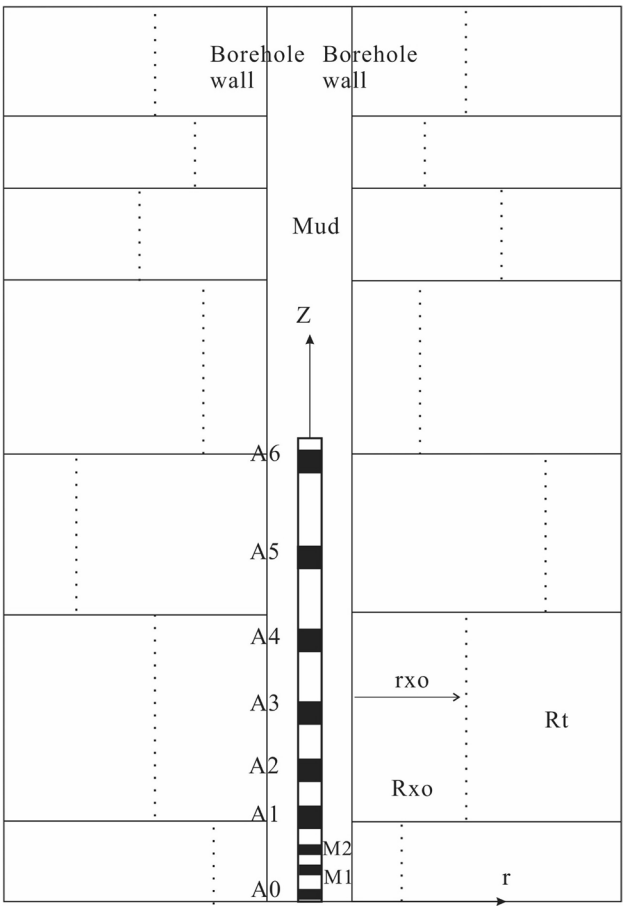


Fig. 1 Array laterolog tool and formation model

where σ is the conductivity of the formation and U is the measured potential.

In the cylindrical coordinate system (r, φ, z), it can be expressed as

$$\frac{\partial}{\partial r}\left(\sigma r\frac{\partial U}{\partial r}\right)+\frac{\partial}{\partial \varphi}\left(\frac{\sigma}{r}\frac{\partial U}{\partial \varphi}\right)+r\frac{\partial}{\partial z}\left(\sigma\frac{\partial U}{\partial z}\right)=0 \tag{2}$$

This equation is applicable under the following boundary conditions.

Dirichlet boundary condition: U is constant on the electrode surfaces, U is known on the current electrodes, but it is unknown on the voltage electrodes. On the infinity boundary, $U=0$.

Neumann boundary condition: For the constant current electrode, $\iint_D \sigma_m \frac{\partial U}{\partial n} ds = I_A$ (where I_A is the supply current of the power supply electrode A, σ_m is the mud conductivity, D is the surface of the electrode, and S is the integral variable). On the insulating boundary surface, $\frac{\partial U}{\partial n} = 0$.

In order to obtain Eq. (2), the problem of the partial differential equation needs to be transformed into the extreme problem of the functional:

$$\begin{aligned}&\boldsymbol{\Phi}=\frac{1}{2}\int_{\Omega}\sigma\left[\left(\frac{\partial U}{\partial r}\right)^2+\frac{1}{r^2}\left(\frac{\partial U}{\partial \varphi}\right)^2+\left(\frac{\partial U}{\partial z}\right)^2\right]rdrdzd\\&\varphi-\sum_E I_E U_E=\boldsymbol{\Phi}_1-\boldsymbol{\Phi}_2\end{aligned} \tag{3}$$

$$\boldsymbol{\Phi}_1=\frac{1}{2}\int_{\Omega}\sigma\left[\left(\frac{\partial U}{\partial r}\right)^2+\frac{1}{r^2}\left(\frac{\partial U}{\partial \varphi}\right)^2+\left(\frac{\partial U}{\partial z}\right)^2\right]rdrdzd\varphi \tag{4}$$

$$\boldsymbol{\Phi}_2=\sum_E I_E U_E \tag{5}$$

where I_E and U_E, respectively are the current and potential of each electrode, E is the boundary of the electrode, and Ω is the boundary of the formation.

After the potential is obtained using the finite element method, the lateral log response of the array satisfies the following conditions:

$$R_a = K\frac{U_{M1}}{I_{A0}} \tag{6}$$

where K is the instrument electrode system constant, U_{M1} is the potential for electrode M_1, I_{A0} is the emission current intensity of electrode A_0, and R_a is the apparent resistivity under different conditions.

When solving the potential distribution of the array laterolog, the principle of electric field superposition is used to transform it into the potential function of multiple corresponding fields (Zhu et al. 2015).

In the first subfield, only electrode A_0 emits a unit current $I=1$ A, and the other electrodes do not emit any current. From the second subfield onwards, only electrodes A_{n-1} and $A_{n-1'}$ emit unit current $I=1$ A in the nth subfield, and the other electrodes do not emit any current. Electrodes A_{n-1} and $A_{n-1'}$ are equipotential $(2 \le n \le 7)$. We calculate the first to seventh field potentials in turn using $U_0(x, y, z)$, $U_1(x, y, z)$, $U_2(x, y, z)$, $U_3(x, y, z)$, $U_4(x, y, z)$, $U_5(x, y, z)$, and $U_6(x, y, z)$. Then, the lateral potential of the array is

$$U = U_0 + C_1U_1 + C_2U_2 + C_3U_3 + C_4U_4 + C_5U_5 + C_6U_6 \tag{7}$$

where C_1, C_2, C_3, C_4, C_5, and C_6 are the focus synthesis coefficients.

R_{A1} is taken as an example to discuss the constraints of the working mode of the array laterolog. Its constraint condition is

$$U_{M1} = U_{M2} \tag{8}$$

$$U_{A2} = U_{A3} = U_{A4} = U_{A5} = U_{A6} \tag{9}$$

$$I_{A0} + I_{A1} + I_{A2} + I_{A3} + I_{A4} + I_{A5} + I_{A6} = 0 \tag{10}$$

where U_{M1} and U_{M2} are the respective potential values of electrodes M_1 and M_2; U_{A2},U_{A3},U_{A4}, U_{A5} and U_{A6} are the respective potential values of electrodes A_2, A_3, A_4, A_5, and A_6; I_{A0}, I_{A1}, I_{A2}, I_{A3}, I_{A4}, I_{A5}, and I_{A6} are the respective current values of electrodes A_1, A_2, A_3, A_4, A_5, and A_6.

Using the constraints noted in Eqs. (8) to (10), the focus synthesis coefficients of each subfield can be obtained.

$$\begin{aligned}&U_{0M1}+C_1U_{1M1}+C_2U_{2M1}+C_3U_{3M1}+C_4U_{4M1}\\&+C_5U_{5M1}+C_6U_{6M1}=U_{0M2}+C_1U_{1M2}+C_2U_{2M2}\\&+C_3U_{3M2}+C_4U_{4M2}+C_5U_{5M2}+C_6U_{6M2}\end{aligned} \tag{11}$$

$$\begin{aligned}&U_{0A2}+C_1U_{1A2}+C_2U_{2A2}+C_3U_{3A2}\\&+C_4U_{4A2}+C_5U_{5A2}+C_6U_{6A2}=U_{0A3}+C_1U_{1A3}\\&+C_2U_{2A3}+C_3U_{3A3}+C_4U_{4A3}+C_5U_{5A3}+C_6U_{6A3}\end{aligned} \tag{12}$$

$$\begin{aligned}&U_{0A2}+C_1U_{1A2}+C_2U_{2A2}+C_3U_{3A2}\\&+C_4U_{4A2}+C_5U_{5A2}+C_6U_{6A2}=U_{0A4}+C_1U_{1A4}\\&+C_2U_{2A4}+C_3U_{3A4}+C_4U_{4A4}+C_5U_{5A4}+C_6U_{6A4}\end{aligned} \tag{13}$$

$$\begin{aligned}&U_{0A2}+C_1U_{1A2}+C_2U_{2A2}+C_3U_{3A2}\\&+C_4U_{4A2}+C_5U_{5A2}+C_6U_{6A2}=U_{0A5}+C_1U_{1A5}\\&+C_2U_{2A5}+C_3U_{3A5}+C_4U_{4A5}+C_5U_{5A5}+C_6U_{6A5}\end{aligned} \tag{14}$$

$$\begin{aligned}&U_{0A2}+C_1U_{1A2}+C_2U_{2A2}+C_3U_{3A2}\\&+C_4U_{4A2}+C_5U_{5A2}+C_6U_{6A2}=U_{0A6}+C_1U_{1A6}\\&+C_2U_{2A6}+C_3U_{3A6}+C_4U_{4A6}+C_5U_{5A6}+C_6U_{6A6}\end{aligned} \tag{15}$$

$$1 + C_1 + C_2 + C_3 + C_4 + C_5 + C_6 = 0 \tag{16}$$

$$U_{M1} = U_{0M1} + C_1U_{1M1} + C_2U_{2M1} + C_3U_{3M1} + C_4U_{4M1} + C_5U_{5M1} + C_6U_{6M1} \tag{17}$$

where U_{iM1}, U_{iM2}, U_{iA2}, U_{iA3}, U_{iA4}, U_{iA5}, and U_{iA6} are the potential values of the corresponding electrodes under the i-1-th subfield.

Substituting Eq. (6) into Eq. (11) provides the value of the apparent resistivity R_{A1}. The other resistivity values of R_{A2}, R_{A3}, R_{A4}, and R_{A5} can be obtained by different focusing methods.

In this paper, the entire section is symmetrical along the lateral direction, only need to calculate 0°–180° space, meshing with 15° in the circumferential direction, and non-uniform meshing in the longitudinal direction and radial direction to reduce the number of grid. For the smallest unit, the representation of the tetrahedron is used, and finally m elements and n nodes are formed. After the assembly of the unit, the overall conductance matrix is formed, and the obtained equations are in the form of

$$K\varphi = f \tag{18}$$

In the formula, $K = \sum_{e=1}^{m} K^e, \varphi = [\phi_1, \ldots, \phi_N]^{\mathrm{T}}, f = [I_1, \ldots, I_N]^{\mathrm{T}}$ where K^e is the coefficient matrix, φ is the voltage value on each element, and f is the current value at each node. Here, the array is assembled with the lateral electrode nodes. The front-line solution only needs to be iterated one step to obtain the voltage value corresponding to the electrode. This method is beneficial to the improvement of the forward calculation speed.

Inversion model

Considering the two-dimensional formation model of horizontal stratification, it is assumed that there are M layers, and each layer has three parameters: invasion zone resistivity, invasion radius, and original formation resistivity. They are labeled as R_{ti} (original formation resistivity), r_{xoi} (invasion radius), and R_{xoi} (invasion zone resistivity) ($i = 1, 2, \ldots, M$). Thus, M layers have 3 M parameters.

The resistivity curves in the log are recorded as $R_{Ai(k)}$ ($k = 1, 2, 3\ldots, n$; $i = 1, 2, 3, \ldots 5$), n is the number of points measured by a single curve, R_{ti}, r_{xoi}, and R_{xoi} ($i = 1, 2, \ldots, M$), and

$$\min \quad Q(s) = \sum_{i=1}^{5} [(y_k - f_k(b))^2] \tag{19}$$

$$y_k = (R_{Ak}(1), R_{Ak}(2), \ldots R_{Ak}(n))^{\mathrm{T}}$$

$$b = (r_{xo1}, R_{xo1}, R_{t1}, \ldots, r_{xoM}, R_{xoM}, R_{tM})^{\mathrm{T}}$$

where y refers to the observed data; b denotes the model parameters of the inversion (invasion radius, invasion zone resistivity and true formation resistivity), and f represents the array lateral log response value calculated by model parameter b (Avdeev 2005; Deng et al. 2005). Since f is a nonlinear function, let it be Taylor-expanded near b (0), ignoring the quadratic terms, and obtaining a linearized function.

$$f(b) = f(b^{(0)}) + \sum_{j=1}^{5M} \left(\frac{\partial f_i}{\partial b_j} \right) \Big| b^{(0)} \Delta b_j \tag{20}$$

and

$$\Delta b_j = b_j - b_j^{(0)}, \quad \Delta f_j = f_j - f_j^{(0)} \tag{21}$$

The following formula can be obtained by the least squares criterion:

$$\frac{\partial Q}{\partial b_j} = 0, \quad j = 1, 2, \ldots, 5M \tag{22}$$

After formulating (21) and (22), we get

$$\Delta f = A\Delta b \tag{23}$$

where A is the Jacobi matrix. The iterative format using the damped least squares method is

$$(AA^{\mathrm{T}} + \lambda I)\Delta b^{(k)} = -A^{\mathrm{T}}(y - f) \tag{24}$$

$$b^{(k+1)} = b^{(k)} + \Delta b^{(k)} \tag{25}$$

where λ is the damping factor. There are many ways to solve ill-posed linear equations such as (24). The following approaches are commonly used in logging inversion calculations: Newton's method, SVD method, generalized inverse method, and damped least squares method (Deng et al. 2015; Tan et al. 2011; Sewell et al. 2010; Smits et al. 1998). The error caused by linearizing the nonlinear problem itself is very large. In addition, the high dependence on the initial value in these methods greatly discounts the accuracy of the inversion result.

Array laterolog inversion based on particle swarm optimization and the Newton–SVD hybrid algorithm

Assume a particle group consisting of S (number of particles). The position of the ith particle in the particle group is represented by a 3 M-dimensional vector as $s_i = (r_{xo_1}^{s_i}, R_{xo_1}^{s_i}, R_{t_1}^{s_i}, \ldots, r_{xo_M}^{s_i}, R_{xo_M}^{s_i}, R_{t_M}^{s_i})^{\mathrm{T}}$, and the optimal position of the ith particle search so far is $p_i = (r_{xo_1}^{p_i}, R_{xo_1}^{p_i}, R_{t_1}^{p_i}, \ldots, r_{xo_M}^{p_i}, R_{xo_M}^{p_i}, R_{t_M}^{p_i})^{\mathrm{T}}$. The optimal position of the particle swarm search so far is $p_g = (r_{xo_1}^{p_g}, R_{xo_1}^{p_g}, R_{t_1}^{pg}, \ldots, r_{xo_M}^{p_g}, R_{xo_M}^{p_g}, R_{t_M}^{p_g})^{\mathrm{T}}$. Using superscripts to indicate the number of iterations, the two basic iterations of the particle swarm optimization algorithm are

$$v_i^{t+1} = \omega v_i^t + c_1 r_1^t (p_i^t - s_i^t) + c_2 r_2^t (p_g^t - s_i^t) \quad (26)$$

$$s_i^{t+1} = s_i^t + v_i^{t+1} \quad (27)$$

where $i = 1, 2, \ldots, S$; ω is the inertia weight; c_1 and c_2 are called the learning factors or the acceleration factors (both are non-negative constants); and r_1 and r_2 are random numbers between [0, 1]. When $\omega = 0$, the speed at which the particles fly depends only on their current positions, their best position in history, and the best position in the history of the particle group; there is no memory of the speed itself. Thus, the particles at the global best position will remain stationary and the other particles will tend to their own best positions and the weighted center of the global best position; that is, the particle group will shrink to the current global best position.

$$s_i^{t+1} = s_i^t + c_1 r_1^t (p_i^t - x_i^t) + c_2 r_2^t (p_g^t - s_i^t) \quad (28)$$

The evolutionary equation of Eq. (28) weakens the global search ability and strengthens the local ability. If the current point is the same as the local best or the global best, the point will stop evolution. However, in order to improve the original global search ability, maintain the original global best, and re-randomly generate the position of the particle in the new search space, other particles were optimized according to the original Eq. (27). In this way, new random particles will appear in the new generation of evolution, which can enhance the global search ability of the algorithm. This method is called a random particle swarm method.

First, the search should be conducted using the random particle swarm method (Biswas et al. 2007; Ziari and Jalilian 2012; Mohammadi et al. 2017; Abordán and Szabó 2018). When $Q(s)$ is less than the given minimum allowable value in Eq. (19), it is used as the initial point of the Newton–SVD approach, and then, the traditional Inversion method based on the Newton–SVD method is adopted. Using this technique to perform array lateral logging inversion not only helps find a good initial point but also facilitates fast inversion.

The following are the steps of the Newton–SVD algorithm based on PSO:

(1) Set particle size m, acceleration factors c_1 and c_2, weighting factor = 0, and search space dimension D, maximum iteration number TN, and minimum fitness value φ.
(2) Calculate the fitness value of the particles.
(3) Find the particle that satisfies the minimum fitness function as the initial value of the Newton–SVD method.
(4) Calculate the new formation parameters by combining the obtained formation parameters (of the particles) with the generalized inverse method.
(5) Calculate the logging response using the formation parameters, and compare the calculated logging response with the measured logging response using the 2-norm method. If the error is within the allowable range, the final formation parameter information has been obtained. Otherwise, repeat (4).

Results and discussion

Electric field distribution of homogeneous formation

The following simulation conditions were considered for this study comprised of uniform formation resistivity R_t of 1 Ω·m, and instrument radius of 0.045 m.

As per the simulation results (Fig. 2), as the number of shield electrodes increases, the potential value of the formation gradually increases at the same position, as does the detection depth of the instrument.

Pseudo-geometry factor calculation

According to the stratum model shown in Table 1, the well diameter dh = 8″ = 0.2032 m, and we need to calculate the pseudo-geometry factor of the lateral logging of the array and determine the depth of detection. The pseudo-geometry factor is expressed as

$$L = \left(R_t - R_a\right) / \left(R_t - R_{xo}\right) \quad (29)$$

The detection depth of the instrument is defined as 50% of the pseudo-geometry factor, as shown in Fig. 3. The detection depth of the five detection modes of array laterolog logging under high-invasion simulation conditions is as follows: $R_{A5} = 0.70$ m, $R_{A4} = 0.45$ m, $R_{A3} = 0.37$ m, $R_{A2} = 0.29$ m, and $R_{A1} = 0.23$ m.

The detection depth of the array laterolog is smaller than the depth of deep lateral logging of the Schlumberger Cyber Service Unit (CSU) (1.27 m); however, the advantage of array laterolog is to reflect the radial gradient of formation resistivity, which can be used for resistivity profile imaging in the radial direction.

Influence of well inclination

According to the stratigraphic model shown in Table 2, the array laterolog logging responses of the two stratum models under different well inclination conditions were simulated.

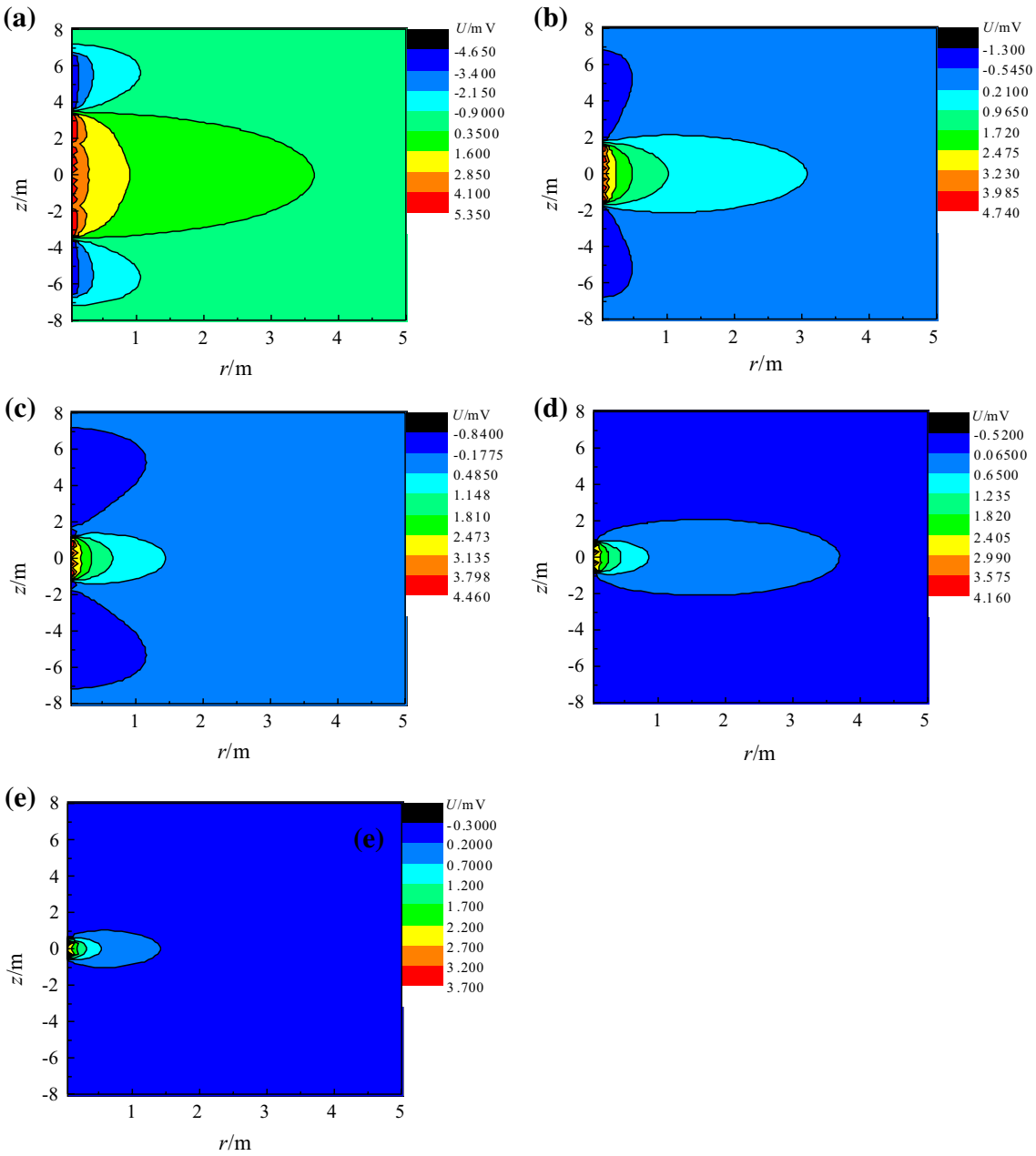


Fig. 2 Electric potential distributions of different measurements in a homogeneous formation ($R_t = 1$ Ω·m)

Table 1 Stratum model

Stratum types	R_m (Ω·m)	R_{xo} (Ω·m)	R_t (Ω·m)
Stratum 1 (low invasion)	0.1	10	1
Stratum 2 (high invasion)	0.1	10	50

R_m is the mud resistivity, R_{xo} is the resistivity of invasion zone, and R_t is the resistivity of the target layer

The considered inclinations of the well were 0°, 15°, 30°, 45°, 60°, 75°, and 90°, and the simulation results are shown in Figs. 4 and 5.

Figure 4 shows that for a formation with a thickness of 1 m, when the well inclination is 0° (that is, the well is vertical), the array laterolog has the advantage of high resolution (0.4 m). Thus, information about its formation information is well reflected, and the response of R_{A5} (with the deepest detection depth) is 18.1 Ω·m. As the depth of detection of the instrument decreases, instruments with shallow detection depth are slightly affected by the wellbore and surrounding rock, which causes the measured values of R_{A4}, R_{A3}, R_{A2}, and R_{A1} to decrease sequentially. When the inclination angle of the well is less than 15°, the influence of well inclination on

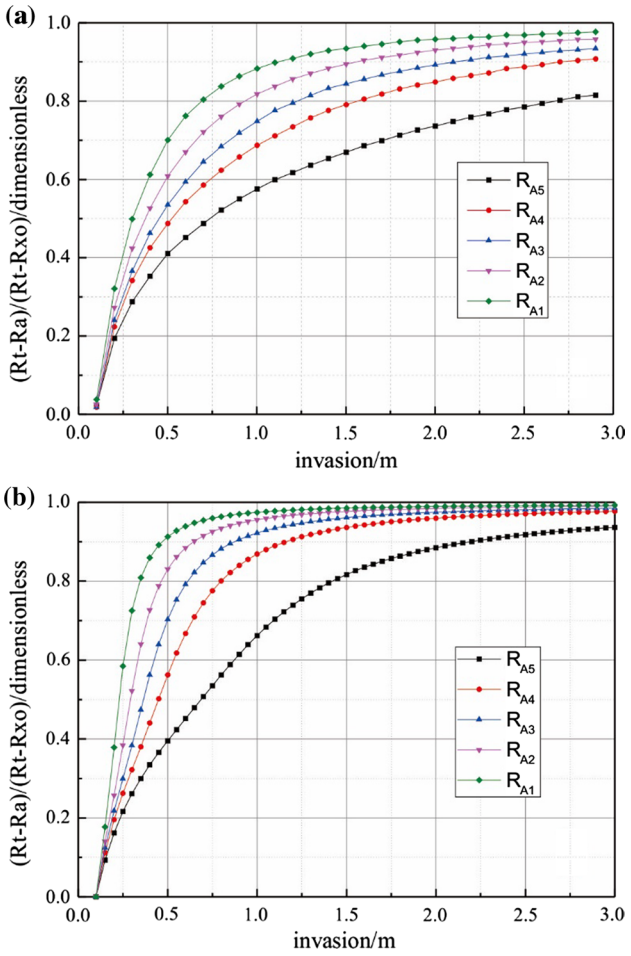


Fig. 3 Pseudo-geometry factor as a function of the invasion zone radius for **a** low-invasion (Stratum 1) and **b** high-invasion (Stratum 2) models at different operation modes of array laterolog

Table 2 Stratum model

Stratum types	R_t (Ω·m)	H (m)	R_b (Ω·m)	R_m (Ω·m)
Stratum 3 (low invasion)	20	1	3	1
Stratum 4 (high invasion)	3	1	20	1

H is the thickness of the target layer. R_b is the resistivity of the surrounding rock

the logging response of array laterolog can be neglected. As the inclination angle of the well increases, the resistivity of array laterolog at the midpoint of the formation decreases significantly. When the well inclination angle exceeds 60°, the logging response of the array laterolog drops to 70% of logging response in the vertical well (Fig. 4a), and using the resistivity information to calculate the water saturation with Archie's formula at this time will certainly result in a large error. The increase in the inclination angle therefore has the greatest influence on R_{A5} at the deepest detection depth, and the least influence on R_{A1} at the shallowest detection depth.

Figure 5 shows that for the low-resistivity mudstone layer, when the local layer thickness is 1 m and the well inclination is 0°, the response of R_{A5} with the deepest depth of detection is close to the resistivity of the target layer, while the measured values of R_{A4}, R_{A3}, R_{A2}, and R_{A1} are greatly affected by the surrounding rock with the decrease in the detection depth, and the response values increase sequentially. When the inclination angle of the well is less than 15°, the influence of well inclination on the logging response of the array laterolog can be neglected; as the inclination angle of the well increases, the resistivity of the array laterolog at the midpoint of the formation also increases gradually, and the logging response of array laterolog is greatly affected by surrounding rock. The increase in the inclination angle also has the greatest influence on R_{A5} at the deepest detection depth, and the least influence on R_{A1} at the shallowest detection depth.

Inversion case

1. Stratum initial model and array laterolog response

Construct the stratum model such that the borehole radius $r_h = 0.1016$ m, resistivity of the mud filtrate $R_m = 0.1\,\Omega \cdot \mathrm{m}$, resistivity of the surrounding rock $R_s = 5\,\Omega \cdot \mathrm{m}$, invasion radius $r_{xo} = 0.5$ m, resistivity of the invasion zone $R_{xo} = 8\,\Omega \cdot \mathrm{m}$, and true resistivity of the formation $R_t = 20\,\Omega \cdot \mathrm{m}$. The thicknesses of the inversion stratum were 0.5 m, 1.0 m, and 3.0 m. Figure 6 shows the array laterolog response produced by the model.

2. Inversion of three electric rock parameters from the array laterolog logs

The inversion of three parameters of the array laterolog was conducted to calculate the invasion radius (r_{xo}), invasion zone resistivity (R_{xo}), and true formation resistivity (R_t) of the model using the obtained array laterolog response. The true resistivity of the formation is an important parameter in logging interpretation, and the calculation accuracy directly affects the calculation of oil saturation. The inversion results are consistent with the stratum model for the thin layer affected by the surrounding rock, and the logging response is effectively corrected (Fig. 7).

Actual data processing

Figure 8 shows the inversion results using the particle swarm-based Newton–SVD method, which inverts the actual data of a well in an oil field. Natural gamma, natural potential, and array lateral curves were measured for this well. The stratum belongs to the oil-bearing stratum, and its permeability is poor. The oil–water layer cannot be judged

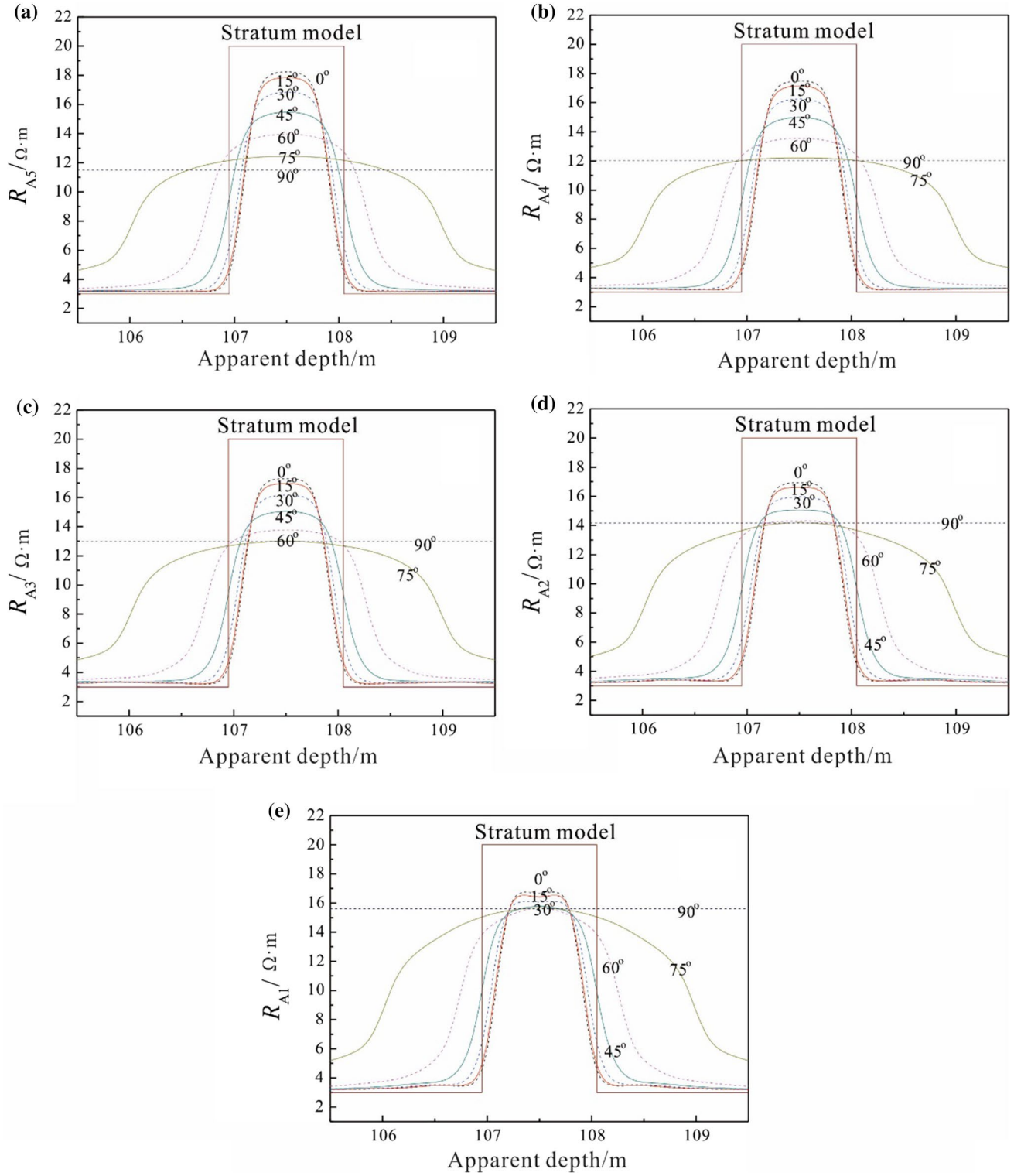


Fig. 4 Logging response of array laterologs for a 1-m-thick formation at different borehole inclinations (0° to 90°) and operation modes of array laterolog (Stratum 3)

intuitively by the conventional curve. The true resistivity and invasion depth of the stratum obtained by the hybrid algorithm can, however, effectively judge the oil–water layer.

At a depth of 1730–1768 m, the value of resistivity is approximately 3 Ω·m, and the response of the five array laterolog curves shows negative differences. After the inversion

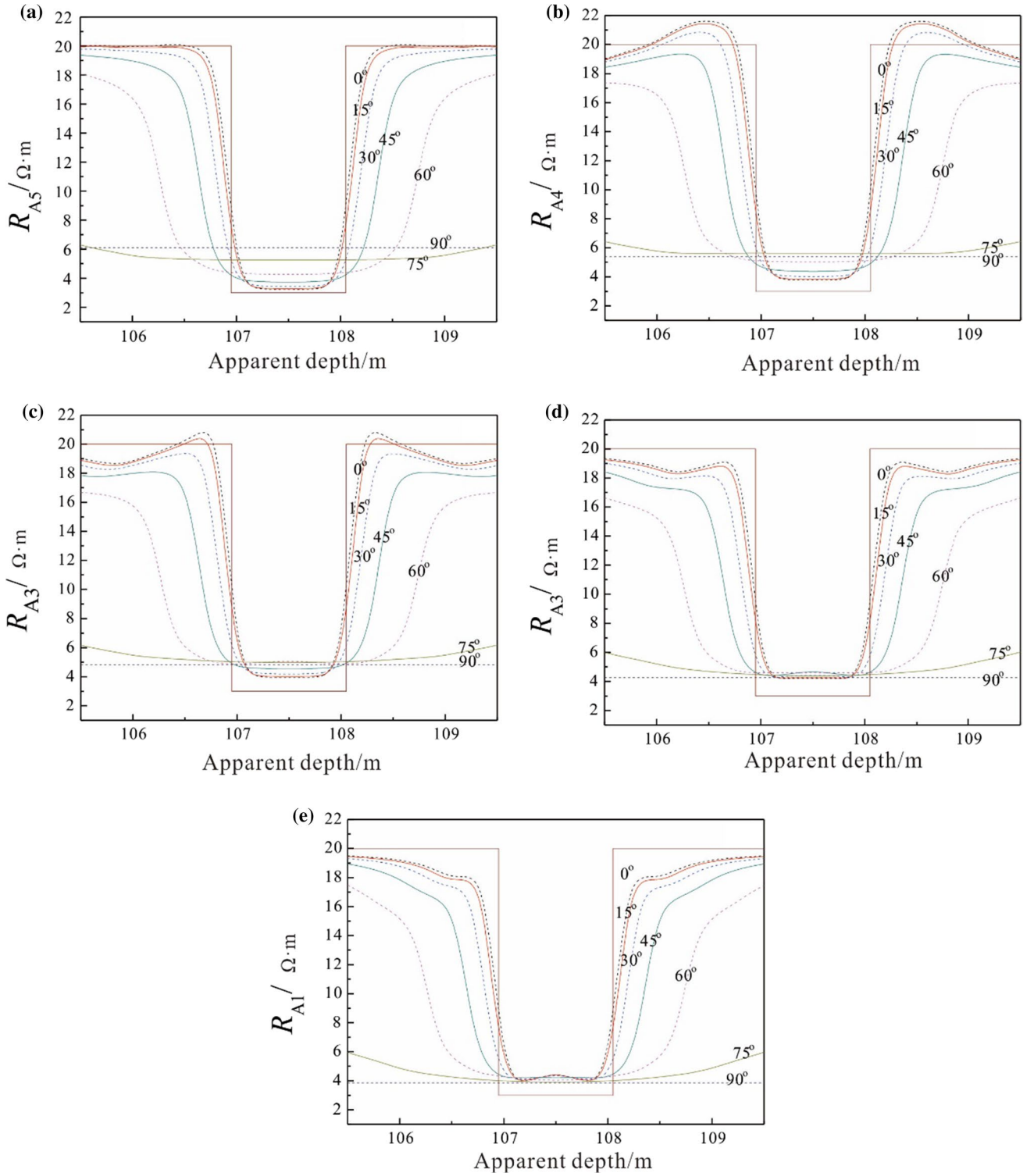


Fig. 5 Logging response of the array laterolog for a 1-m-thick formation at different borehole inclinations (0° to 90°) and operation modes of array laterolog (Stratum 4)

process, the layer presents obvious invasion properties, and R_{xo} is greater than R_t. Thus, this layer is comprehensively judged to be a water layer. At depths of 1641–1645 m, after the inversion process, the layer is characterized by invasion, and value of R_t is greater than that of R_{xo}. At depths of 1645–1670 m, after the inversion process, the layer

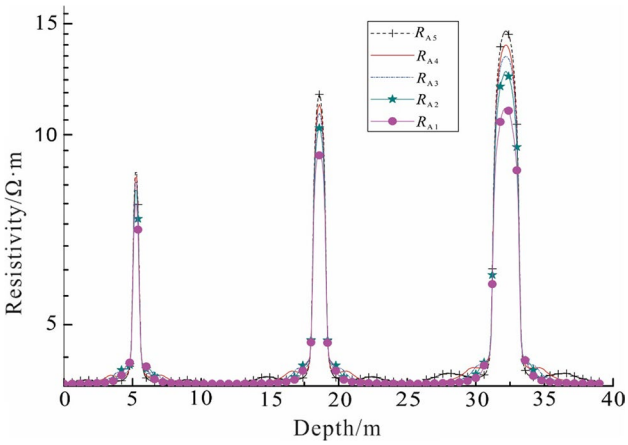


Fig. 6 Array laterolog response for the formation computed by forward modeling

continues to show certain invasion properties, and R_{xo} is greater than R_t. Thus, it is judged that the layer from 1641 to 1645 m is an oil-bearing water layer, while that from 1645 to 1670 m contains water. At depths of 1515–1518 m, the value of resistivity is 4 Ω·m, and the difference in the responses of the five array laterolog curves is not obvious. After the inversion treatment, the layer shows certain invasion characteristics, and R_t exceeds R_{xo}. Thus, the layer is comprehensively judged to be an oil layer of thickness of 3 m.

Conclusions

1. The 3D finite element method can be used to calculate the logging response of array laterologs in horizontal/highly deviated wells, and to analyze the influencing factors in 3D formation environments.
2. The detection depth of an array laterolog is smaller than that of a deep laterolog. However, the array laterolog can perform resistivity profile imaging in the radial direction, thereby reflecting the changing process of formation resistivity more intuitively.
3. The influence of well inclination on the logging response of the array laterolog is mainly reflected in the vicinity of the formation interface. As the inclination angle of the well increases, the influence of the surrounding rock increases gradually, and the influence of well inclination

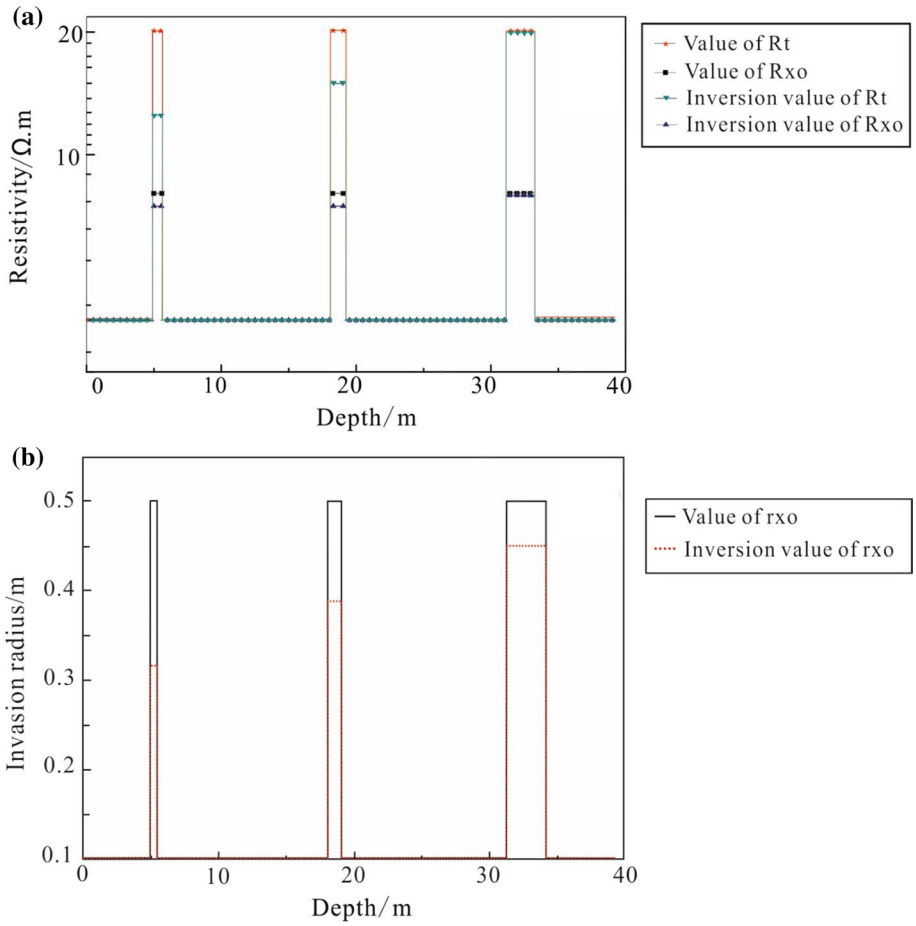


Fig. 7 Comparison of inversion results of formation resistivity (**a**) and comparison of inversion results of invasion radius (**b**)

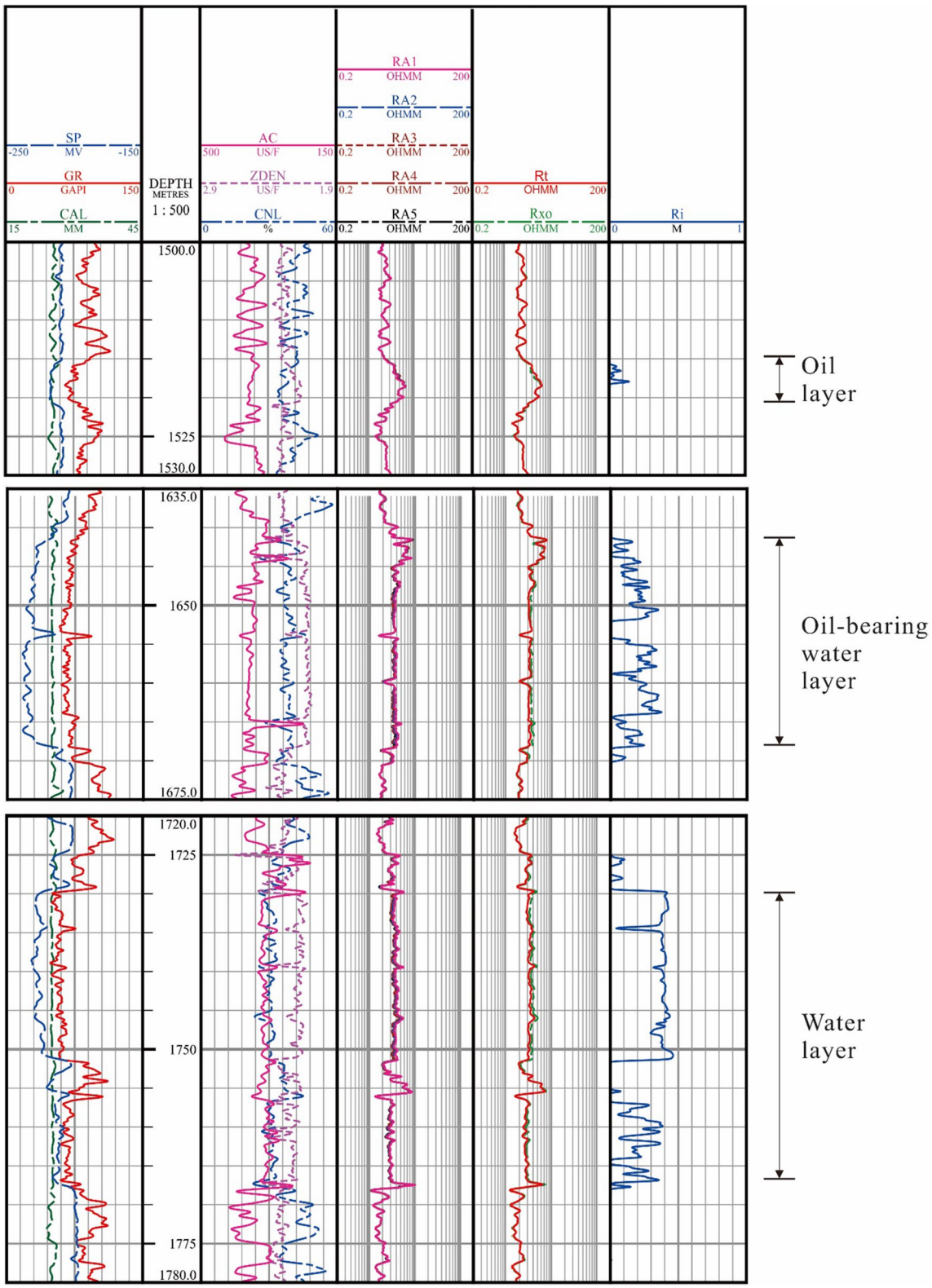


Fig. 8 Inversion results using the particle swarm-based Newton–SVD method for an actual oil field

on the logging curve at deep detection depth becomes greater than that at shallow detection depths.

4. When the well inclination is less than 15°, its influence can be neglected. However, the correction for well inclination must be performed for highly deviated wells. Otherwise, the accuracy of the water saturation calculation will be affected adversely.
5. This paper presents a Newton–SVD method based on PSO. The proposed optimization outperforms the conventional Newton–SVD method. It inherits the latter's advantages with regard to fast calculation speed and good convergence. It also overcomes two problems: slow convergence in the later period of calculation in the PSO method and the fact that the Newton method is strongly affected by the initial value. Moreover, the particle swarm-based Newton–SVD method does not require the mathematical form or derivative information of the objective function. It only needs an interval including the true parameters of the formation. Therefore, this method is suitable for large-scale, nonlinear, and multi-extreme complex problems. It is easy to program and suitable for parallel computing problems.
6. The calculations and comparison with actual data presented in this study show that the proposed method is feasible for oil exploration as it can effectively improve the stability of the logging inversion. It can eliminate the influence of complex logging environments and enhance the resolution and interpretation accuracy of oil and water layers in oil formations.

Acknowledgements This work is sponsored by Demonstration project of large carbonate gas field development in Sichuan Basin (No. 2016ZX05052).

References

Abordán A, Szabó NP (2018) Particle swarm optimization assisted factor analysis for shale volume estimation in groundwater formations. Geosci Eng 6(9):85–95

Avdeev DB (2005) Three-dimensional electromagnetic modelling and inversion from theory to application. Surv Geophys 26(6):767–799

Biswas A, Dasgupta S, Das S, Abraham A (2007) Synergy of PSO and bacterial foraging optimization—a comparative study on numerical benchmarks. Adv Soft Comput 44:255–263

Chauhan P, Deep K, Pant M (2013) Novel inertia weight strategies for particle swarm optimization. Memet Comput 5(3):229–251

Chen H, Deng S, Li Z, Fan Y (2007) The PSO (particle swarm optimization) method for dual-lateral log inversion. Geophys Geochem Explor 31(Supp):29–31

Clavier C (1991) The challenge of logging horizontal wells. Log Analyst 32(2):63–84

Datta-Gupta A, Lake LW, Pope GA (1995) Characterizing heterogeneous permeable media with spatial statistics and tracer data using sequential simulated annealing. Math Geol 27(6):763–787

Davydycheva S, Druskin V, Habashy T (2003) An efficient finite-difference scheme for electromagnetic logging in 3D anisotropic inhomogeneous media. Geophysics 68(5):1525–1536

Deng S, Li G, Fan Y, Zhang X, Dai S (2005) Preprocessing method of dual laterolog data in the low-resistivity reservoir. Well Logging Technol 29(3):220–222

Deng S, Li L, Li Z, He X, Fan Y (2015) Numerical simulation of high-resolution azimuthal resistivity laterolog response in fractured reservoirs. Pet Sci 12(2):252–263

Ding Z, Yang C, Tao H, Deng G, Chen G, Ma E (2002) Physical restriction on model selection of neural network in dual-lateral log inversion. Prog Geophys 17(2):331–336

Li D (1980) Finite element method in the application of the electrical logging. Petroleum industry press, Beijing, pp 72–129

Li S (1998) Inverting fracture porosity and dip of limestone fractured reservoir using 3-D FEM. Well Log Technol 22(6):412–415

Li S, Xiao C, Wang H, Zhang G (1996) Mathematical model of dual laterolog response to fracture and quantitative interpretation of fracture porosity. Acta Geophys Sinica 39(6):845–852

Li Z, Fan Y, Deng S, Ji X (2010) Inversion of array laterolog by improved difference evolution. J Jilin Univ Earth Sci Edit 40(5):1199–1204

Liu Z, Hu Q, Zhang J (1994) Calculating intrusion by the finite element method under the condition of dual laterolog tool response. Pet Instrum 8(3):149–152

Liu F, Li S, Zhang G (1997) Computation of borehole correction curves of dual laterolog tool by integral equation method. Acta Geophys Sinica 40(6):857–866

Mohammadi M, Montazeri M, Abasi S (2017) Bacterial graphical user interface oriented by particle swarm optimization strategy for optimization of multiple type DFACTS for power quality enhancement in distribution system. J Cent South Univ 24(3):569–588

Sewell P, Noroozi S, Vinney J, Amali R, Andrews S (2010) Improvements in the accuracy of an inverse problem engine's output for the prediction of below-knee prosthetic socket interfacial loads. Eng Appl Artif Intell 23(6):1000–1011

Sibbit AM, Faivre O (1985) The dual laterolog response in fractured rocks[C]. In: SPWLA 26th annual logging symposium, Dallas, Texas, 17–20

Smits J, Dubourg I, Luling M (1998) Improved resistivity interpretation utilizing a new array laterolog tool and associated inversion processing. SPE 49328:831–842

Tan Y, Wang J (2005) Calculate the geophysical inverse problems using genetic algorithm. Chin J Eng Math 22(3):427–434

Tan M, Gao J, Wang X, Zhang S (2011) Numerical simulation of the dual laterolog for carbonate cave reservoirs and response characteristics. Appl Geophys 8(1):79–85

Zhang G (1996) Electrical logging. University of Petroleum Press, Dongying

Zhang L, Tang Y, Hua C, Guan X (2015) A new particle swarm optimization algorithm with adaptive inertia weight based on Bayesian techniques. Appl Soft Comput 28:138–149

Zhu D, Deng S, Fan Y, Wang X (2005) Numerical simulation of dual laterolog response in highly deviated wells. Well Log Technol 29(3):208–211

Zhu P, Lin C, Li Z, Zhao W, Zhang H (2015) Numerical simulation of array laterolog response in horizontal and highly deviated wells. J Jilin Univ 45(6):1862–1869

Ziari I, Jalilian A (2012) Optimal placement and sizing of multiple aplcs using a modified discrete PSO. Int J Electr Power Energy Syst 43(1):630–639

Acta Geophysica (2019) 67:1319–1327
https://doi.org/10.1007/s11600-019-00339-6

RESEARCH ARTICLE - APPLIED GEOPHYSICS

On application of fractal magnetization in Curie depth estimation from magnetic anomalies

Chun-Feng Li[1,2] · **Duo Zhou**[1] · **Jian Wang**[3]

Received: 16 April 2019 / Accepted: 1 August 2019 / Published online: 10 August 2019

Abstract

As an independent geothermal proxy, the Curie-point depth has important geodynamic implications, but its estimation from magnetic anomalies requires an understanding of the spatial correlation of source magnetization, mathematically characterized by a fractal exponent. In this paper, we show that fractal exponent and Curie depth are so strongly inter-connected that attempts to simultaneous or iterative estimation of both of them often turn out to be futile. In cases of true large Curie depths, the iterative "de-fractal" method has a tendency of overcorrecting fractal exponents and thereby producing erroneously small Curie depths and smearing out true geological trends. While true fractal exponent can no way be constant over a large area, a regionally fixed fractal exponent is better than any mathematical treatments that are beyond the limit of data resolution and the underlying physics.

Keywords Curie depth · Geothermal structure · Heat flow · Fractal magnetization · Magnetic anomalies · Inversion · North America

Introduction

A wide variety of spectral techniques has been proposed to detect the depth to the bottom of the magnetic layer of the lithosphere from near-surface (or sometimes satellite) magnetic anomalies. This is also coined the Curie-point depth, where the temperature reaches the Curie point and rocks lose their ferromagnetism. Curie depths reflect deep thermal structure of the lithosphere assuming that the Curie temperature can be restricted to a narrow range (520–580 °C) for different mineralogy (Friedman et al. 2014). Curie temperatures decrease linearly with an increasing Ti content for natural terrestrial titanomagnetites (O'Reilly 1984), which are chemically stable at crust and upper mantle temperatures/pressures (Sauerzapf et al. 2008).

Among various techniques of detecting Curie depths, the linearized stepwise centroid method (Okubo et al. 1985; Tanaka et al. 1999), and its various extensions to fractal magnetization (Bansal et al. 2011; Li et al. 2009, 2010, 2013; Salem et al. 2014; Wang and Li 2015; Ravat et al. 2016), are theoretically simple but computationally stable. The reasoning behind this technique is simple; rather than seeking to directly estimate the Curie depth from nonlinear fitting between calculated and observed spectra of magnetic anomalies, Curie depth can be estimated indirectly from relatively easy linear inversion of the depths to the top and centroid of the magnetic layer, which are shallower than the magnetic bottom (Curie depth) and thereby more tractable computationally.

Curie depths are independent of Moho depths (or crustal thickness) and shallow radiogenic heat production, because the Moho is a lithological boundary, and radiogenic heat production decreases with depth (Turcotte and Schubert 2002), to have minimal effects at the Curie depth.

Examining amplitude (or power) spectra of magnetic anomalies offers by far the most valid and efficient means of estimating Curie depths over a large region. There are also geothermal methods based on the temperature-depth

✉ Chun-Feng Li
cfli@zju.edu.cn

[1] Department of Marine Sciences, Zhejiang University, Zhoushan 316021, China

[2] Laboratory for Marine Mineral Resources, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China

[3] Key Laboratory of Crustal Dynamics, Institute of Crustal Dynamics, China Earthquake Administration, Beijing 100085, China

relationship (Lachenbruch 1968; Negi et al. 1987). Curie depth can be correlated with surface heat flow (Li et al. 2010, 2017), and thereby is a good proxy to lithospheric thermal structure, particularly where heat flow measurements are sparse and hydrothermal activities could prevail (e.g., Li et al. 2017). Curie depths have been successfully applied to infer thermal evolution of oceanic lithosphere, global heat loss, lithospheric thermal conductivity, and regional geodynamics (e.g., Bansal et al. 2011; Li et al. 2009, 2010, 2013; Salem et al. 2014; Ravat et al. 2016; Wang and Li 2015), and to correlate with regional magmatism and seismicity (Tanaka and Ishikawa 2002, 2005; Manea and Manea 2011; Wang and Li 2015; Wang et al. 2016).

Despite these important applications, there are many caveats in estimating Curie-point depth, particularly in the application of fractal exponent of source magnetization. This paper is to clarify some of the confusions in applying fractal exponent in Curie depth estimation and outline the pitfalls that should be avoided in future applications.

Numerical backgrounds

Three-dimensional source magnetizations are spatially correlated and can be characterized by a scaling law.

$$\phi_M\left(k_x, k_y, k_z\right) \propto k^{-\beta_{3\mathrm{D}}^p} \tag{1}$$

in which $\beta_{3\mathrm{D}}^p$ is the fractal exponent and the superscript p refers to the definition for power spectrum, $\phi_M\left(k_x, k_y, k_z\right)$ is the 3D power spectrum of the magnetization, k_x, k_y and k_z are wavenumbers in x, y, and z directions, respectively, and their Euclidean norm $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$.

Curie depths are dependent on wavenumber (or wavelength) distribution of magnetic anomalies, because the deeper the base of the magnetic layer, the more present are longer wavelength components. Theoretically, the radially averaged power (or amplitude) spectrum of total-field magnetic anomalies $A_{\Delta T}$ is linked to the spectrum of the magnetization and can be represented as a function of depths to the bottom (Z_b) and top (Z_t) of the magnetic layer, and the fractal exponent (Maus et al. 1997; Bouligand et al. 2009; Blakely 1995; Li et al. 2009).

To estimate Curie depths based on spectral methods, magnetic anomalies are interpolated and gridded and then divided into overlapping windows. Within each window, a radially averaged amplitude (or power) spectrum is calculated, from which a Curie depth is estimated from fitting the calculated spectrum to the theoretical models of Blakely (1995) or Maus et al. (1997). This windowing scheme can be skipped using wavelet transform (Gaudreau et al. 2019).

On the assumption of constant fractal exponent

In general, the fractal exponent of magnetization (named α in Ravat et al. 2016 in 2D, or β in 3D in many other papers, e.g., Bouligand et al. 2009; Li et al. 2013) is not known. In order to keep the inversion result stable and manageable, the fractal exponent is often assumed to be a known constant in a study area. Of course, it is unlikely that fractal exponent keeps constant across a large area due to the differing magnetic source characteristics. Undoubtedly, any method for Curie depth estimation is biased by the lack of knowledge of the fractal exponent (Audet and Gosselin 2019), or by using a single fractal exponent. Previous attempts have been made to estimate the fractal exponent $\left(\beta_{3\mathrm{D}}^p\right)$ simultaneously with the depths to the top and bottom (Z_b, Z_t) from magnetic anomalies based on a nonlinear inversion scheme, but it turned out to be very difficult as these parameters are strongly inter-dependent (Ravat et al. 2007; Li et al. 2010). The best constraints on estimated Curie depths and the fractal exponent are from known geology, such as shallow geotherms associated with mid-ocean ridges and active volcanoes (Li et al. 2013, 2017). Alternatively, Mather and Fullea (2019) combined independent geophysical data with magnetic anomaly data in a probabilistic framework to constrain geotherms. Gaudreau et al. (2019) determine the fractal exponent a posteriori by comparing Monte Carlo simulations of predicted heat flow with observed heat flow in various regions.

One of the strategies is a stepwise linearized inversion for Z_t at intermediate to high wavenumbers and the depth to the centroid (Z_o) at small wavenumbers (Tanaka et al. 1999), assuming a regionally constant fractal exponent. This constant assumption is found effective and valid in previous regional and global studies, giving useful geothermal information conformable to real geology (e.g., Bouligand et al. 2009; Li et al. 2013, 2017).

Li et al. (2009, 2013) applied the centroid method assuming a constant fractal exponent. Bouligand et al. (2009) applied a one-step nonlinear fitting in their western North America study, also assuming a constant fractal exponent. Uncertainties are involved in selecting the best fitting intervals in the two-step linearized method. However, with fixed fitting intervals, Li et al. (2010) showed that the two-step centroid method can give more stable Curie depth results than the one-step nonlinear simultaneous inversion.

On the "de-fractal" method

Salem et al. (2014) and Ravat et al. (2016) argued that they could test a set of fractal exponents β by visual inspection of fit between observed and modeled power spectra,

and found the optimal β for correcting observed spectrum before estimating Curie depth (Fig. 1). They coined their methods "de-fractal" spectral depth determination, which involves forward modeling to fit spectral peaks, and numerical iteration and visual inspection to select β. These iterative procedures are time-consuming.

Correction of the spatially correlated magnetization has been practiced in many other studies (e.g., Bouligand et al. 2009; Bansal et al. 2011; Li et al. 2013). The terminology "de-fractal" is unnecessary for the following reasons.

(1) Magnetization is universally fractal, i.e., spatially correlated. Fitting an observed spectrum with a theoretical model can be done without "de-fractal" because the theoretical models can handle fractal magnetization neatly (Blakely 1995; Maus et al. 1997; Li et al. 2013).
(2) "De-fractal" is operated only in a relative sense because two-dimensional spatially uncorrelated magnetization can be spatially correlated in three dimensions. "De-fractal" is meaningful only when the dimension of the reference space is identified. Randomness (uncorrelation) is just a special form of fractal.

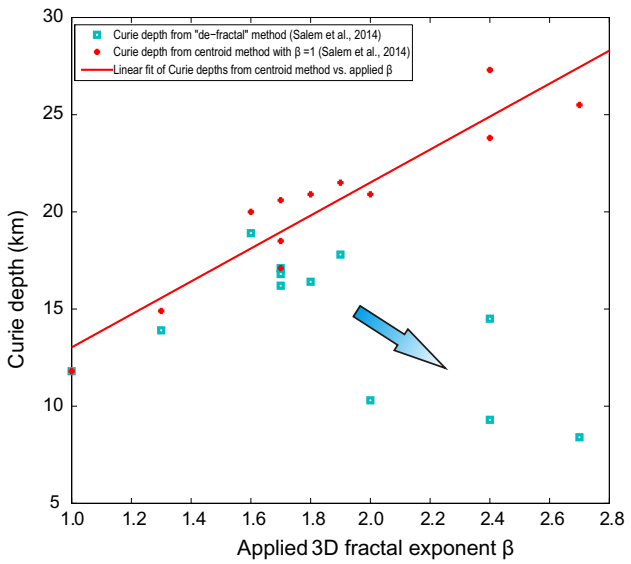


Fig. 1 A positive correlation is noticed between Curie depths estimated with a constant fractal exponent (red cross) and applied fractal exponents in the "de-fractal" method in the central Red Sea. In other words, the "de-fractal" method tends to apply a larger fractal exponent where the Curie depths could be larger if keeping a constant fractal exponent, and consequently gives smaller Curie depth estimates (data in blue squares). This over-correlation tendency (marked by the blue arrow) leads to systematic computational errors in the "de-fractal" method. The straight line is from least square fitting. Depths are below sea level from Salem et al. (2014)

There are other three more important issues in both the visual (Salem et al. 2014) and semiautomatic (Ravat et al. 2016) "de-fractal" method.

(1) The first is the often subjective and random selection of fitting intervals and wavenumbers of spectrum. Theoretical and numerical models suggested that the fitting intervals for estimating the centroid depth should be fixed to the smallest wavenumbers, unless a peak occurs in the fractal-corrected and wavenumber-scaled spectrum, which is likely due to windowing (Li et al. 2013). In the case a peak occurs, points to the smallest wavenumber side of the spectral peak should be simply ignored in data fitting (Li et al. 2013). While estimating Curie depth from the steepest segment is theoretically sound (Li et al. 2013), fitting just on the steepest segment of the spectrum with only 2 or 3 controlling points is prone to large fitting uncertainties.
(2) The second major uncertainty rests upon changing fractal exponent β from window to window (Fig. 1). In each iteration, a modeled power spectrum is produced to match with the observed one, and when an acceptable visual fit is found with a particular fractal exponent, that fractal exponent is chosen as an a priori input for the next step of Curie depth estimation (Ravat et al. 2016). Identical to the number of unknowns in the inversion, an equal number of parameters are needed in the forward modeling of the "de-fractal" scheme. In other words, forward modeling is dependent not just on fractal exponent, but also on depths to the magnetic layer, which are also unknowns. A large Curie depth will induce magnetic anomalies with more long-wavelength signals, as if from a highly correlated magnetization of large fractal exponent, which also equivalently induces more long-wavelength magnetic signals. The reverse is also true. There is essentially no work-around to know the best-fit fractal exponent. Therefore, the "de-fractal" method is circular and does not have anything internal to the Curie depth calculation to tie results to. Consequently, the "de-fractal" method results in low resolution and likely high error (Fig. 1).

Changing fractal exponents β that are not well constrained from window to window introduces additional error, because this will smear out Curie depth anomalies associated with true geological features. Although the "de-fractal" method appears to give a mechanism to constrain the fractal exponent β in an iterative way, in reality it can do more harm than help.

By plotting Curie depths from Table 2 of Salem et al. (2014), who applied the "de-fractal" method, we further demonstrate that their applied fractal exponent is strongly correlated with Curie depth estimated

with a constant fractal exponent (Fig. 1). The larger the Curie depth, the larger were their applied fractal exponents for spectral correction, and the smaller were their finally estimated Curie depths. The consequence is that, wherever there are large Curie depths, this "de-fractal" operation tends to systematically pick large fractal exponents and obtain small Curie depths with overcorrection. This can distort the final Curie depth map. In other words, the "de-fractal" method can mistakenly interpret large Curie depths as from large fractal exponents. The best and likely the only constraints on estimated Curie depths are from known geology, such as shallow geotherms associated with mid-ocean ridges and active volcanoes (Li et al. 2013, 2017), or from other independent geophysical measurements such as heat flow (e.g., Mather and Fullea 2019; Gaudreau et al. 2019), but not from some calculated mathematical operations. With a tendency of overcorrection and almost a random selection of fractal exponent that is strongly dependent on the correlation of the treated magnetic anomalies (Fig. 1), the "de-fractal" method cannot map, in a consistent and systematic manner, true geological units of similar scaling in spatial magnetization.

(3) The "de-fractal" method compensates for the fractal parameter of the magnetic anomaly field such that a spectral peak is formed. Whether a peak could occur or not is not solely dependent on the fractal parameter, but also on the Curie depth and applied window size (Li et al. 2010, 2013). For the same fractal parameter, shallow Curie depths can also give spectral peaks. Occurrence of a spectral peak is not a correct criterion for judging the fractal exponent of the underlying magnetization.

On the detection limit

With a 500 km window length and the recommended wavenumber range of Li et al. (2013), Ravat et al. (2016) compared their results from the "de-fractal" method with those of Li et al. (2013), who applied an automatic fractal centroid method. Ravat et al. (2016) showed that they can get even more accurate Curie depth estimate with the smallest error bar for the deepest 40 km depth test.

It is all known in geophysics that the deeper the target, the more uncertainties and difficulties are in geophysical inversion. Ravat et al. (2016) did not state how they obtained the Curie depths from the method of Li et al. (2013) and showed neither numerical/synthetic models (like Fig. 4 of Li et al. 2013), upon which these tests were performed, nor power spectra for fitting. Numerical and synthetic models of 3D magnetization and corresponding magnetic anomalies and power spectra are needed to validate their argument. A regional map of their applied fractal exponents should also be presented to aid in the interpretation and assessment of their results, because the degree of correction affects the estimated Curie depths.

Ravat et al. (2016) showed that they estimated the centroid depth directly from fitting the spectrum itself, not from the required wavenumber-scaled spectrum, because the vertical axes of these two figures are labeled with "Annular Average of ln of Amplitude (nT)." This might be just a typo, and they stated in the caption that the calculation was based on wavenumber-scaled spectrum. However, the labeled unit "nT" is misleading, because the amplitude here is the spectral strength at certain wavenumbers, surely no longer the original magnetic anomaly amplitudes with the unit "nT." Furthermore, using only 2 or 3 controlling points for linear regression for the steepest segment of the spectrum introduces large uncertainties and inconsistencies.

Numerical synthetic modeling with known and fixed fractal exponents showed that, with a set of input Curie depths of 10.0, 20.0, 30.0 and 40.0 km, the inverted depths are 9.5, 13.1, 26.2, and 35.0 km, respectively (Li et al. 2013). Plotting this early result of synthetic test on Fig. A2 of Ravat et al. (2016) shows that the two-step linearized method captures the overall trend of input Curie depth, but tends to underestimate it (Fig. 2). There are several reasons behind this underestimation.

Firstly, magnetic anomalies from deeper sources are more attenuated by the Earth filter, and we have to deal with a narrow band of long wavelengths and work on very small wavenumbers containing the centroid depth information. Secondly, we apply windowing in practice on magnetic anomalies, whereas the mathematical models assume infinite horizontal extension (Blakely 1995; Maus et al. 1997). Thirdly, the linearized centroid technique is itself based on an approximation of the nonlinear system.

These theoretical and practical limitations apply to all Curie depth inversion techniques. We can partly circumvent these issues of underestimation by choosing a smaller fractal exponent in spectrum correction. Nonlinear inversion can be tested on synthetic models to expect larger, albeit unstable, Curie depths (Li et al. 2010). However, there is a mutual dependence of depths to the top and bottom of the magnetic layer in the nonlinear inversion, and solutions can be non-unique. In addition, the nonlinear method also requires data fitting only at the very narrow-banded low-wavenumber portion of the spectrum, producing highly fluctuating results with just a few controlling data points (Li et al. 2010).

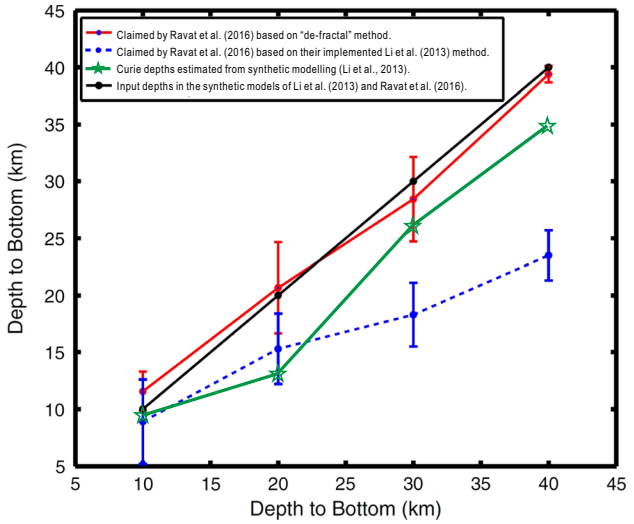


Fig. 2 Comparison of numerical results from the "de-fractal" method with a 500 km window by Ravat et al. (2016), from Li et al. (2013) fractal correction centroid method implemented by Ravat et al. (2016), and from synthetic modeling of 3D magnetization of Li et al. (2013)

Comparison between known Curie depth results in North America

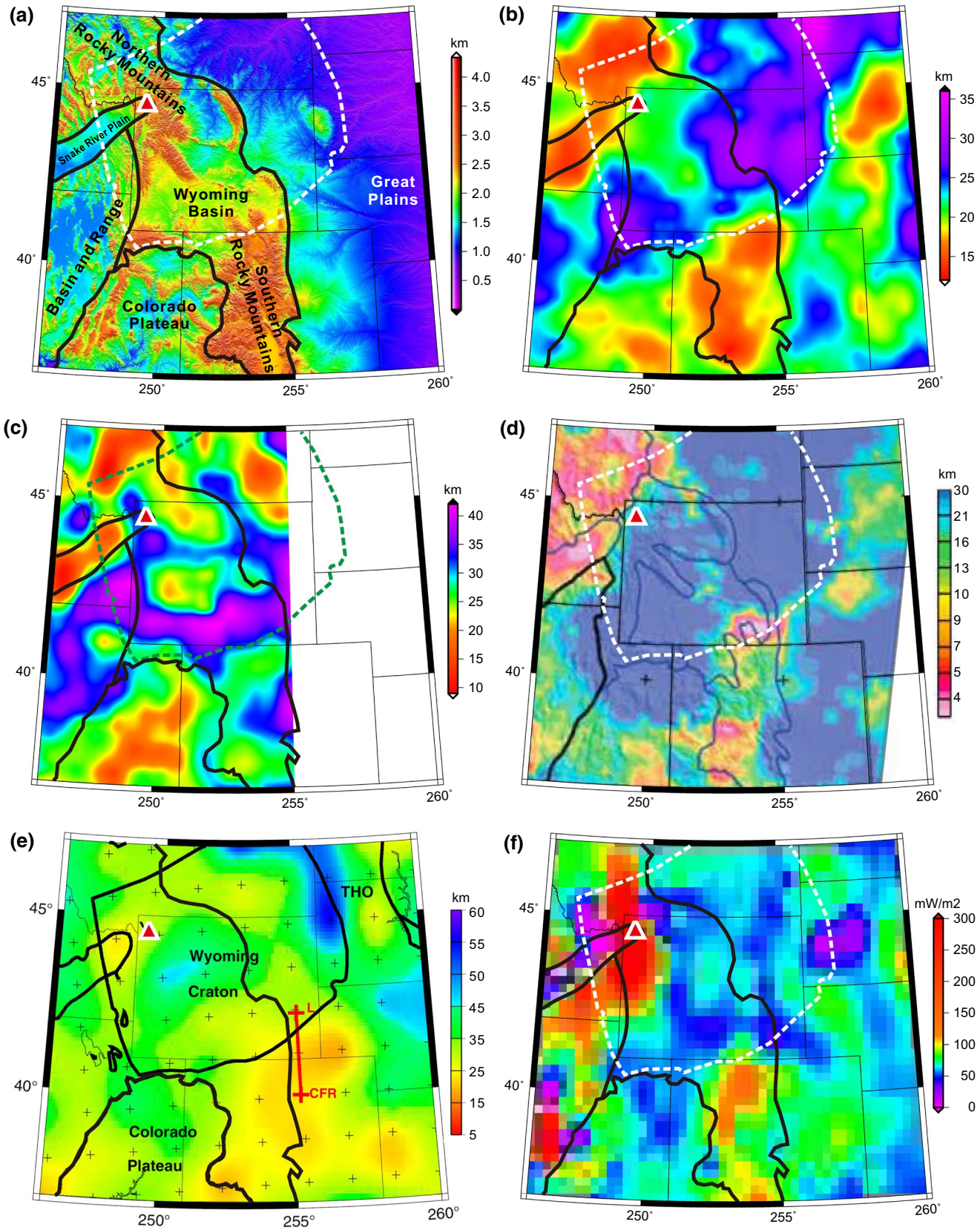
We here make a comparison among published Curie depth results in the Northern Colorado–Wyoming Craton area (Fig. 3; Table 1). Bouligand et al. (2009) mapped Curie depths in the western USA with a fractal magnetization model based on nonlinear inversion (Fig. 3d). Based on the linearized centroid method, Li et al. (2017) developed a global Curie depth model (GCDM), using nearly the same window size range and fractal exponent as Bouligand et al. (2009), but a different magnetic dataset of lower resolution (Maus et al. 2009) (Table 1). Part of the GCDM is shown here for comparison (Fig. 3b). One can easily notice that these two maps show very similar features. The Yellowstone hotspot trail, the northern and southern Rocky Mountains, and a large part of Colorado Plateau have smaller Curie depths. There is also a belt of small Curie depths to the east margin of the study area in the Great Plains. By contrast, the Wyoming Craton shows mostly large Curie depths (Fig. 3b, d). It can also be seen that, as expected and mentioned above, nonlinear inversion resulted in more fluctuating estimates (Fig. 3d) than the centroid method (Fig. 3b), producing many points shallower than 10 km and deeper than 30 km. The apparent higher resolution of Fig. 3d is mostly likely due to the higher resolution of the North America magnetic grid (NAMAG 2002) applied by Bouligand et al. (2009), as well as to more fluctuating nonlinear estimates.

Also based on this high-resolution NAMAG, Wang and Li (2015) examined Curie depths with smaller windows in western North America (Fig. 3c). Figure 3b, c is from different data sources of different resolution and from applying different window sizes, and thereby some differences in resolution and values between them are expected. Both the high-resolution input data and smaller window size gave high resolution in the mapped Curie depths that conform to real geology (Fig. 3c). The central eroded and rifted drainage basin of the Colorado River in the Colorado Plateau shows smaller Curie depths (Fig. 3c), which could indicate thermal rejuvenation at depth. The two areas of smaller Curie depths of the Snake River Plain and the northern Rocky Mountains can be distinguished from each other on the high-resolution result. Overall, these three Curie depth results (Fig. 3b–d) show similar features that are consistent to known geology and can be correlated with surface heat flow (Fig. 3f).

The Yellowstone hotspot turns out not to be a good control point on Curie depth because presently it has very strong hydrothermal activity (Bryan 2008), which can lower the regional deep temperature considerably, like in young oceanic lithospheres. Li and Wang (2018) have shown that strong hydrothermal activity along the fast spreading mid-ocean ridge can lower the mantle temperature and increase the Curie depth. Our reasoning of strong hydrothermal influence on the deep temperature is also drawn from the discrepancy between heat flow (Fig. 3f) and Curie depth (Fig. 3b–d) along the Snake River Plain. Instead of in the central Snake River Plain of the smallest Curie depths, the highest heat flow is found in the surrounding uplifted shoulders of the plain, where fractures, evident on the topographic map (Fig. 1), may drain deep hot hydrothermal fluids.

Again, one cannot guarantee that Curie depth estimation using different window sizes and data of different resolution can give identical result at the same single location, because different window sizes focus on different anomalies, and different data resolution focuses on different wavelengths. Nonetheless, regional features should remain the same and should be captured, such as the shallow Curie depths of the Yellowstone hotspot trail (Snake River Plain, Fig. 3).

The "de-fractal" result (Fig. 3e) is quite different from the other three, neither revealing the large Curie depth contrast between the Colorado Plateau and the Wyoming Craton, nor showing small Curie depth zones associated with the Snake River Plain and the northern Rocky Mountains. Instead, Fig. 3e shows smaller Curie depths to the east of the southern Rocky Mountains, which are not present on other three maps. Without knowing areal distribution of the fractal exponents, it is very difficult to assess the "de-fractal" result. Surface heat flow (Fig. 3f) has better correlations to the Curie depths of Bouligand et al. (2009) and Li et al. (2017) than to the "de-fractal" result (Fig. 3e). This demonstrates again that the "de-fractal" method, with variable fractal exponents that cannot be accurately determined and are strongly correlated with Curie depths themselves (Fig. 1),


(a)
Northern Rocky Mountains
Snake River Plain
Wyoming Basin
Great Plains
Basin and Range
Rocky Mountains
Southern
Colorado Plateau
km
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
45°
40°
250°
255°
260°
(b)
km
35
30
25
20
15
(c)
km
40
35
30
25
20
15
10
(d)
km
30
21
16
13
10
9
7
5
4
(e)
THO
Wyoming
Craton
L
CFR
Colorado
Plateau
km
60
55
50
45
35
25
15
5
(f)
mW/m2
300
250
200
150
100
50
0

◂**Fig. 3** Comparison between known Curie depth results in the western North America. **a** Topography of the study area. The white dashed line outlines the Wyoming Craton shown in Ravat et al. (2016). Thick black lines outline major tectonic units. The red triangle marks the Yellowstone hotspot. **b** Curie depths from the global reference Curie depth model (GCDM) of Li et al. (2017). **c** Curie depths from Wang and Li (2015). The green dashed line outlines the Wyoming Craton shown in Ravat et al. (2016). No Curie depths were obtained to the east of the 255° longitude line. **d** Curie depths from Bouligand et al. (2009). **e** Curie depths from the "de-fractal" method, and the black solid line outlines the Wyoming Craton (Ravat et al. 2016). **f** Surface heat flow gridded in a 30′ interval using the minimum curvature algorithm with tension (Briggs 1974) (heat flow data from the international heat flow commission database https://www.heatflow.und.edu/; last updated in January 2011). No preselection or preprocessing is done on the original raw heat flow data. Red line shows the location of a thermal property profile in Ravat et al. (2016). Data mapping is supported by GMT (Wessel and Smith 1995)

can smear out true geological features with distorted Curie depth maps.

Application of "de-fractal" Curie depths casts doubt on the derived geothermal result, which shows that Wyoming geotherm is hotter by ~ 200 °C at the Curie depth than the Decker et al. (1988) model. The Wyoming Craton shows largely low heat flow (Fig. 3f) and deep and heterogeneous Curie depths (Fig. 3b–d). The heterogeneity in the Wyoming Craton stems from late tectonism, which has already divided it into different subunits of contrasting topography (Fig. 3a) and heat flow (Fig. 3f). Strictly, the Wyoming Craton can no longer be regarded as a typical craton. The anomalously hotter and uniform geotherm of Ravat et al. (2016) is, in our opinion, due to their estimated shallow Curie isotherm, which makes no distinction of the Wyoming Craton on their Curie depth map (Fig. 3e).

A very large applied window size of 500.0 km may also contribute to this loss of information. Ideally, the long-wavelength components carrying the information of Curie depth can be best captured by using very large windows, but the incorporation of fractal exponent, which deals with the wide (correlated) but shallow anomalies, partially relieves this requirement. Increasing window size does not appreciably increase calculated Curie depth, but merely leads to a low resolution (Li et al. 2010, 2013). This is because the extra information we can gain at the smallest wavenumbers (or longest wavelengths) is rather minimal (Fig. 4); one can never approach the theoretically required scale of infinity by just attempting to increase the window length by several hundred kilometers. The strong averaging effect of large windows can decrease, not increase, locally large Curie depths, for example, those associated with a cold accretionary wedge. Features smaller than the chosen window size will not be properly imaged, because normally only one Curie depth is estimated in each window, and the small feature contributes only partially to the radially averaged spectrum in that window. It can also be seen that most of the "de-fractal" Curie depths in the study area are between 15 and 40 km, although the color bar shows much larger values (Fig. 3e), and are not appreciably larger than those from using the centroid method (Fig. 3b, c). A Curie depth comparison between using a larger window size (Fig. 3b) and a smaller window sizes (Fig. 3c) also shows that a significantly large window size at 500.0 km is not necessary.

Conclusion

This paper intends to clarify that mathematical treatment beyond the limit of data resolution and the underlying physics could introduce additional errors to Curie depth estimation. Wherever there are true large Curie depths, the "de-fractal" method, by its very nature, has a tendency of overcorrecting fractal exponents and thereby producing small Curie depths and smearing out true geological trends.

For Curie depth estimation in a large area, the fractal exponent cannot be a constant, but it can be better fixed than variable but just loosely controlled purely by mathematical overtreatment. At long wavelengths containing primarily the Curie depth information, fractal exponents of source magnetizations are expected to be rather stable over a large area. Long-distance spatial correlation in source magnetization

Table 1 Comparison of parameters applied in four different Curie depth results

Author	Method	Window size (km)	3D fractal exponent β	Data source
Bouligand et al. (2009)	Nonlinear inversion	100.0 to 300.0	3.0	Magnetic anomaly map of North America (NAMAG 2002) and the state map of Nevada (Kucks et al. 2006)
Wang and Li (2015)	Centroid	80.0, 100.0 and 120.0	2.5	Magnetic anomaly map of North America (NAMAG 2002)
Li et al. (2017)	Centroid	98.8, 195.0, and 296.4	3.0	Earth Magnetic Anomaly Grid of 2′ resolution (EMAG2, Maus et al. 2009)
Ravat et al. (2016)	"De-fractal," based on centroid	500.0	Variable but unknown	Unknown

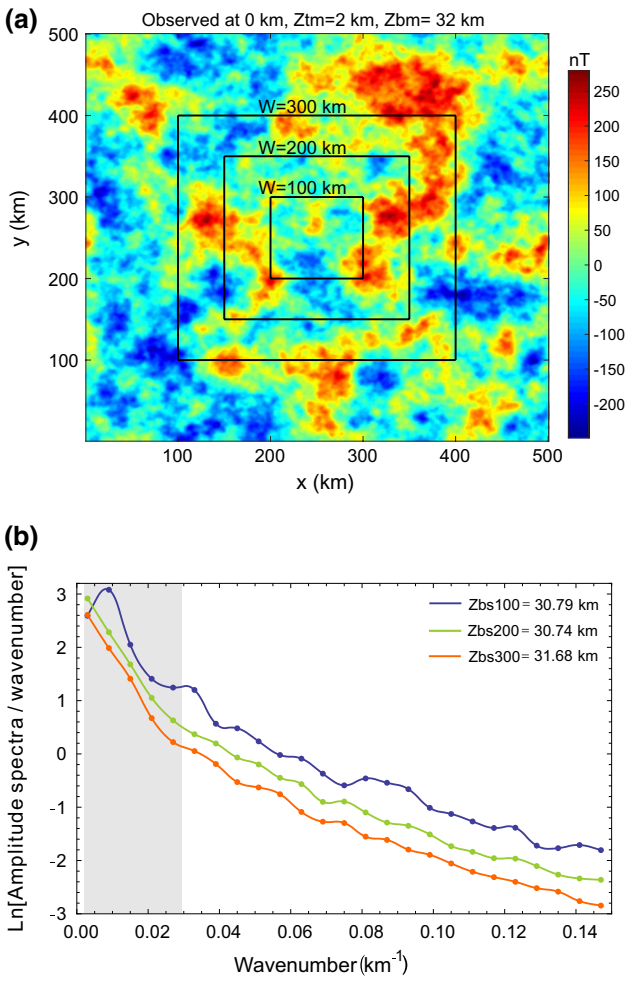


Fig. 4 Testing of window size effect on Curie depth estimation. **a** Magnetic anomalies calculated at the surface of the North Pole from a 3D synthetic fractal magnetization model. The magnetization has a Gaussian distribution with a mean value of 0 A/m and a standard deviation of 0.3 A/m. The fractal exponent, the top (*Z*tm) and bottom (*Z*bm) depths to the magnetic sources of the input model are taken as 3, 2 km and 32 km, respectively. Black squares depict the three window lengths of 100, 200 and 300 km. **b** Wavenumber-scaled radially averaged amplitude spectra of the synthetic magnetic anomalies using the three different window sizes shown in **a**. Estimated Curie depths indicate that increasing window size does not considerably lead to larger Curie depths. The gray shadow zone indicates where the centroid depths are estimated from linear regressions. The centroid depth predominates the Curie depth

is related to more regional geological features, whose geothermal conditions and source magnetizations are unlikely to alter swiftly in short distances or between two nearby windows. Therefore, a constant fractal exponent constrained by geology is preferred, when the true fractal exponent is not obtainable, over a method trying to vary the fractal exponents but in an overcorrection tendency.

Significantly large window lengths at ~ 500.0 km are not necessary for capturing large Curie depths. Normally using multiple window sizes ranging from tens of kilometers to 200.0 km is sufficient, and an average from these different windows can suppress random noise and increase the resolution of Curie depths. With magnetic anomalies of increasing resolutions to be available in the future, the calculated Curie depths should be improved, mostly from better calibrating the depths to the magnetic top in the intermediate to large wavenumbers. But since Curie depths are more dependent on long wavelengths, better data coverage is even more critical.

The linearized stepwise centroid method has proven to be stable and efficient, and gained more applications. In recent years, new techniques, such as Bayesian inversion (Mather and Fullea 2019; Audet and Gosselin 2019), multitaper spectral analysis (Audet and Gosselin 2019), and wavelet transform (Gaudreau et al. 2019), are being applied in Curie depth estimation. By statistically incorporating independent geological and geophysical constraints, the fractal exponent could be better estimated prior to the inversion of Curie depth. Ensemble-based approaches can produce probability distributions and provide greater confidence for the recovered parameters.

Acknowledgements This study is funded by National Natural Science Foundation of China (Grant Nos. 41776057, 41761134051, 41704086 and 91858213). Data mapping is supported by GMT (Wessel and Smith 1995). We thank anonymous reviewers for their critical but positive suggestions.

References

Audet P, Gosselin JM (2019) Curie depth estimation from magnetic anomaly data: a re-assessment using multitaper spectral analysis and Bayesian inference. Geophys J Int 218:494–507. https://doi.org/10.1093/gji/ggz166

Bansal AR, Gabriel G, Dimri VP, Krawczyk CM (2011) Estimation of depth to the bottom of magnetic sources by a modified centroid method for fractal distribution of sources: an application to aeromagnetic data in Germany. Geophysics 76:L11–L22. https://doi.org/10.1190/1.3560017

Blakely RJ (1995) Potential theory in gravity and magnetic applications. Cambridge University Press, Cambridge, pp 1–464

Bouligand C, Glen JMG, Blakely RJ (2009J) Mapping Curie temperature depth in the western United States with a fractal model for crustal magnetization. J Geophys Res 114:B11104. https://doi.org/10.1029/2009JB006494

Briggs IC (1974) Machine contouring using minimum curvature. Geophysics 9:39–48

Bryan TS (2008) The geysers of Yellowstone. University Press of Colorado, Niwot, pp 1–456

Decker ER, Heasler HP, Buelow KL, Baker KH, Hallin JS (1988) Significance of past and recent heat-flow and radioactivity studies in the Southern Rocky Mountains region. Colo Geol Soc Am Bull 100:1851–1885

Friedman SA, Feinberg JM, Ferré EC, Demory F, Martín-Hernández F, Conder JA, Rochette P (2014) Craton vs. rift uppermost mantle contributions to magnetic anomalies in the United States interior. Tectonophysics 624–625:15–23

Gaudreau É, Audet P, Schneider DA (2019) Mapping Curie depth across western Canada from a wavelet analysis of magnetic anomaly data. J Geophys Res 124:4365–4385. https://doi.org/10.1029/2018JB016726

Kucks RP, Hill PL, Ponce DA (2006) Nevada magnetic and gravity maps and data: a website for the distribution of data, U.S. geological survey data series, 234. https://pubs.usgs.gov/ds/2006/234. Accessed Aug 2017

Lachenbruch AH (1968) Preliminary geothermal model of the Sierra Nevada. J Geophys Res 73:6977–6988

Li C-F, Wang J (2018) Thermal structures of the Pacific lithosphere from magnetic anomaly inversion. Earth Planet Phys 2:52–66

Li C-F, Chen B, Zhou Z (2009) Deep crustal structures of eastern China and adjacent seas revealed by magnetic data. Sci China Ser D Earth Sci 52:984–993. https://doi.org/10.1007/s11430–009–0096-x

Li C-F, Shi XB, Zhou ZY, Li JB, Geng JH, Chen B (2010) Depths to the magnetic layer bottom in the South China Sea area and their tectonic implications. Geophys J Int 182:1229–1247. https://doi.org/10.1111/j.1365-246X.2010.04702.x

Li C-F, Wang J, Lin J, Wang T (2013G) Thermal evolution of the North Atlantic lithosphere: new constraints from magnetic anomaly inversion with a fractal magnetization model. Geochem Geophys Geosyst 14:5078–5105. https://doi.org/10.1002/2013GC004896

Li C-F, Lu Y, Wang J (2017) A global reference model of Curie-point depths based on EMAG2. Sci Rep 7:45129. https://doi.org/10.1038/srep45129

Manea M, Manea VC (2011) Curie point depth estimates and correlation with subduction in Mexico. Pure Appl Geophys 168:1489–1499

Mather B, Fullea J (2019) Constraining the geotherm beneath the British Isles from Bayesian inversion of Curie depth: integrated modelling of magnetic, geothermal, and seismic data. Solid Earth 10:839–850. https://doi.org/10.5194/se-2019-9

Maus S, Gordan D, Fairhead D (1997) Curie-temperature depth estimation using a self-similar magnetization model. Geophys J Int 129:163–168. https://doi.org/10.1111/j.1365-246X.1997.tb00945.x

Maus S, Barckhausen U, Berkenbosch H, Bournas N, Brozena J, Childers V, Dostaler F, Fairhead JD, Finn C, von Frese RRB, Gaina C, Golynsky S, Kucks R, Luhr H, Milligan P, Mogren S, Müller D, Olesen O, Pilkington M, Saltus R, Schreckenberger B, Thébault E, Caratori Tontini F (2009G) EMAG2: a 2-arc-minute resolution earth magnetic anomaly grid compiled from satellite, airborne and marine magnetic measurements. Geochem Geophys Geosyst 10:Q08005. https://doi.org/10.1029/2009GC002471

Negi JG, Agrawal PK, Pandey OP (1987) Large variation of Curie depth and lithospheric thickness beneath the Indian subcontinent and a case for magnetothermometry. Geophys J R Astr Soc 88:763–775

North American Magnetic Anomaly Group (NAMAG) (2002) Digital data grids for the magnetic anomaly map of North America, U.S. geological survey open file rep. 02–414. https://pubs.usgs.gov/of/2002/ofr-02-414/. Accessed Aug 2017

O'Reilly W (1984) Rock and mineral magnetism. Blackie Academy and Professor, London

Okubo Y, Graf RJ, Hansen RO, Fytikas M (1985) Curie point depths of the Island of Kyushu and surrounding areas, Japan. Geophysics 53:481–494. https://doi.org/10.1190/1.1441926

Ravat D, Pignatelli A, Nicolosi I, Chiappini M (2007) A study of spectral methods of estimating the depth to the bottom of magnetic sources from near-surface magnetic anomaly data. Geophys J Int 169:421–434

Ravat D, Morgan P, Lowry A (2016) Geotherms from the temperature-depth–constrained solutions of 1-D steady-state heat-flow equation. Geosphere 12:1187–1197. https://doi.org/10.1130/GES01235.1

Salem A, Green C, Ravat D, Singh KH, East P, Fairhead JD, Mogren S, Biegert E (2014) Depth to Curie temperature across the central Red Sea from magnetic data using the de-fractal method. Tectonophysics 624–625:75–86

Sauerzapf U, Lattard D, Burchard M, Engelmann R (2008) The titanomagnetite-ilmenite equilibrium: new experimental data and thermooxybarometric application to the crystallization of basic to intermediate rocks. J Petrol 49:1161–1185

Tanaka A, Ishikawa Y (2002) Temperature distribution and focal depth in the crust of the northeastern Japan. Earth Planets Space 54(11):1109–1113

Tanaka A, Ishikawa Y (2005) Crustal thermal regime inferred from magnetic anomaly data and its relationship to seismogenic layer thickness: the Japanese islands case study. Phys Earth Planet Int 152(4):257–266

Tanaka A, Okubo Y, Matsubayashi O (1999) Curie point depth based on spectrum analysis of the magnetic anomaly data in East and Southeast Asia. Tectonophysics 306:461–470. https://doi.org/10.1016/S0040-1951(99)00072-4

Turcotte DL, Schubert G (2002) Geodynamics. Cambridge University Press, New York

Wang J, Li C-F (2015) Crustal magmatism and lithospheric geothermal state of western North America and their implications for a magnetic mantle. Tectonophysics 638:112–125. https://doi.org/10.1016/j.tecto.2014.11.002

Wang J, Li CF, Lei JS, Zhang GW (2016) Relationship between seismicity and crustal thermal structure in North China. Acta Seismol Sin 38(4):618–631

Wessel P, Smith WHF (1995) New version of the generic mapping tools (GMT) version 3.0 released. Trans Am Geophys Union EOS 76:329

Acta Geophysica (2019) 67:1329–1340
https://doi.org/10.1007/s11600-019-00258-6

RESEARCH ARTICLE - APPLIED GEOPHYSICS

Direct inversion for sensitive elastic parameters of deep reservoirs

Shengjun Li[1,2] · Jinyong Gui[2] · Jianhu Gao[2] · Shangxu Wang[1] · Hailiang Li[2]

Received: 28 September 2018 / Accepted: 29 January 2019 / Published online: 22 June 2019

Abstract
The deep reservoir is usually a type of tight reservoir with high pressure, high stress, low permeability and low porosity. The elastic parameters including Poisson's ratio and Young's modulus are important sensitive parameters to the tight reservoir, and the Gassmann fluid term is frequently used in the field of fluid identification as a highly sensitive fluid factor. Such parameters can be obtained by the common prestack seismic inversion method, but not directly. It must first invert for other elastic parameters and then convert them into the Poisson's ratio, Young's modulus and Gassmann fluid term by some formula. The errors will be accumulated in the conversion step, and the inversion results will have a large deviation. We propose a one-step inversion method to solve this problem. Firstly, a new form of P-wave reflection coefficient equation in terms of Poisson's ratio, Young's modulus and Gassmann fluid term is derived which can directly establish the functional relationship between the P-wave reflection coefficient and these elastic parameters. Considering seismic data of deep reservoir generally have a lower signal-to-noise ratio (*S*/*N*) and the partial angle stack gather has a higher *S*/*N* than single angle gather, we then derive a stack impedance equation which is suitable for the partial angle stack gather. By using three stacked impedance inversion data with different angle stack ranges, we can directly get the Poisson's ratio, Young's modulus and Gassmann fluid term simultaneously. Model and real data tests both prove that the one-step direct inversion method can reduce the cumulative errors effectively and has higher inversion accuracy.

Keywords Deep reservoir · Direct inversion · P-wave reflection coefficient equation · Stack impedance

Introduction

Elastic parameters, such as velocity, density and impedance, play a significant role in oil and gas exploration (Shi et al. 2018; Yuan et al. 2019b). Reservoirs with different types of lithology and fluids usually have different elastic properties. The proposal of the Zoeppritz equation makes the elastic parameters of rocks related to seismic amplitude, which is the theoretical basis for current prestack seismic inversion and fluid identification (Zoeppritz 1919; Hilterman 2001). Bortfeld (1961) gave the approximate equation of Zoeppritz equation for the first time since the Zoeppritz equation is highly nonlinear and the solution is not often stable. This makes the relationship between the seismic amplitude implied in the Zoeppritz equation and the elastic properties of rocks more clear. Aki and Richards (1980) derived the approximate equation including the P- and S-wave velocities and density term, now became the most common Zoeppritz approximate equation in prestack seismic inversion. Shuey (1985) deduced the direct functional relationship between P-wave reflection coefficient and Poisson's ratio. He also proved that the gradient of reflection coefficient with the incident angle was mainly determined by the change in Poisson's ratio. More importantly, the concept of amplitude versus offset (AVO) intercept and gradient were proposed and the approximate equation of the reflection coefficient expressed by different angle terms was given. On the basis of Castagna's mudstone baseline (Castagna et al. 1985), Smith and Gidlow (1987) integrated the P- and S-wave velocity reflectivity into a new parameter and found that the parameter is very sensitive to fluid, and then, the concept of "fluid factor" was first proposed. Fatti et al. (1994) reconstituted the Aki–Richards approximate equation, inverted for the P- and S-wave impedance reflectivity by weighted superposition

✉ Jinyong Gui
guijy@petrochina.com.cn

[1] China University of Petroleum, Changping, Beijing 102249, China

[2] Research Institute of Petroleum Exploration and Development-Northwest, Petrochina, Lanzhou 730020, Gansu, China

method, and improved the Smith–Gidlow's fluid factor. Goodway et al. (1997) put forward Lambda–Mu–Rho (LMR) method, which is of great significance in the field of fluid identification. The shear modulus and Lame parameter are calculated by P- and S-wave impedance conversion and used as an indicator tool for fluid identification. Gray (1999) re-deduced Zoeppritz approximate equation and proposed a new P-wave reflection coefficient equation including Lame parameter, shear modulus and bulk modulus. It can directly and simultaneously invert for Lame parameter, shear modulus and density without P- and S-wave impedance conversion. Russell et al. (2011) found that the most of the existing elastic parameter are related to the rock skeleton, which limits the ability of fluid identification. They proposed the definition formula of the Gassmann fluid term, which can eliminate the influence of pore, and used it as an indicator for fluid identification. Subsequent studies proved that the Gassmann fluid term as a new fluid factor has higher fluid sensitivity than the common fluid factors (Feng et al., 2007; Zhang et al., 2009; Liu and Yin 2014; Zhang et al., 2018). Recently, Yuan et al. (2019a) proposed a new prestack inversion to directly invert for the frequency-dependent velocity to favorably detect hydrocarbons.

With the increasing demand for oil and natural gas, the deep reservoir has become one of the most important potential exploration targets in the world and the corresponding exploration technology has become a hotspot of current research in the field of geophysics (Wang et al. 2018). The deep reservoir is usually characterized by high pressure, stress, and low permeability, porosity, which is a type of tight reservoir. Seismic rock physics experiments and exploration practices have proved that rock brittleness is an important index in hydraulic fracturing evaluation of tight reservoir and can be characterized by Poisson's ratio and Young's modulus (Sena et al. 2011; Harris et al. 2011; Zong et al. 2013; Nebojsa and Nina 2017). Generally, the "sweet spot" of deep reservoir has a relatively low Poisson's ratio value, but a high young's modulus value. In practical applications, the common two-step inversion method is typically adopted to obtain Poisson's ratio, Young's modulus and Gassmann fluid term. Firstly, to invert for P- and S-wave velocity and density by prestack seismic inversion method. Secondly, converting the P- and S-wave velocity and density into the Poisson's ratio, Young's modulus and Gassmann fluid term by some conversion formula. However, in the target area of deep reservoir, seismic data have a small illumination angle, low effective coverage and low *S*/*N*, as well as lack of large-offset information, which brings new challenges to prestack seismic inversion. Nevertheless, influenced by data quality and inversion algorithm in the first step, the prestack seismic inversion results of P- and S-wave velocity and density may have a certain degree of errors. Besides, the inversion accuracy of P- and S-wave velocity and density is different. Their accuracy usually decreases from P-velocity to density (Mallick 2001). Therefore, the errors will be accumulated in the conversion step, and the inversion results will have a large deviation.

In this study, we derive a new kind of P-wave reflection coefficient equation. Based on this new equation, we propose a sensitive elastic parameter inversion method which is suitable for deep reservoir. Tests of model and real data show that we can not only invert for the Poisson's ratio and Young's modulus with high accuracy directly, but also can invert for the Gassmann fluid term simultaneously by using this method.

Methods

Based on the P-wave reflection coefficient equation of Gray (1999), Russell et al. (2011) derived a P-wave reflection coefficient equation which is a function of the shear modulus, density and Gassmann fluid term, as

$$R_{pp}(\theta) = \left[\left(\frac{1}{4} - \frac{\gamma_{\text{dry}}^2}{4\gamma_{\text{sat}}^2}\right)\sec^2\theta\right]\frac{\Delta f}{f} + \left(\frac{\gamma_{\text{dry}}^2}{4\gamma_{\text{sat}}^2}\sec^2\theta - \frac{2}{\gamma_{\text{sat}}^2}\sin^2\theta\right)\frac{\Delta\mu}{\mu} + \left(\frac{1}{2} - \frac{\sec^2\theta}{4}\right)\frac{\Delta\rho}{\rho} \tag{1}$$

where θ stands for the incident angle. γ_{dry} and γ_{sat} stand for the ratio of P-wave velocity to S-wave velocity in dry rocks and saturated rocks, respectively. $\Delta f/f$ stands for the Gassmann fluid term reflectivity. $\Delta\mu/\mu$ stands for the shear modulus reflectivity. $\Delta\rho/\rho$ stands for the density reflectivity.

Equation (1) is usually used to invert for the Gassmann fluid term which is frequently used in the field of fluid identification as a highly sensitive fluid factor. Our purpose is to make the P-wave reflection coefficient directly related to the Poisson's ratio, young's modulus and Gassmann fluid term, and can be used for deep reservoir prediction.

For isotropic medium, the relationship among shear modulus, Young's modulus and Poisson's ratio is

$$\mu = \frac{E}{2(1+\sigma)} \tag{2}$$

where μ is shear modulus, E is Young's modulus and σ is Poisson's ratio.

Defining $X = \frac{1}{2(1+\sigma)}$, then we have

$$\mu = EX \tag{3}$$

Taking complete differential operation of μ, then

$$\mathrm{d}\mu = X\mathrm{d}E + E\mathrm{d}X \tag{4}$$

Both sides of Eq. (4) are divided by μ, then Eq. (4) can be written as

$$\frac{\mathrm{d}\mu}{\mu} = \frac{\mathrm{d}E}{\mu}X + \frac{\mathrm{d}X}{\mu}E \tag{5}$$

According to Eq. (3), we have

$$X = \frac{\mu}{E} \tag{6}$$

$$E = \frac{\mu}{X} \tag{7}$$

Substituting Eqs. (6) and (7) into Eq. (5), then we have

$$\frac{\Delta\mu}{\mu} = \frac{\Delta E}{E} + \frac{\Delta X}{X} \tag{8}$$

The $\frac{\Delta X}{X}$ in Eq. (8) can be written as

$$\frac{\Delta X}{X} = 2\frac{\sigma_1 - \sigma_2}{2 + \sigma_1 + \sigma_2} \tag{9}$$

where σ_1 and σ_2 stand for the Poisson's ratio of upper strata and lower strata of the reflection interface, respectively.

Because

$$\sigma_1 = \sigma - \frac{\Delta\sigma}{2} \tag{10}$$

$$\sigma_2 = \sigma + \frac{\Delta\sigma}{2} \tag{11}$$

Substituting Eqs. (10) and (11) into Eq. (9), then

$$\frac{\Delta X}{X} = \frac{-\sigma}{1 + \sigma}\frac{\Delta\sigma}{\sigma} \tag{12}$$

The σ and γ_{sat} have the following relationship

$$\sigma = \frac{\gamma_{sat}^2 - 2}{2\gamma_{sat}^2 - 2} \tag{13}$$

Substituting Eq. (13) into Eq. (12), then Eq. (12) can be written as

$$\frac{\Delta X}{X} = \frac{2 - \gamma_{sat}^2}{3\gamma_{sat}^2 - 4}\frac{\Delta\sigma}{\sigma} \tag{14}$$

Substituting Eq. (14) into Eq. (8), then we have

$$\frac{\Delta\mu}{\mu} = \frac{\Delta E}{E} + \frac{2 - \gamma_{sat}^2}{3\gamma_{sat}^2 - 4}\frac{\Delta\sigma}{\sigma} \tag{15}$$

According to Liu and Yin (2014), the relationship between the density and shear modulus can be written as

$$\frac{\Delta\rho}{\rho} = \frac{1}{1 + 2r}\frac{\Delta\mu}{\mu} \tag{16}$$

Substituting Eqs. (15) and (16) into Eq. (1), a new reflection coefficient equation is derived as

$$R_{pp}(\theta) = a(\theta)\frac{\Delta f}{f} + b(\theta)\frac{\Delta E}{E} + c(\theta)\frac{\Delta\sigma}{\sigma} \tag{17}$$

where $a(\theta) = \left(\frac{1}{4} - \frac{\gamma_{dry}^2}{4\gamma_{sat}^2}\right)\sec^2\theta$, $b(\theta) = \frac{1}{4}\left(\frac{\gamma_{dry}^2}{\gamma_{sat}^2} - \frac{1}{1+2r}\right)\sec^2\theta - \frac{2}{\gamma_{sat}^2}\sin^2\theta + \frac{1}{2+4r}$, $c(\theta) = \frac{2-\gamma_{sat}^2}{12\gamma_{sat}^2-16}\frac{\gamma_{dry}^2(1+2r)-\gamma_{sat}^2}{\gamma_{sat}^2(1+2r)}\sec^2\theta - \frac{4-2\gamma_{sat}^2}{3\gamma_{sat}^4-4\gamma_{sat}^2}\sin^2\theta + \frac{2-\gamma_{sat}^2}{(2+4r)(3\gamma_{sat}^2-4)}$.

By Eq. (17), we finally get the direct functional relationship between the P-wave reflection coefficient and the Poisson's ratio, Young's modulus and Gassmann fluid term.

For conventional reservoirs, based on Eq. (17), the common AVO inversion method can be used to obtain the Poisson's ratio, Young's modulus and Gassmann fluid term. However, for deep reservoir, due to the low *S*/*N* of angle gathers, the AVO inversion may be unstable and have large errors. The elastic impedance (EI) inversion which uses partial angle stack gathers has the advantages of both the post-stack seismic inversion and prestack seismic AVO inversion, and is widely used for elastic parameter inversion (Connolly, 1999; Mallick 2001; Zong et al. 2013). Compared with the single angle gather, the partial angle stacked gather has a higher *S*/*N*, since the stack processing can remove some random noise and enhance the effective signal. Therefore, the EI inversion may be suitable for deep seismic data. According to Connolly's idea of EI equation, we can derive an EI equation related to the Poisson's ratio, Young's modulus and Gassmann fluid term based on Eq. (17), as

$$\mathrm{EI}(\theta) = \mathrm{EI}_0\left(\frac{f}{f_0}\right)^{2a(\theta)}\left(\frac{E}{E_0}\right)^{2b(\theta)}\left(\frac{\sigma}{\sigma_0}\right)^{2c(\theta)} \tag{18}$$

where EI_0 stands for the normalization factor. σ_0, E_0 and f_0 stand for the average of Poisson's ratio, Young's modulus and Gassmann fluid term of target areas, respectively.

We can see that the EI equation is a function of single incident angle; therefore, single incident angle gather need be used for precise EI inversion. However, in practical applications, when extracting single incident angle gather from common midpoint gather, it is often strongly influenced by noise. Therefore, the partial angle stack gather with the higher *S*/*N* is generally used as an input data for EI inversion in practical applications. Thereby, there is a contradiction between EI equation and the gather applied in practical application, which may cause some errors and inevitably influence the subsequent elastic parameters inversion results. Especially in the target area of deep reservoir, seismic data have a small illumination angle, which make the extraction of single incident angle gather more difficult. Li et al. (2008) proposed a solution by rewriting the Connolly's EI equation into a new form, called stack impedance (SI) equation. Unlike the EI equation which is a function of a single incident angle, the SI equation is a function of the starting and ending angles of partial angle stack gather.

According to the idea of SI, the reflection coefficient of partial angle stack is an arithmetic average of all incident angles, as

$$\mathrm{SR}(\boldsymbol{\Phi}_0,\boldsymbol{\Phi})=\frac{\int_{\boldsymbol{\Phi}_0}^{\boldsymbol{\Phi}} R_{pp}(\theta)\mathrm{d}\theta}{\int_{\boldsymbol{\Phi}_0}^{\boldsymbol{\Phi}} \mathrm{d}\theta} \tag{19}$$

where $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Phi}$ the starting and ending angles of partial angle stack gathers.

Substituting Eq. (17) into Eq. (19) and according to the derivation of SI equation (Li et al. 2008), we can deduce Eq. (18) into the form of SI, as

$$\mathrm{SI}(\boldsymbol{\Phi}_0,\boldsymbol{\Phi})=\mathrm{SI}_0\left(\frac{f}{f_0}\right)^{m(\boldsymbol{\Phi}_0,\boldsymbol{\Phi})}\left(\frac{E}{E_0}\right)^{n(\boldsymbol{\Phi}_0,\boldsymbol{\Phi})}\left(\frac{\sigma}{\sigma_0}\right)^{p(\boldsymbol{\Phi}_0,\boldsymbol{\Phi})} \tag{20}$$

where SI_0 is the normalization factor and can be the mean value of the acoustic impedance of target layer, and the exponential terms $m(\boldsymbol{\Phi}_0,\boldsymbol{\Phi})=\left(\frac{1}{2}-\frac{\gamma_{\mathrm{dry}}^2}{2\gamma_{\mathrm{sat}}^2}\right)\frac{\sin(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)}{(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)\cos(\boldsymbol{\Phi})\cos(\boldsymbol{\Phi}_0)}$, $n(\boldsymbol{\Phi}_0,\boldsymbol{\Phi})=\frac{1}{2}\left(\frac{\gamma_{\mathrm{dry}}^2}{\gamma_{\mathrm{sat}}^2}-\frac{1}{1+2r}\right)\frac{\sin(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)}{(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)\cos(\boldsymbol{\Phi})\cos(\boldsymbol{\Phi}_0)}-\frac{2}{\gamma_{\mathrm{sat}}^2}\left[1-\frac{\cos(\boldsymbol{\Phi}+\boldsymbol{\Phi}_0)\sin(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)}{(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)}\right]+\frac{1}{1+2r}$, $p(\boldsymbol{\Phi}_0,\boldsymbol{\Phi})=\frac{2-\gamma_{\mathrm{sat}}^2}{6\gamma_{\mathrm{sat}}^2-8}\frac{\gamma_{\mathrm{dry}}^2(1+2r)-\gamma_{\mathrm{sat}}^2}{\gamma_{\mathrm{sat}}^2(1+2r)}\frac{\sin(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)}{(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)\cos(\boldsymbol{\Phi})\cos(\boldsymbol{\Phi}_0)}-\frac{8-4\gamma_{\mathrm{sat}}^2}{3\gamma_{\mathrm{sat}}^4-4\gamma_{\mathrm{sat}}^2}\left[1-\frac{\cos(\boldsymbol{\Phi}+\boldsymbol{\Phi}_0)\sin(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)}{(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)}\right]+\frac{4-2\gamma_{\mathrm{sat}}^2}{(2+4r)(3\gamma_{\mathrm{sat}}^2-4)}$.

When the starting angle is equal to the ending angle, Eq. (20) becomes Eq. (18), which means that the input seismic gather is a precise single angle stack gather.

After Eq. (20) is obtained, when the partial angle stack gathers, well logging data and horizon interpretation data are input, the SI inversion can be taken by using the conventional constrained sparse spike inversion (CSSI) method which is the same as EI inversion. The advantage of CSSI is that it can obtain broadband reflection coefficients (Latimer et al. 2000; Yuan et al. 2017; Ma et al. 2019). It has been proven that the SI equation is more suitable for partial angle stack data and the inversion result is more accurate (Li et al. 2008; Gui et al. 2014). Since there are three parameters needed to be inverted, at least three SI data with different angle stack ranges are required. It can be found that the exponential terms of SI equation $m(\boldsymbol{\Phi}_0, \boldsymbol{\Phi})$, $n(\boldsymbol{\Phi}_0, \boldsymbol{\Phi})$ and $p(\boldsymbol{\Phi}_0, \boldsymbol{\Phi})$ are only related to the range of stack angle, so when the starting and ending angles are determined, the exponential terms will not change with time t. For n sample points, logarithmic calculation is performed on both sides of Eq. (20)

$$\begin{bmatrix}\ln\frac{\mathrm{SI}(t_1,\boldsymbol{\Phi}_0,\boldsymbol{\Phi})}{\mathrm{SI}_0}\\ \ln\frac{\mathrm{SI}(t_2,\boldsymbol{\Phi}_0,\boldsymbol{\Phi})}{\mathrm{SI}_0}\\ \vdots\\ \ln\frac{\mathrm{SI}(t_n,\boldsymbol{\Phi}_0,\boldsymbol{\Phi})}{\mathrm{SI}_0}\end{bmatrix}=\begin{bmatrix}\ln\frac{f(t_1)}{f_0} & \ln\frac{E(t_1)}{E_0} & \ln\frac{\sigma(t_1)}{\sigma_0}\\ \ln\frac{f(t_2)}{f_0} & \ln\frac{E(t_2)}{E_0} & \ln\frac{\sigma(t_2)}{\sigma_0}\\ \vdots & \vdots & \vdots\\ \ln\frac{f(t_n)}{f_0} & \ln\frac{E(t_n)}{E_0} & \ln\frac{\sigma(t_n)}{\sigma_0}\end{bmatrix}\begin{bmatrix}m(\boldsymbol{\Phi}_0,\boldsymbol{\Phi})\\ n(\boldsymbol{\Phi}_0,\boldsymbol{\Phi})\\ p(\boldsymbol{\Phi}_0,\boldsymbol{\Phi})\end{bmatrix} \tag{21}$$

Table 1 Model parameters

Strata	V_p (m/s)	V_s (m/s)	ρ (kg/m^3)
Shale	2898	1290	2425
Gas sand	2857	1666	2275

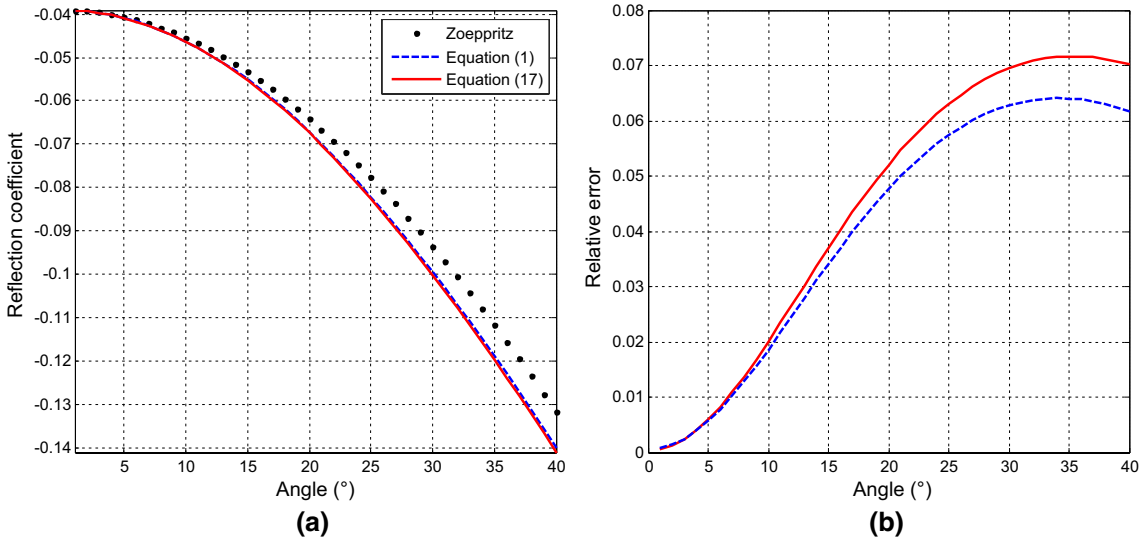


Fig. 1 Comparison of approximation accuracy. **a** Reflection coefficients and **b** relative errors

For Eq. (21), we can use well logging data and borehole side SI inversion curves to get the exponential terms $m(\Phi_0, \Phi)$, $n(\Phi_0, \Phi)$, $p(\Phi_0, \Phi)$ by the linear fitting. For three different angle stack ranges Φ_0–Φ_1, Φ_2–Φ_3 and Φ_4–Φ_5, nine exponential terms can be obtained in the same way. Then, for the sample point t_n, we have the following matrix equation

$$\begin{bmatrix} m(\Phi_0,\Phi_1) & n(\Phi_0,\Phi_1) & p(\Phi_0,\Phi_1) \\ m(\Phi_2,\Phi_3) & n(\Phi_2,\Phi_3) & p(\Phi_2,\Phi_3) \\ m(\Phi_4,\Phi_5) & n(\Phi_4,\Phi_5) & p(\Phi_4,\Phi_5) \end{bmatrix} \begin{bmatrix} \ln\frac{f(t_n)}{f_0} \\ \ln\frac{E(t_n)}{E_0} \\ \ln\frac{\sigma(t_n)}{\sigma_0} \end{bmatrix} = \begin{bmatrix} \ln\frac{\mathrm{SI}(t_n,\Phi_0,\Phi_1)}{\mathrm{SI}_0} \\ \ln\frac{\mathrm{SI}(t_n,\Phi_2,\Phi_3)}{SI_0} \\ \ln\frac{\mathrm{SI}(t_n,\Phi_4,\Phi_5)}{SI_0} \end{bmatrix} \tag{22}$$

Since nine exponential terms can be obtained by using Eq. (21), when three SI inversion data are input, the unknown elastic parameters f, E and σ at t_n will be easily obtained by solving the linear Eq. (22).

Examples

A two-level reference model is used to test the approximation accuracy of the new reflection coefficient equation. The elastic parameters of the model are shown in Table 1 (Goodway et al. 1997). The P-wave reflection coefficients at elastic interface are obtained by Eqs. (1), (17) and the exact Zoeppritz equation, respectively. Figure 1a shows the reflection coefficient curves, and Fig. 1b shows the relative error curves. From Fig. 1b, we can see that the curves calculated by Eqs. (1) and (17) almost coincide with each other and they are very close to the curve calculated by exact Zoeppritz equation. As shown in Fig. 1b, when the incident angle is below 30°, the relative errors between Eq. (17) and exact Zoeppritz equation are less than 7%. In

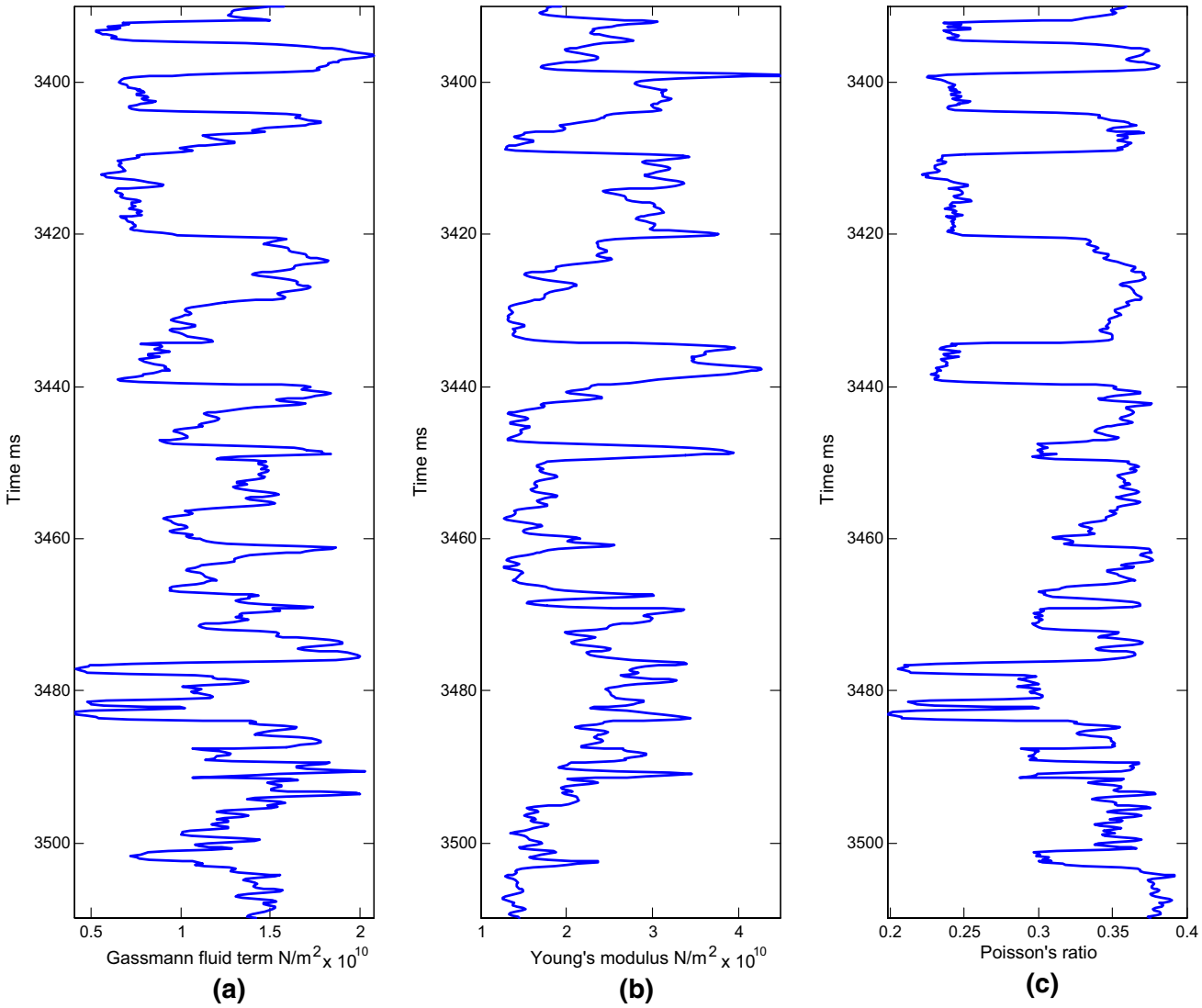


Fig. 2 Well log curves. **a** Gassmann fluid term, **b** Young's modulus and **c** Poisson's ratio

practical applications, the incident angle of deep seismic gather rarely exceeds 30°. Therefore, Eq. (17) is suitable for seismic inversion of real deep reservoir.

The real measured data are used as model test data, as shown in Fig. 2. Using the curves in Fig. 2, we synthesized three SI curves with angle ranges 0°–10°, 5°–15°, 10°–20° by Eq. (20), as shown in Fig. 3. In order to test the inversion accuracy of our method, the random Gaussian noise with a different intensity is added to the SI curves, with *S*/*N* of 5 and 2 (*S*/*N* is the ratio of the root mean square amplitude of the signal to that of noise), respectively, and the common two-step inversion and the one-step direct inversion are performed. The inversion results are shown in Figs. 4 and 5. By comparing inversion results with different *S*/*N*, we can find that the accuracy decreases with the decrease in *S*/*N*. When $S/N=5$, the inversion results of our method agree well with the true curves. Although the inversion results of common two-step inversion method are not as good as that of one-step direct inversion method, it can still be roughly consistent with the true curves. When $S/N=2$, the inversion results of our method still agree well with the true curves while the inversion results of common two-step inversion method have large deviation.

Our method has been applied to real prediction of deep tight sand reservoir. The 3D cross-plot of logging samples in this area is shown in Fig. 6. We can see that the

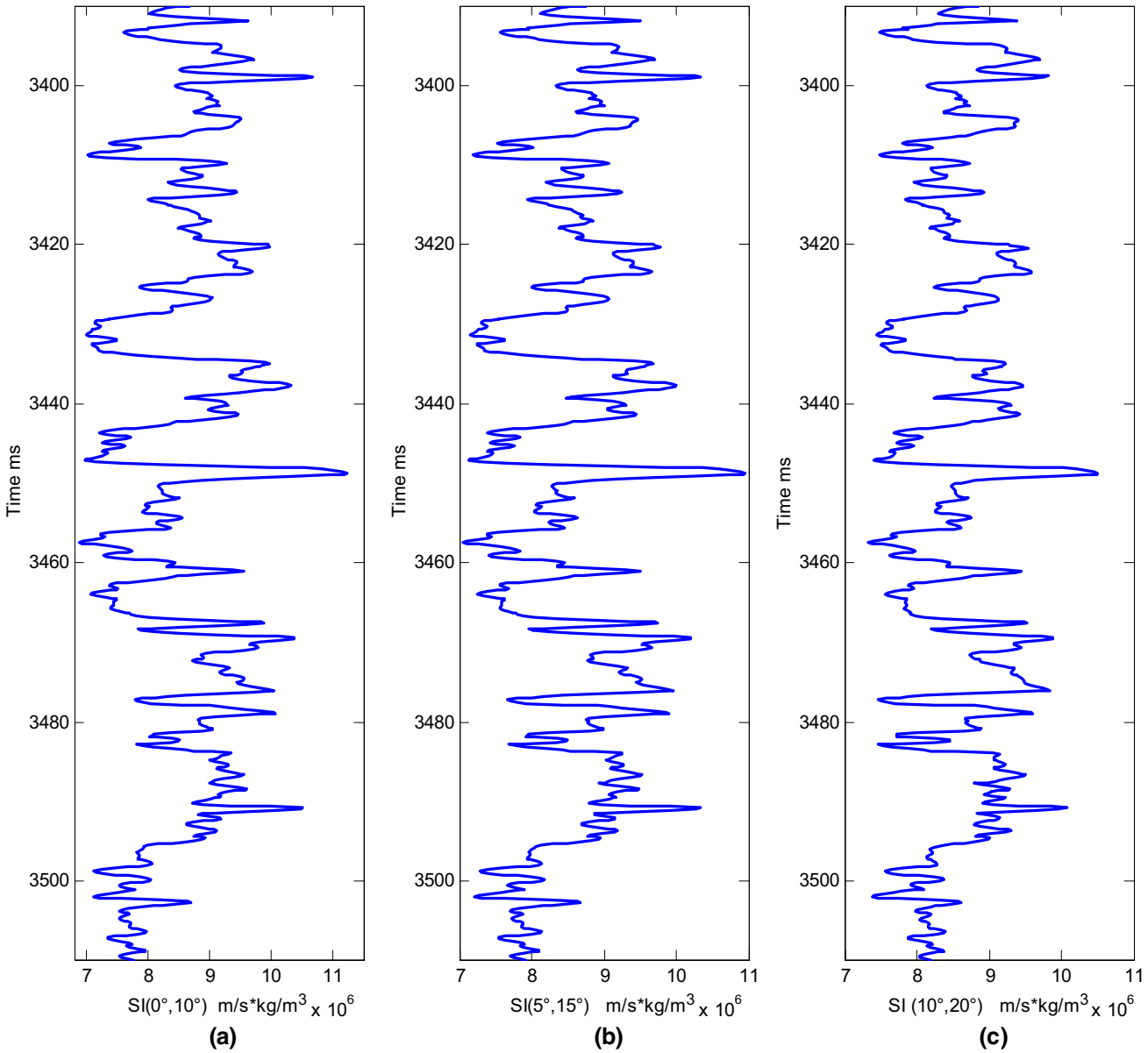


Fig. 3 Synthetic SI curves. **a** SI curves with angle range 0°–10°, **b** SI curves with angle range 5°–15° and **c** SI curves with angle range 10°–20°

sand reservoir has a relatively low Poisson's ratio and high Young's modulus value than that of shale, and the gas-bearing reservoir has a relatively low Gassmann fluid term value than that of water-bearing reservoir. Such characteristics indicate that the gas-bearing tight sand reservoir can be predicted by jointly using Poisson's ratio, Young's modulus and Gassmann fluid term.

In this study area, the maximum incident angle of the seismic data in target layer is around 20°, and can be divided into three partial angle stack gathers. The partial angle stack gather profiles of one through-well seismic line are shown in Fig. 7. The position of water-bearing and gas-bearing sand is circled by dashed ellipse and solid ellipse, respectively. The CSSI algorithm is adopted to the common two-step inversion and one-step direct inversion method, and the inversion results are shown in Figs. 8, 9 and 10. On the whole, the differences between the two are not very large, but there is certain degree of differences in details. Figure 11a–c shows the comparison of real logging curves and inversion results at the well location. We can see that the inversion results of one-step direct inversion method agree well with the real curves than the common two-step inversion method, especially in the regions of actual tight sand reservoirs.

In addition, for one-step direct inversion results, we noticed that the regions with relatively low Poisson's ratio value and high Young's modulus value are in good agreement with the regions of actual tight sand reservoirs. However, the differences between the regions of water-bearing and gas-bearing tight sand in Poisson's ratio and Young's modulus inversion profiles are not obvious, but there are

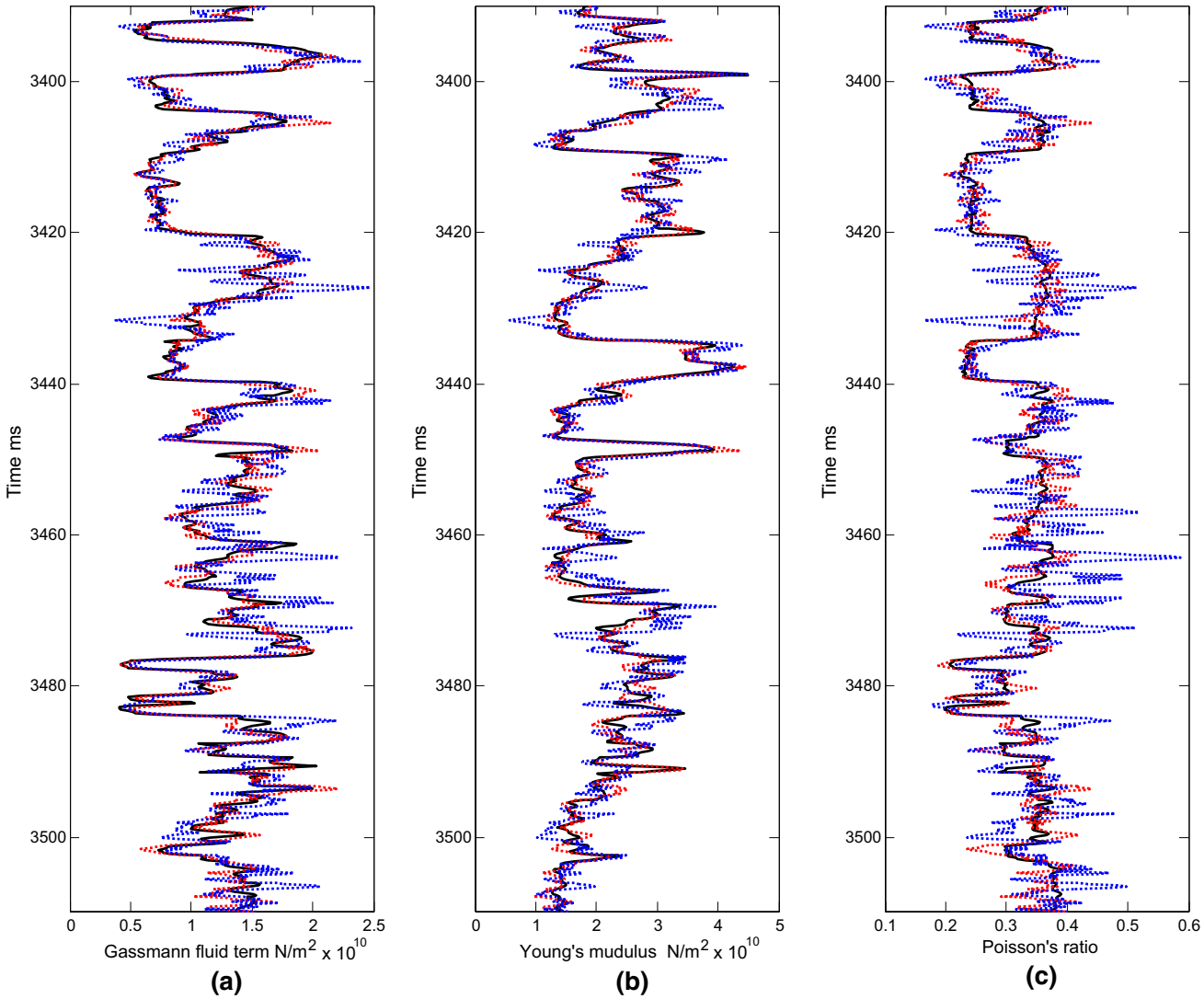


Fig. 4 Comparison of the common two-step inversion and one-step direct inversion methods with SNR = 5. **a** Gassmann fluid term, **b** Young's modulus and **c** Poisson's ratio. The black, blue and red curves represent the exact curves, inversion curves by common two-step method and inversion curves by one-step direct method, respectively

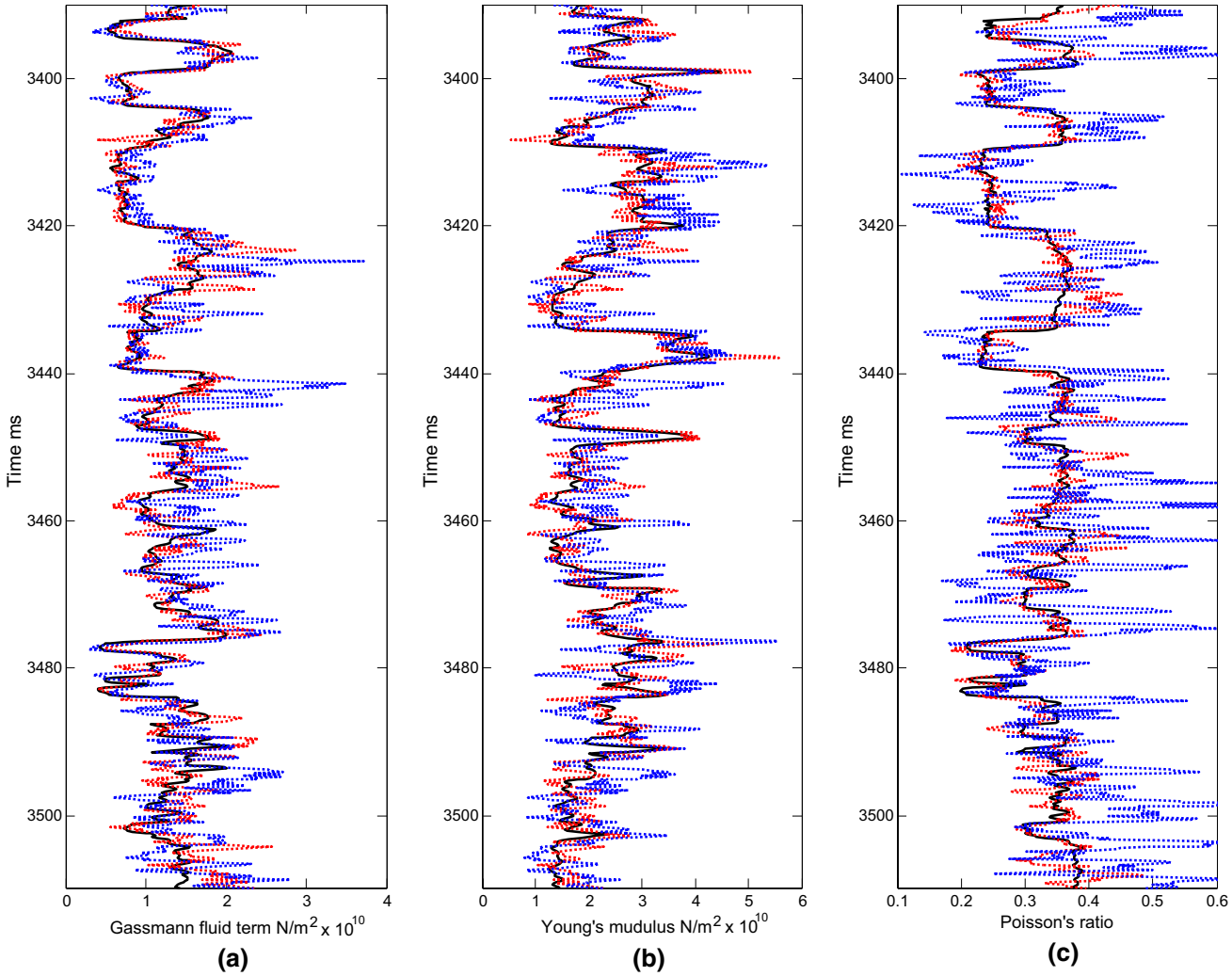


Fig. 5 Comparison of the common two-step inversion and one-step direct inversion methods with SNR = 2. **a** Gassmann fluid term, **b** Young's modulus and **c** Poisson's ratio. The black, blue and red curves represent the exact curves, inversion curves by common two-step method and inversion curves by one-step direct method, respectively

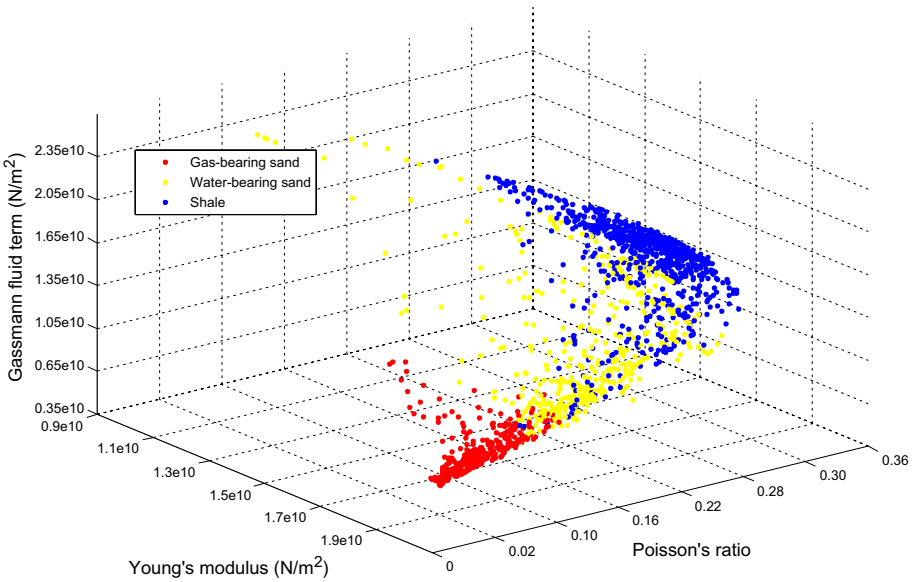


Fig. 6 Cross-plot of fluid term, Young's modulus and Poisson's ratio. The red, yellow and blue dots represent gas-bearing sand, water-bearing sand and shale, respectively

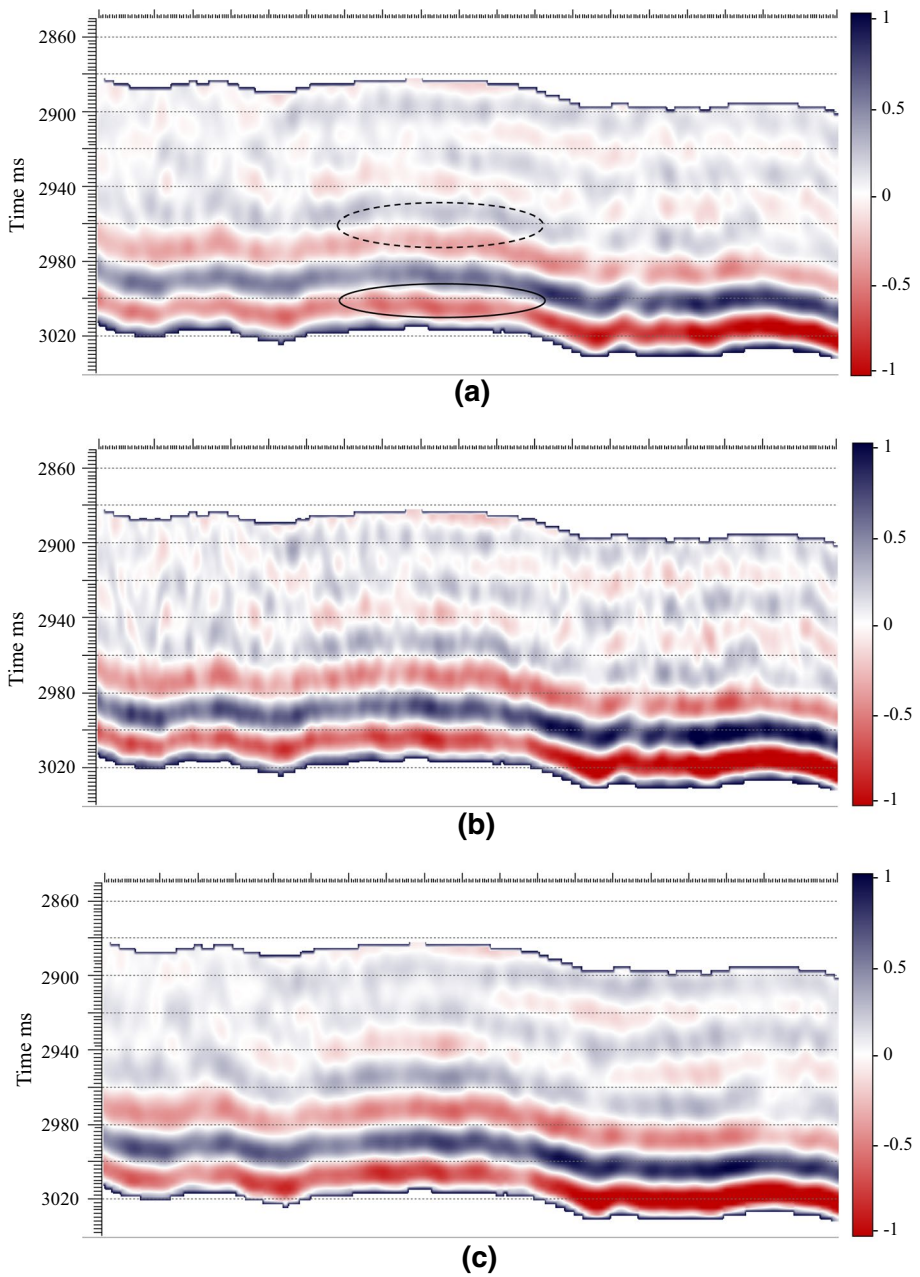


Fig. 7 Partial angle stack seismic profile. From **a**–**c**, respectively, for small to large partial angle stack seismic profile

very obvious differences in Gassmann fluid term inversion profile. The gas-bearing sand has obvious lower value of Gassmann fluid term. This phenomenon is consistent with the 3D cross-plot analysis result. Therefore, by jointly utilizing the Poisson's ratio, Young's modulus and Gassmann fluid term, we can not only predict the areas of deep reservoir development, but also can identify the fluid type in reservoirs.

Conclusions

A direct inversion method for sensitive elastic parameters of deep reservoirs is put forward throughout the study. On the basis of Russell's P-wave reflection coefficient equation, we derived a new reflection coefficient equation directly related to the Poisson's ratio, Young's modulus and Gassmann fluid term. The error analysis of reflection coefficient showed that the accuracy of the new equation is very close to the original equation at small incident angle. Based on the new reflection coefficient equation, we can obtain the SI equation by means of angle integration to solve the contradiction between the EI equation and the

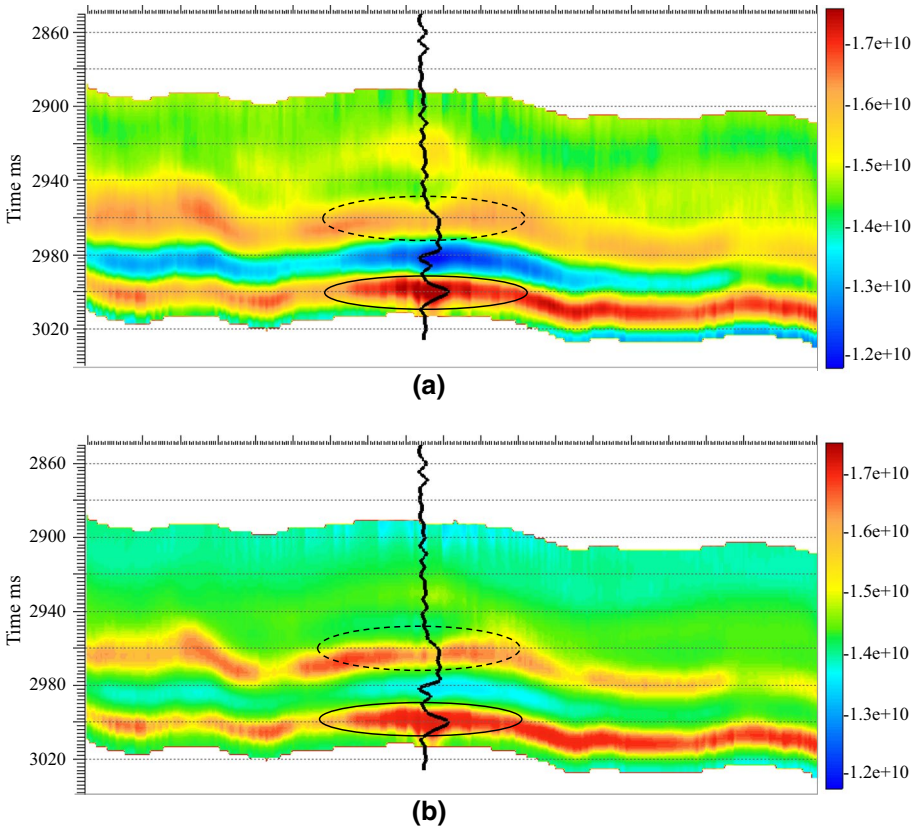


Fig. 8 Inversion results of Young's modulus. **a** Common two-step inversion and **b** one-step direct inversion

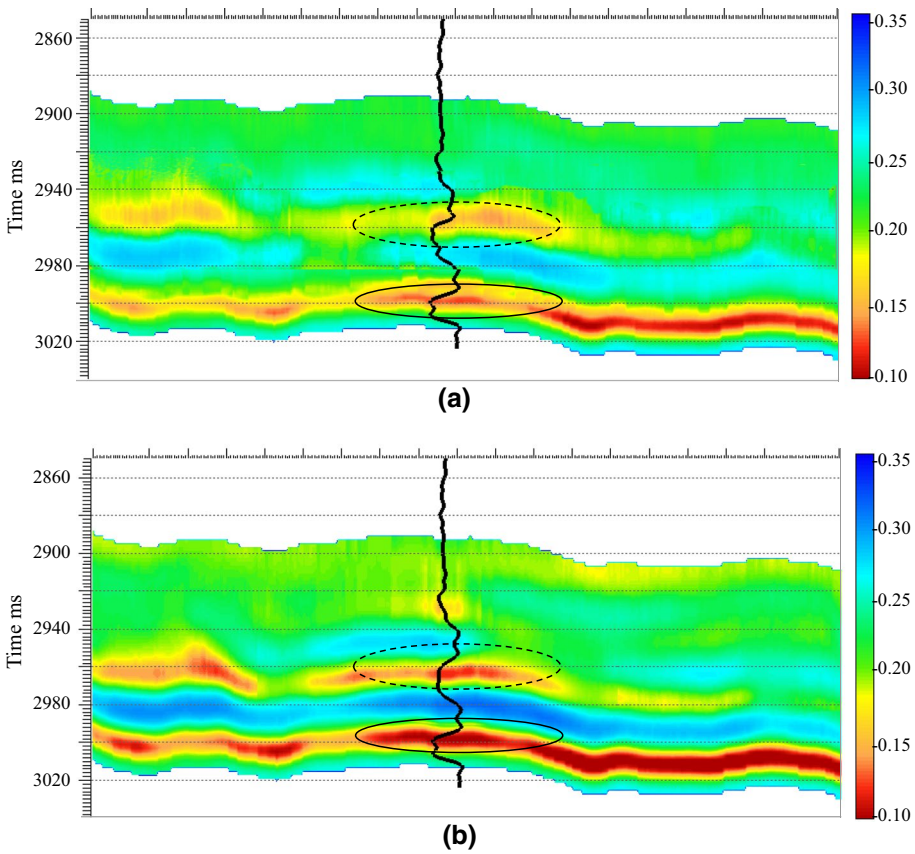


Fig. 9 Inversion results of Poisson's ratio. **a** Common two-step inversion and **b** one-step direct inversion

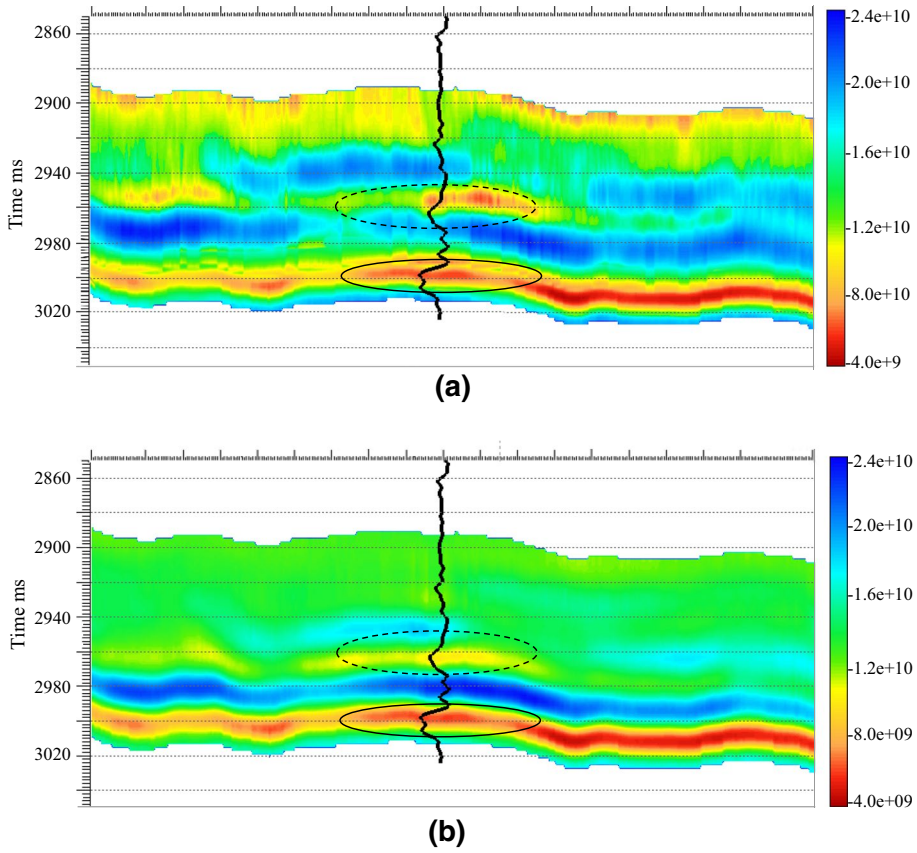


Fig. 10 Inversion results of Gassmann fluid term. **a** Common two-step inversion and **b** one-step direct inversion

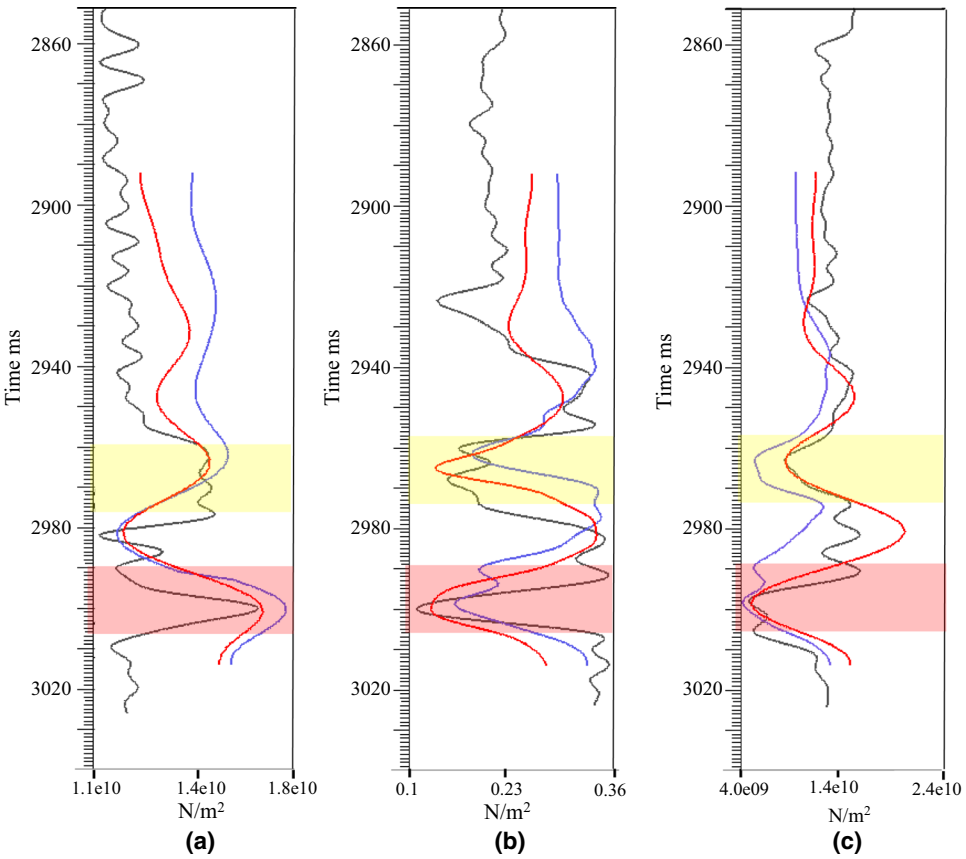


Fig. 11 Comparison of real logging curves and inversion results at the well location. **a** Young's modulus, **b** Poisson's ratio and **c** Gassmann fluid term. The red, blue and black curves represent the result of one-step direct inversion, common two-step inversion and real logging curves, respectively. The yellow and red translucent rectangle represents the area of water-bearing sand and gas-bearing sand, respectively

actual input seismic data with low *S/N*. After obtaining three SI inversion data with different angle stack ranges, we can directly get the Poisson's ratio, Young's modulus and Gassmann fluid term by solving a linear equation. Examples show that the approach is more suitable for deep reservoirs and can reduce the cumulative errors effectively compared with the common two-step inversion method.

Acknowledgements This work was financially supported by the Major Scientific Research Program of Petrochina Science and Technology Management Department "Comprehensive Seismic Prediction Software Development and Applications of Natural Gas" (2016B-0603), and the National Science and Technology Major Project (2016ZX05007-006).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Aki K, Richards PG (1980) Quantitative seismology: theory and methods, vol 1. WH Freeman & Co, San Francisco, pp 123–192

Bortfeld R (1961) Approximation to the reflection and transmission coefficients of plane longitudinal and transverse waves. Geophys Prospect 9(4):485–502. https://doi.org/10.1111/j.1365-2478.1961.tb01670.x

Castagna JP, Batzle ML, Eastwood RO (1985) Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks. Geophysics 50(4):571–581. https://doi.org/10.1190/1.1441933

Connolly P (1999) Elastic impedance. Lead Edge 18(4):438–452. https://doi.org/10.1190/1.1438307

Fatti JL, Smith GC, Vail PJ, Strauss PJ, Levitt PR (1994) Detection of gas in sandstone reservoirs using AVO analysis: a 3-D seismic case history using the Geostack technique. Geophysics 59(9):1362–1376. https://doi.org/10.1190/1.1443695

Feng H, Russell BH, Bancroft JC (2007) A comparison of hydrocarbon indicators derived from AVO analysis. Consort Res Elastic Wave Explor Seismol Res Rep 19:1–9. https://doi.org/10.1190/1.2792426

Goodway B, Chen T, Downton J (1997) Improved AVO fluid detection and lithology discrimination using Lamé petrophysical parameters; "λρ", "μρ", & "λ/μ fluid stack", from p and s inversions. SEG Technical Program Expanded Abstracts, pp 183–186. https://doi.org/10.1190/1.1885795

Gray D (1999) Bridging the gap: using AVO to detect changes in fundamental elastic constants. SEG Technical Program Expanded Abstracts, pp 852–855. https://doi.org/10.3997/2214-4609.201407877

Gui JY, Gao JH, Li SJ, Wang WL (2014) Using the stack impedance equation to invert elastic parameters. In: 76th EAGE conference and exhibition. https://doi.org/10.3997/2214-4609.20141045

Harris NB, Miskimins JL, Mnich CA (2011) Mechanical anisotropy in the woodford shale, Permian Basin: origin, magnitude, and scale. Geophysics 30(3):284–291. https://doi.org/10.1190/1.3567259

Hilterman F (2001) Seismic amplitude interpretation Distinguished instructor short course. Geophysical Development Corporation, Houston

Latimer RB, Davidson R, Riel PV (2000) An interpreter's guide to understanding and working with seismic-derived acoustic impedance. Lead Edge 19(3):242–256. https://doi.org/10.1190/1.1438580

Li H, Cui XF, Huang WF (2008) Stack impedance. SEG Technical Program Expanded Abstracts, pp 2021–2025. https://doi.org/10.1190/1.3059289

Liu XJ, Yin XY (2014) Direct estimation for fluid factor by two term AVO inversion for deep reservoir. SEG Technical Program Expanded Abstracts, pp 574–579. https://doi.org/10.1190/segam2014-0507.1

Ma M, Zhang R, Yuan SY (2019) Multichannel impedance inversion for nonstationary seismic data based on the modified alternating direction method of multipliers. Geophysics 84(1):A1–A6. https://doi.org/10.1190/geo2018-0319.1

Mallick S (2001) AVO and elastic impedance. Lead Edge 20(10):1094–1104. https://doi.org/10.1190/1.1487239

Nebojsa T, Nina G (2017) Well-log based rock physics template of the Vienna Basin and the underlying Calcereous Alps. Acta Geophys 65:441–451. https://doi.org/10.1007/s11600-017-0037-6

Russell BH, Gray D, Hampson DP (2011) Linearized AVO and poroelasticity. Geophysics 76(3):C19–C29. https://doi.org/10.1190/1.3555082

Sena A, Castillo G, Chesser K, Voisey S, Estrada J, Carcuz J, Carmona E, Hodgkins P (2011) Seismic reservoir characterization in resource shale plays: stress analysis and sweet spot discrimination. Lead Edge 30(7):758–764. https://doi.org/10.1190/1.3609090

Shi PD, Yuan SY, Wang TY, Wang YY, Liu T (2018) Fracture identification in a tight sandstone reservoir: a seismic anisotropy and automatic multisensitive attribute fusion framework. IEEE Geosci Remote Sens Lett 15(10):1525–1529. https://doi.org/10.1109/LGRS.2018.2853631

Shuey RT (1985) A simplification of the Zoeppritz equations. Geophysics 50(4):609–614. https://doi.org/10.1190/1.1441936

Smith GC, Gidlow PM (1987) Weighted stacking for rock property estimation and detection of gas. Geophys Prospect 35(5):993–1014. https://doi.org/10.1111/j.1365-2478.1987.tb00856.x

Wang SX, Yuan SY, Wang TY, Gao JH, Li SJ (2018) Three-dimensional geosteering coherence attributes for deep-formation discontinuity detection. Geophysics 83(6):O105–O113. https://doi.org/10.1190/geo2017-0642.1

Yuan SY, Wang SX, Ma M, Ji YZ, Den L (2017) Sparse Bayesian learning-based time-variant deconvolution. IEEE Trans Geosci Remote Sens 55(11):6182–6194. https://doi.org/10.1109/TGRS.2017.2722223

Yuan SY, Liu Y, Zhang Z, Luo CM, Wang SX (2019a) Prestack stochastic frequency-dependent velocity inversion with rock-physics constraints and statistical associated hydrocarbon attributes. IEEE Geosci Remote Sens Lett 16(1):140–144. https://doi.org/10.1109/LGRS.2018.2868831

Yuan SY, Wang SX, Luo YN, Wei WW, Wang GC (2019b) Impedance inversion by using the low-frequency full-waveform inversion result as a priori model. Geophysics. https://doi.org/10.1190/geo2017-0643.1

Zhang S, Yin X, Zhang F (2009) Fluid discrimination study from fluid elastic impedance (FEI). SEG Technical Program Expanded Abstracts, pp 2437–2441. https://doi.org/10.1190/1.3255350

Zhang GL, Hao CT, Yao C (2018) Analytical study of the reflection and transmission coefficient of the submarine interface. Acta Geophys 66:449–460. https://doi.org/10.1007/s11600-018-0153-y

Zoeppritz K (1919) On the reflection and penetration of seismic waves through unstable layers. Goettinger Nachr 1:66–84

Zong Z, Yin X, Wu G (2013) Elastic impedance parameterization and inversion with young's modulus and poisson's ratio. Geophysics 78(6):N35–N42. https://doi.org/10.1190/geo2012-0529.1

Acta Geophysica (2019) 67:1341–1348
https://doi.org/10.1007/s11600-019-00318-x

RESEARCH ARTICLE - APPLIED GEOPHYSICS

Multiples inversion imaging using a one-way propagation operator

Qiang Li[1] · Deli Wang[1]

Received: 25 March 2019 / Accepted: 10 June 2019 / Published online: 26 June 2019

Abstract
The one-way propagation operator in the frequency-space domain has the advantages of fast calculation speed and good adaptability to medium with lateral velocity variation. The full wavefield model constructed by the one-way propagation operator is iterative. As the number of iterations increases, the components of wavefield are more and more abundant. In the full wavefield model, the propagation and scattering processes are independent of each other. The former is determined by the propagation operator, while the latter is determined by the scattering operator. As each iteration increases, the wavefield component will increase by one order. As an inverse migration operator, the full wavefield model could feed back the imaging result to the data. By calculating the residual between the simulated data and the actual data, the reflectivity is updated. This is an inversion process. In this process, multiples will be imaged. In this way, the subsurface information contained in multiples is utilized and the imaging quality is greatly improved. The L_1-norm is used to constrain the imaging result, which further suppresses the artifacts and improves the imaging resolution. We have made some numerical examples in 2D case, explaining the principles and advantages of this methodology.

Keywords Modeling · Imaging · Inversion · Multiples · L_1-norm

Introduction

The traditional views regard multiples as interference information, which need to be removed or attenuated before imaging. There are many methods to remove multiples, such as the surface-related multiple elimination proposed by Verschuur (1990) and Verschuur et al. (1991, 1992). Another method estimating primaries by sparse inversion is also proposed by Van Groenestijn and Verschuur (2009). Other related research work to the removal of surface related multiple were also carried out by Yuan et al. (2018). However, multiples contain important information of the subsurface, which, if utilized, will greatly improve the imaging quality of the subsurface. The surface-related multiple can improve the illumination of the shallow and middle locations of the subsurface due to the smaller incident angle than primary. The internal multiple could contribute to the imaging of salt dome structure because it can reflect below the layer. The migration of surface-related multiple was achieved by Verschuur and Berkhout (2011a, b), Ning and Herrmann (2015) and Li et al. (2018). Seismic imaging with internal multiple was performed by Malcolm et al. (2009) and Fleury and Snieder (2012) and Wang et al. (2019). Another method (NLRTM) using internal multiple is implemented by Broggini et al. (2013), Wapenaar et al. (2013), which achieves imaging of internal multiple by estimating model perturbations. The Marchenko method is used to reconstruct the Green's function from the data, and then, the wavefields are reconstructed. But, it requires a dense source and detector arrangement. And this method can only be applied to non-free surface, that is, it cannot handle surface-related multiple. In these methods, surface-related multiple and internal multiple are imaged separately. The full wavefield migration (FWM) was proposed by Berkhout (2011, 2012, 2014b), which utilizes the full wavefield, including the multiples (surface-related multiple and internal multiple) in the migration. FWM is an inversion process. The residual between the simulated data and the actual data is minimized to update the imaging result. FWMod (2011, 2012, 2014a) (full wavefield model) works as a forward engine in the FWM, which simulates data using the imaging result. FWMod is also an iterative process. In FWMod, the one-way propagation operator and the two-way scattering operator are used to simulate the two-way wavefields, which means that the propagating process and the scattering process are decoupled from each other. The principle of

✉ Deli Wang
wangdeli@jlu.edu.cn

[1] Jilin University, Changchun, China

FWM has been demonstrated by Soni and Verschuur for VSP data (2014) and Davydenko and Verschuur (2016).

Since the one-way propagation operator is an approximation solution to the wave equation, artifacts will occur during the propagation process. In order to suppress this part of artifacts and improve the resolution of the subsurface imaging, different sparse constraint functions could be used. Soni and Verschuur (2014) used the Cauchy-norm to constrain the imaging result. Similarly, Davydenko and Verschuur (2016) also constrained the imaging result with the Cauchy-norm during the imaging process.

In this paper, a one-way propagation operator in the frequency-space domain is used to construct the full wavefield model. Then, the L_1-norm is used to constrain the imaging result during the internal multiple migration, improving the quality of the imaging.

Forward model

According to Berkhout's matrix description (1982), the monochromatic component wavefield recorded at any jth grid point on depth z_n emitted by the kth source on depth z_m can be expressed as $P_{j,k}(z_n, z_m)$. The wavefield on the whole depth z_n for a single shot can be expressed as a column vector $\vec{P}(z_n, z_m)$. In the case of multi-shots, the wavefield can be expressed as a matrix $\mathbf{P}(z_n, z_m)$. In the case of single shot, the wavefield on depth z_n can be divided into four types, $\vec{P}^{+}(z_n)$ for the downward (+) incident wavefield, $\vec{P}^{-}(z_n)$ for the upward (−) incident wavefield, $\vec{Q}^{+}(z_n)$ for the downward outgoing wavefield, and $\vec{Q}^{-}(z_n)$ for the upward outgoing wavefield (see Fig. 1).

The relationship between the incident and outgoing wavefield is defined by the scattering source $\delta\vec{S}(z_n)$:

$$\begin{aligned}\vec{\mathbf{Q}}^{+}(z_n) &= \vec{\mathbf{P}}^{+}(z_n) + \vec{\boldsymbol{\delta\mathbf{S}}}^{+}(z_n)\\ \vec{\mathbf{Q}}^{-}(z_n) &= \vec{\mathbf{P}}^{-}(z_n) + \vec{\boldsymbol{\delta\mathbf{S}}}^{-}(z_n).\end{aligned} \tag{1}$$

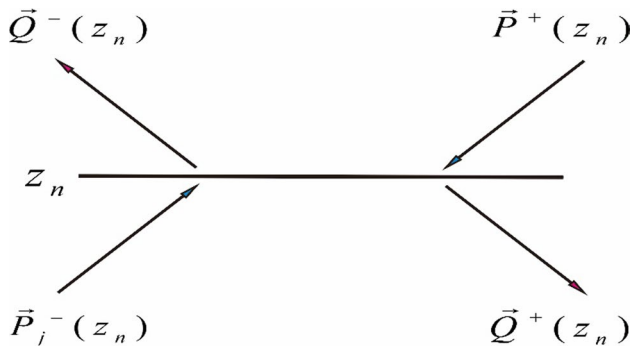


Fig. 1 Wavefields on depth z_n contain two incident wavefield and two outgoing wavefield. The incident wavefields include the upward wavefield $\vec{P}^{-}(z_n)$ and the downward wavefield $\vec{P}^{+}(z_n)$. And the outgoing wavefields include the upward wavefield $\vec{Q}^{-}(z_n)$ and the downward wavefield $\vec{Q}^{+}(z_n)$

$\vec{\boldsymbol{\delta\mathbf{S}}}^{+}(z_n)$ and $\vec{\boldsymbol{\delta\mathbf{S}}}^{-}(z_n)$ are the scattering sources generated by the indent wavefield on depth z_n:

$$\begin{aligned}\delta\vec{S}^{+}(z_n) &= \vec{P}^{-}(z_n)\mathbf{R}^{\cap}(z_n) + \vec{P}^{+}(z_n)\mathbf{R}^{\cup}(z_n)\\ \delta\vec{S}^{-}(z_n) &= \vec{P}^{+}(z_n)\mathbf{R}^{\cup}(z_n) + \vec{P}^{-}(z_n)\mathbf{R}^{\cap}(z_n),\end{aligned} \tag{2}$$

where the $\mathbf{R}^{\cup}$ and $\mathbf{R}^{\cap}$ are reflectivity matrices, representing the upper and lower reflectivity of the layer, respectively. In the acoustic case, we have $\mathbf{R}^{\cap}(z_n) = -\mathbf{R}^{\cup}(z_n)$.

When the wavefields are extrapolated from one depth to another, the outgoing wavefields of one depth will become the incident wavefield of another depth, as shown below (see Fig. 2):

$$\begin{aligned}\vec{P}^{+}(z_{n+1}) &= \mathbf{W}^{+}(z_{n+1}, z_n)\vec{Q}^{+}(z_n)\\ \vec{P}^{-}(z_{n-1}) &= \mathbf{W}^{-}(z_{n-1}, z_n)\vec{Q}^{-}(z_n),\end{aligned} \tag{3}$$

where $\mathbf{W}^{\pm}$ are the one-way propagation operators. Here, the weighted least-squares propagation operator with a smooth function is used. (Thorbecke et al. 2004).

Given a source, the downward record and the upward record for depth z_n can be obtained after a few iterations:

$$\begin{aligned}\vec{P}^{+}(z_n) &= \sum_{m<n}\mathbf{W}^{+}(z_n, z_m)[\vec{S}^{+}(z_m) + \delta\vec{S}^{+}(z_m)]\\ \vec{P}^{-}(z_n) &= \sum_{m>n}\mathbf{W}^{-}(z_n, z_m)\delta\vec{S}^{-}(z_m),\end{aligned} \tag{4}$$

where the $\vec{S}^{+}$ represents the physical source.

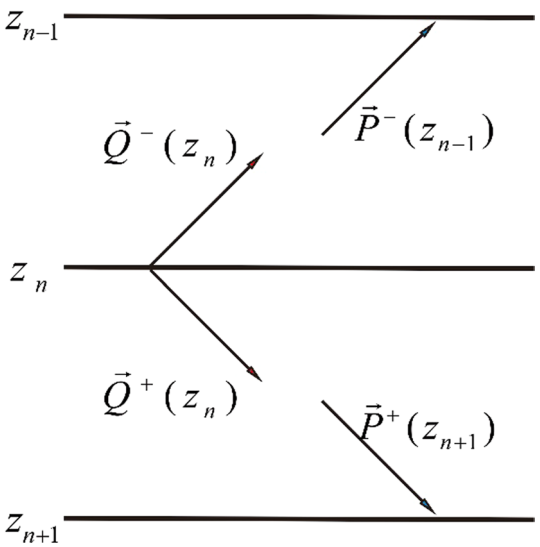


Fig. 2 Relationship of the wavefield between the depth: The outgoing wavefields $\vec{\mathbf{Q}}^{+}(z_n)$ and $\vec{\mathbf{Q}}^{-}(z_n)$ of depth z_n will become the incident wavefield $\vec{P}^{+}(z_{n+1})$ and $\vec{P}^{-}(z_{n-1})$ of the adjacent depth

Therefore, in the first iteration, the downward wavefields only include the wavefield propagated directly from the source (without transmission effect). And the upward wavefields only include the reflected wavefield caused by the previous downward wavefield. If we do not consider multiples, the incident wavefields can be expressed as follows:

$$\begin{aligned}\vec{P}^{+}(z_n) &= \mathbf{W}^{+}(z_n, z_0)\vec{S}^{+}(z_0)\\ \vec{P}^{-}(z_n) &= \sum_{m>n} \mathbf{W}^{-}(z_n, z_m)\vec{P}^{+}(z_m)\mathbf{R}^{\cup}(z_m).\end{aligned} \tag{5}$$

$\vec{S}^{+}(z_0)$ means that the source is only located at the surface.

In the iterative process, using Eq. (4), the sources (physical source and scattering source) are propagated down and up to every depth, and the incident wavefields of each depth need to be preserved. Then, the incident wavefields are substituted into Eq. (2) to update the scattering source. The updated scattering source continues to be propagated to update the incident wavefield for each depth. Finally, a seismic record is obtained at the receiver side. The wavefields are extrapolated in the depth direction by Eq. (3).

It can be seen that the main contribution of multiples comes from the scattering source ($\delta\vec{S}$). The generation and update of the scattering source depend on the incident wavefield obtained in the previous iteration. The updated scattering source in turn further updates the incident wavefield of each depth. This is iterative. Therefore, as the number of iterations increases, the components of wavefield become more and more abundant.

Imaging

The purpose of the inversion imaging is to use the imaging result to interpret the migration data. The imaging result is used in the full wavefield model to interpret the observed full wavefield. The objective function in the sense of L_2-norm can be expressed as:

$$J = \sum_{\omega} \left\|\mathbf{P}_{\text{obs}} - \mathbf{P}_{\text{mod}}\right\|_2^2, \tag{6}$$

where $\|\|_2^2$ represents the L_2-norm, $\mathbf{P}_{\text{obs}}$ is the observed data, and $\vec{P}_{\text{obs}}$ for the single shot. $\mathbf{P}_{\text{mod}}$ is the modeled data obtained by the forward model and $\vec{P}_{\text{mod}}$ for the single shot. And we need to calculate all the frequencies of wavefield. As a forward simulation operator, the full wavefield model could feed back the imaging result to the data. The reflectivity is updated by calculating the residual between the simulated data and the actual data. As each iteration increases, an additional one-order wavefield will be imaged. This is an inversion process. During this process, multiples will be imaged. In this way, the subsurface information contained in multiples is utilized, thereby greatly improving the quality of the imaging.

We can solve the objective function by the gradient descent method. Pseudo-code is shown in Table 1.

The gradient of the objective function, that is, the gradient of the reflectivity above the layer, is:

$$\Delta\mathbf{R}^{\cup}(z_n) = \sum_{\omega} -2 \sum_{k} \Delta\vec{P}^{-}(z_n)\left[\vec{P}^{+}(z_n)\right]^{H}, \tag{7}$$

where k is the shot number and H is the conjugate transpose. $\Delta\vec{P}^{-}(z_n)$ is the residual wavefield propagated back to z_n. The reflectivity is a non-diagonal matrix. Here, we only consider the case of angle-independent, which means that only the diagonal elements of the matrix are not zero. There is only one value for each grid point in the subsurface. The case of angle-independent for the gradient can be expressed as:

$$\Delta\mathbf{R}^{\cup}(z_n) = \sum_{\omega} -2 \sum_{k} \text{diag}\left(\Delta\vec{P}^{-}(z_n)\left[\vec{P}^{+}(z_n)\right]^{H}\right). \tag{8}$$

L_1-norm sparse constraint

Since the one-way propagation operator is an approximation solution to the wave equation, artifacts will occur during the propagation process. To further remove the artifacts remained in the imaging result and improve the resolution of the imaging result, we use the L_1-norm to constrain the imaging result. The objective function with sparse constraint can be expressed as:

$$J = \sum_{\omega} \left\|\mathbf{P}_{\text{obs}} - \mathbf{P}_{\text{mod}}\right\|_2^2 + \lambda f(\mathbf{R}), \tag{9}$$

where $f(\mathbf{R})$ is a penalty function acting on the imaging result, which can help improve the resolution of the imaging.

Table 1 Pseudo-code for the inversion imaging using the gradient descent method

PSEUDO-CODE

set $\mathbf{R}_0 = 0$;
for i=0, $i < i_{max}$
  for $\omega_{min} \le \omega \le \omega_{max}$
    forward model: calculate the residual $\Delta\mathbf{P}$;
    imaging: calculate the gradient $\Delta\mathbf{R}$;
  end ω
  search the α;
  update the $\mathbf{R}_i$=$\mathbf{R}_{i-1} + \alpha\Delta\mathbf{R}$;
end i
output $\mathbf{R}$

λ is a hyper-parameter used to balance the loss term and the regulation term. The $f(\mathbf{R})$ for the L_1-norm can be expressed as:

$$f(\mathbf{R}) = \sum_n \sum_j \left| R_{jj} \right|, \tag{10}$$

where R_{jj} is a sample of the reflectivity matrix (a diagonal element from matrix $\mathbf{R}(z_n)$). The gradient of the objective function with L_1-norm constraint can be expressed as:

$$\Delta \mathbf{R}^{\cup}(z_n) = \sum_{\omega} -2 \sum_k \operatorname{diag}\left(\Delta \vec{P}^{-}(z_n) \left[\vec{P}^{+}(z_n) \right]^H \right) + \lambda f'(\mathbf{R}^{\cup}). \tag{11}$$

The second term can be written as:

$$f'(\mathbf{R}) = \begin{cases} \operatorname{sgn}(R_{jj}), & \text{if } R_{jj} \neq 0; \\ \{h : |h| \leq 1, h \in \mathbb{R}\}, & \text{otherwise.} \end{cases} \tag{12}$$

where the sgn is a symbolic function:

$$\operatorname{sgn}(R_{jj}) = \begin{cases} 1 & R_{jj} > 0; \\ 0, & R_{jj} = 0; \\ -1, & R_{jj} < 0; \end{cases} \tag{13}$$

Numerical example

We use the 2D velocity model shown in Fig. 3 to illustrate the iterative property of the full wavefield model. The density is constant, and the receiver spacing is 5 m covering the whole surface. Figure 4 shows a single-shot record for different iterations. As mentioned above, in the first iteration, only the physical source is propagated, so the wavefields only include primary. In the second iteration, the components of wavefield are increased by one order due to the generation of the scattering source. In the third iteration, the second-order wavefields are added to the record and so on. This is an iterative process. As the number of iterations increases, the components of wavefield become more and more abundant.

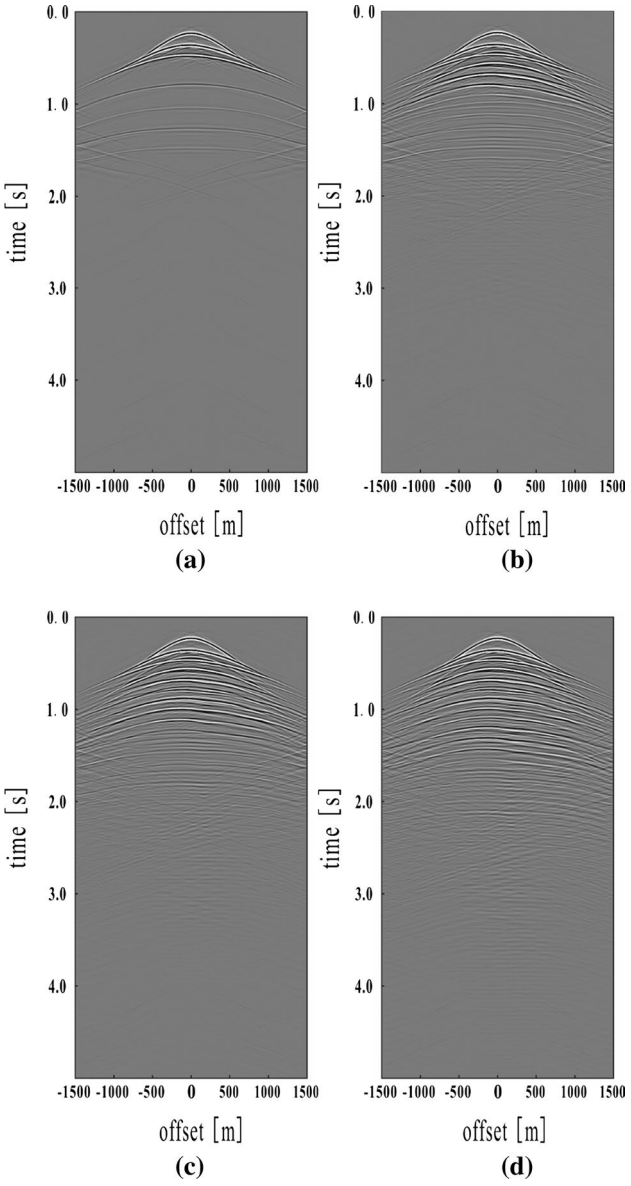


Fig. 4 **a** A shot gather with one iteration; **b** a shot gather with two iterations; **c** a shot gather with three iterations; and **d** a shot gather with four iterations. As the number of iterations increases, the record becomes more and more abundant

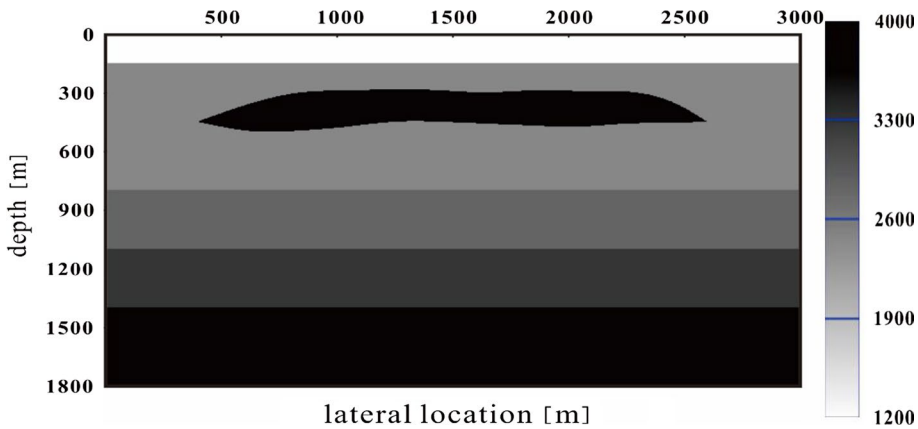


Fig. 3 Velocity model (m/s)

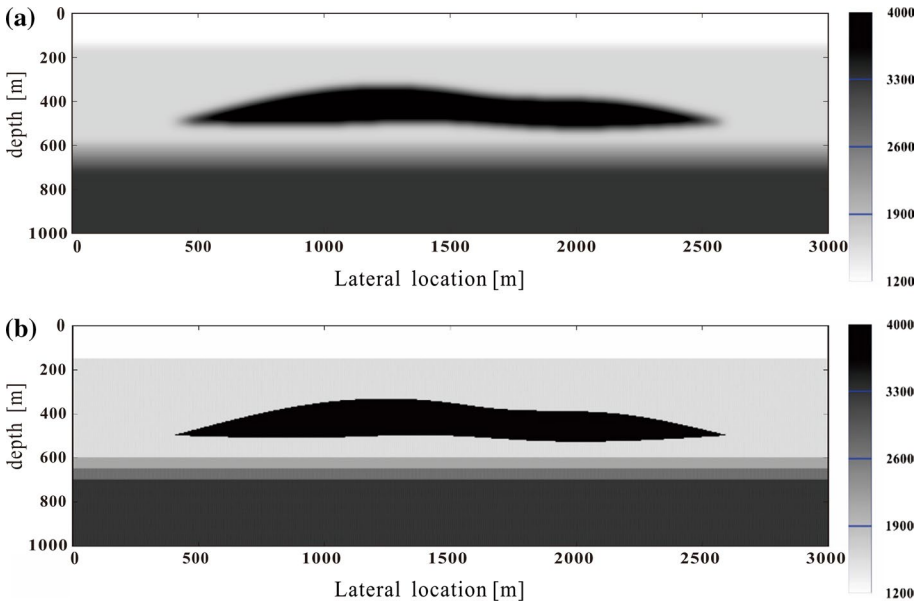


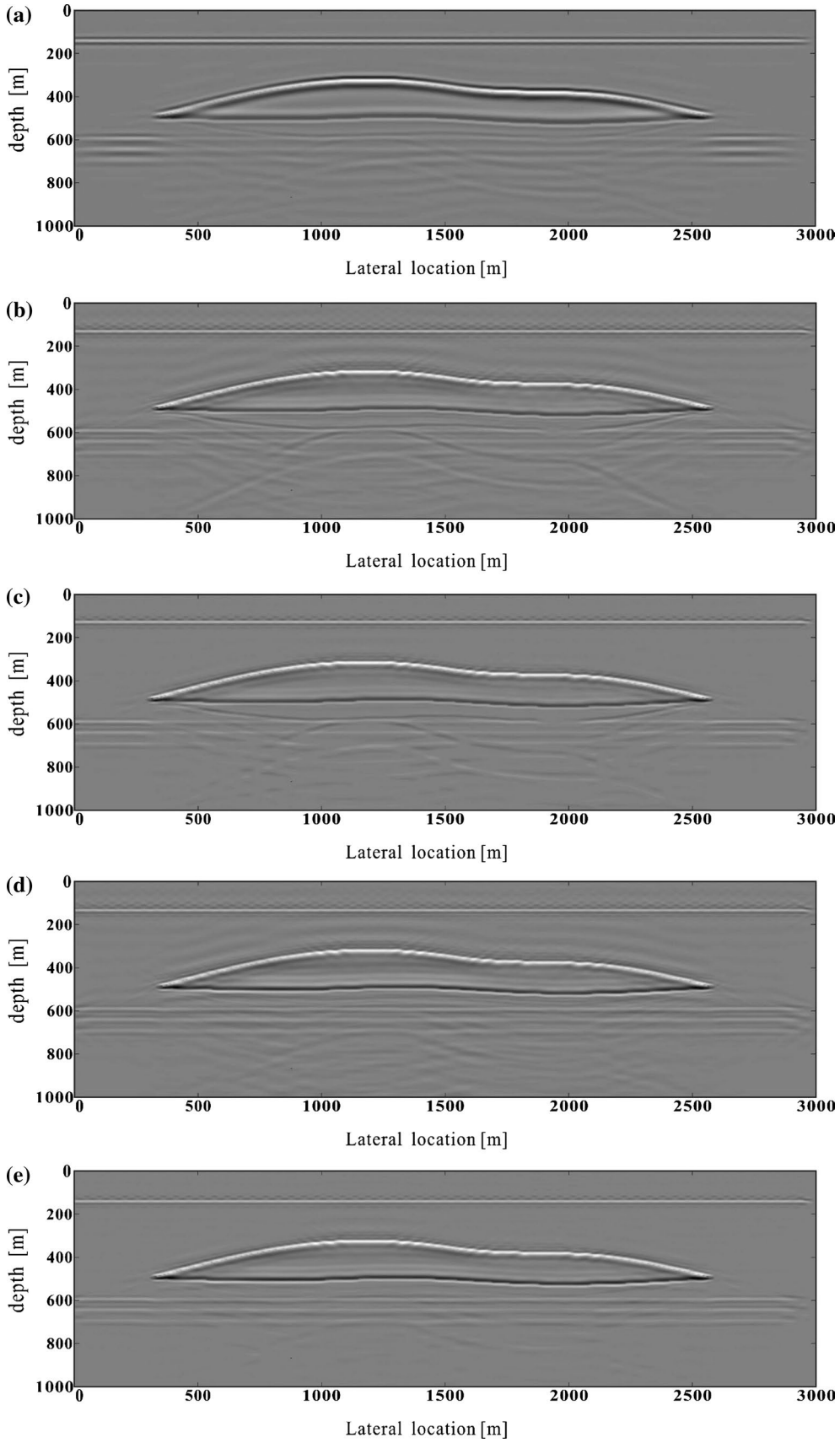
Fig. 5 **a** Velocity model (m/s); **b** density model (kg/m^3)

Figure 5 shows the velocity and density model we used to simulate the migration data (as the actual data here) by the acoustic FD method. The data contain primary and internal multiple. The source spacing is 50 m, and the receiver spacing is 5 m. We implement the migration using the velocity model in Fig. 5 which is assumed to vary smoothly in the spatial direction. Figure 6 shows the imaging result for different cases. Figure 6a shows the conventional imaging result which only considers the primary in the migration. We can see that the cross talk generated by the internal multiple appears in the imaging result, and the cross talk causes the layers below the salt dome to be deformed, resulting in an unclear image. Figure 6b, d shows the results of inversion imaging, and the imaging quality of both is improved. The difference is that Fig. 6d considers the internal multiple in the migration process, that is, the forward model used for the migration process of Fig. 6d is the full wavefield model. Figure 6b shows the least-squares migration of primary based on the one-way propagation operator. In Fig. 6b, the cross talk still exists. It can be clearly seen that since the internal multiple is considered in the imaging process, the cross talk generated by the internal multiple is suppressed in Fig. 6d. Figure 6c, e shows the results of Fig. 6b, d with the L_1-norm sparse constraint. The artifacts in the imaging results are further removed, and the resolution of the imaging results is better.

Discussion

The one-way propagation operator in the frequency-space domain has the advantages of fast calculation speed and good adaptability to medium with lateral velocity variation. The one-way propagation operator used here has an angular limitation. In order to achieve large-angle propagation of the wavefield, the spatial length of the propagation operator needs to be longer, which will lead to an increase in computation. Furthermore, if the spatial length of the operator is too long, it will not be well adapted to medium with lateral velocity variation.

In this paper, only the angle-independent reflectivity is considered, which means that the reflectivity matrix is a diagonal matrix and only the elements on the diagonal are not zero. It is equivalent to the case of normal incidence and does not contain information that varies with angle. If we want to perform AVO inversion, then we need to consider the angle-dependent reflectivity.


(a)
(b)
(c)
(d)
(e)
depth [m]
0
200
400
600
800
1000
0
500
1000
1500
2000
2500
3000
Lateral location [m]

◂**Fig. 6** **a** Imaging result after one iteration for primary; **b** primary inversion imaging result after 20 iterations; **c** primary inversion imaging result after 20 iterations with L_1-norm constraint; **d** multiples inversion imaging result after 20 iterations; and **e** multiples inversion imaging result after 20 iterations with L_1-norm constraint

Conclusion

One-way propagation operators can be used to solve the problem of multiples imaging. The source (including the scattering source) is propagated forward and is then cross-correlated with the back propagation residual wavefield. Multiples imaging is an iterative inversion process. Each iteration will have a higher-order wavefield to be imaged, so the multiples will also be imaged. The residual is obtained by comparing the simulated data modeled by the full wavefield model with the actual data, and the imaging result is continuously corrected in this way.

The full wavefield model is also an iterative process. As the number of iterations increases, the order of the wavefield increases accordingly. In this process, the propagation operator and the scattering operator are independent of each other. The propagation operator is only responsible for propagation, and the scattering operator is only responsible for scattering. The full wavefield model updates the subsurface wavefield by updating the scattering source. The updated scattering source in turn further updates the subsurface wavefield. In the first iteration, the simulated wavefields contain only primary (without transmission effect). In the second iteration, the subsurface scattering source is updated, and the first-order multiples as well as the transmission effect are added to the simulated wavefield. The generation of multiples is mainly due to the presence of the scattering source.

Because the one-way propagation operator is an approximation solution of the wave equation, there must be some propagation artifacts in the propagation process. The L_1-norm is used in the imaging process to constrain the imaging result in order to further remove the artifacts and improve the resolution of the imaging.

Acknowledgements The author would like to thank the supervisor Professor Deli Wang for his guidance and help as well as many useful discussions and suggestions from Professor Eric Verschuur. The author would also like to thank the National Natural Science Foundation (No. 41374108) and the National Major Science and Technology Project of China (2016ZX05026-002-003) for their sponsorship.

References

Berkhout AJ (1982) Seismic migration: imaging of acoustic energy by wave field extrapolation a: theoretical aspects, 2nd edn. Elsevier, Amsterdam. ISBN 0-444-41904-7

Berkhout AJ (2011) Combining full wavefield migration and full waveform inversion. In: 81st SEG meeting, San Antonio, USA, expanded abstracts, pp 3321–3325

Berkhout AJ (2012) Combining full wavefield migration and full waveform inversion: a glance into the future of seismic imaging. Geophysics 77(2):S43–S50

Berkhout AJ (2014a) Review paper: an outlook on the future of seismic imaging, part II: full-wavefield migration. Geophys Prospect 62(5):931–949

Berkhout AJ (2014b) Review paper: an outlook on the future of seismic imaging, part I: forward and reverse modelling. Geophys Prospect 62(5):911–930

Broggini F, Snieder R, Wapenaar K (2013) Data-driven Green's function retrieval and imaging with multidimensional deconvolution: numerical examples for reflection data with internal multiples. In: 83rd SEG meeting, Houston, USA, pp 4156–4161

Davydenko M, Verschuur DJ (2016) Full-wavefield migration: using surface and internal multiples in imaging. Geophys Prospect 65(1):7–21

Fleury C, Snieder R (2012) Increasing illumination and sensitivity of reverse-time migration with internal multiples. In: 74th EAGE annual conference and exhibition, Copenhagen, Denmark, expanded abstracts, pp 4–7

Li Q, Wang D, Zhou J (2018) One-way propagation operator to simulate the two-way wavefield and application on surface-related multiples image. In: 80th EAGE conference and exhibition

Malcolm AE, Ursin BR, de Hoop MV (2009) Seismic imaging and illumination with internal multiples. Geophys J Int 176(3):847–864

Ning T, Herrmann FJ (2015) Fast least-squares imaging with surface-related multiples: application to a North Sea data set. Lead Edge 34(7):788–794

Soni AK, Verschuur DJ (2014) Full-wavefield migration of vertical seismic profiling data: using all multiples to extend the illumination area. Geophys Prospect 62(4):740–759

Thorbecke JW, Wapenaar K, Swinnen G (2004) Design of one-way wavefield extrapolation operators, using smooth functions in WLSQ optimization. Geophysics 69(4):1037–1045

Van Groenestijn GJ, Verschuur DJ (2009) Estimating primaries by sparse inversion and application to near-offset data reconstruction. Geophysics 74(3):A23–A28

Verschuur DJ (1990) Surface-related multiple elimination in the presence of near-surface anomalies. SEG Technical Program Expanded Abstracts, vol 9, no 1, p 1622

Verschuur DJ, Berkhout AJ (2011a) Seismic migration of blended shot records with surface-related multiple scattering. Geophysics 76(1):A7–A13

Verschuur DJ, Berkhout AJ (2011b) Seismic migration of blended shot records with surface-related multiple scattering. Geophysics 76(1):A7–A13

Verschuur DJ, Berkhout AJ, Wapenaar CPA (1991) Surface-related multiple elimination: application on real data. SEG Technical Program Expanded Abstracts, pp 1476–1479

Verschuur DJ, Berkhout AJ, Wapenaar CPA (1992) Adaptive surface-related multiple elimination. Geophysics 57(9):1166–1177

Wang G, Alkhalifah T, Guo Q, Wang S (2019) Frequency domain generalized internal multiple imaging with energy norm. In: 81st EAGE meeting

Wapenaar K, Broggini F, Slob E, Snieder R (2013) Three dimensional single-sided Marchenko inverse scattering, data-driven focusing, Green's function retrieval, and their mutual relations. Phys Rev Lett 110:084301

Yuan SY, Wang SX, Yuan FF, Liu Y (2018) The influence of errors in the source wavelet on inversion-based surface-related multiple attenuation. Geophys Prospect 66:55–73

Acta Geophysica (2019) 67:1349–1363
https://doi.org/10.1007/s11600-019-00320-3

RESEARCH ARTICLE - APPLIED GEOPHYSICS

Multichannel antileakage least-squares spectral analysis for seismic data regularization beyond aliasing

Ebrahim Ghaderpour[1]

Received: 26 February 2019 / Accepted: 26 June 2019 / Published online: 2 July 2019

Abstract

The antileakage least-squares spectral analysis is a new method of regularizing irregularly spaced data series. This method mitigates the spectral leakages in the least-squares spectrum caused by non-orthogonality of the sinusoidal basis functions on irregularly spaced series, and it is robust when data series are wide-sense stationary. An appropriate windowing technique can be applied to adapt this method to non-stationary data series. When data series present mild aliasing, this method can effectively regularize the data series; however, additional information or assumption is needed when the data series is coarsely sampled. In this paper, we show how to incorporate the spatial gradients of the data series into the method to regularize data series presenting severe aliasing and show its robust performance on synthetic and marine seismic data examples.

Keywords Antileakage spectrum · Gradient · Least-squares spectral analysis · Regularization · Seismic data

Introduction

Regularization, a typical spectral interpolation and/or extrapolation, is a crucial problem in seismology. Marine seismic data sets are usually irregularly sampled along spatial directions because of cable feathering, editing bad traces, economy, etc. They are usually more coarsely sampled along the crossline direction than the inline direction as only a limited number of streamers can be towed. Regularly sampled seismic data are required for various purposes including wave equation migration, seismic inversion, amplitude versus azimuth or offset analyses, and surface-related multiple elimination (Weglein et al. 1997; Dragoset et al. 2010).

In seismic signal processing, a wavenumber is the number of cycles per unit distance, and a (cyclic) frequency is the number of cycles per unit time. Aliasing is an effect that causes different signals to be indistinguishable when sampled. For instance, assume that a sinusoid is sampled at equally spaced intervals. If another sinusoid has the same amplitude and phase as the original but its wavenumber differs from the first one by a multiple of the sampling rate, then these two sinusoids will have exactly the same samples over the equally spaced intervals (Craymer 1998; Ghaderpour 2018). Since seismic data are usually well sampled in time, regularization is usually performed on samples along the spatial directions, referred to data series (Ghaderpour et al. 2018).

There is a vast number of regularization methods, addressing practical issues in seismology, such as the prediction error filters (Spitz 1991; Crawley 2000; Fomel 2002; Wang 2002; Liu and Chen 2017), projection onto convex sets (Abma and Kabir 2006; Gao et al. 2012; Yang et al. 2012; Wang et al. 2016), minimum weighted norm interpolation (Liu and Sacchi 2004), compressive sensing (Wang et al. 2011), damped rank-reduction (Chen et al. 2016), nonlinear shaping regularization (Fomel 2007; Chen et al. 2015), antileakage Fourier transform (Xu et al. 2005, 2010), arbitrarily sampled Fourier transform (Guo et al. 2015), antileakage least-squares spectral analysis (Ghaderpour et al. 2018), interpolation by matching pursuit and its generalizations (Vassallo et al. 2010; Özbek et al. 2010a, b, 2012).

The antileakage least-squares spectral analysis (ALLSSA) is an iterative method based on the least-squares spectral analysis (LSSA) that uses a preselected set of wavenumbers to accurately estimate the statistically significant spectral peaks in the spectrum (Vaníček 1969; Pagiatakis 1999; Ghaderpour et al. 2018). After simultaneously suppressing several significant spectral peaks, an iteration process will be performed to estimate the previously estimated

✉ Ebrahim Ghaderpour
ebrahim.ghaderpour@ucalgary.ca

[1] University of Calgary, 2500 University Dr NW, Calgary, AB T2N 1N4, Canada

wavenumbers more accurately, mitigating the spectral leakages in the spectrum and reducing the computational cost. The ALLSSA simultaneously considers the covariance matrix associated with data series and the constituents of known forms, such as datum shifts, trends, sinusoids of known wavenumbers. Ghaderpour (2018, Chapters 3 and 4) showed that the ALLSSA performs more efficiently than the state-of-the-art methods, such as the antileakage Fourier transform, arbitrary sampled Fourier transform, and interpolation by matching pursuit (IMAP).

When data series presents mild aliasing in the spectrum, the ALLSSA can regularize the aliased data series effectively. However, certain assumptions, such as linearity of seismic events (Schonewille et al. 2009), or additional data sets, such as spatial gradients of the seismic data (Vassallo et al. 2010), are further needed to aid the regularization beyond aliasing.

In a truly multicomponent marine seismic data acquisition, pressure measurements are complemented by particle velocity measurements. The particle velocity measurements can increase the effective Nyquist wavenumber by a factor of two or three, depending on how they are used (Robertsson et al. 2008). The equation of motion states that the particle acceleration vector $\mathbf{a}$ (measured by an accelerometer) is proportional to the gradient of pressure P by equation $\nabla P = -\rho\mathbf{a}$, where ρ is the density of the medium. In particular, in the crossline direction (denoted by x here), $P_x = -\rho\dot{V}_x$, where the dot above V_x denotes the temporal derivative. Since particle motion can easily be converted into a pressure gradient by using the equation of motion, such streamers would enable acquisition of both pressure and the gradient of pressure simultaneously (Vassallo et al. 2010; Özbek et al. 2010b).

The multichannel interpolation by matching pursuit (MIMAP) is based on the IMAP that uses the multicomponent seismic measurements to reconstruct the seismic wavefield at any desired crossline position between towed streamers (Vassallo et al. 2010). The generalized matching pursuit (GMP) is based on the MIMAP (three-component generalization of the MIMAP) that iteratively reconstructs the signal as a combination of optimal basis functions (Özbek et al. 2010b). For coarsely sampled data with low signal-to-noise ratio, the GMP results can be improved by considering soft priors (Özbek et al. 2010a, 2012).

In this paper, a similar method, namely multichannel antileakage least-squares spectral analysis (MALLSSA), is introduced that incorporates the spatial gradients of seismic data (if available) into the ALLSSA to regularize the data series beyond aliasing. The MALLSSA can estimate the wavenumbers of data series more accurately compared to the MIMAP, resulting in less number of iterations and better and faster regularization results. The MALLSSA simultaneously considers the covariance matrices associated with the pressure and gradient data series as well as the constituents of known forms, such as datum shifts, trends, and sinusoids of known wavenumbers. The robust performance of the MALLSSA is shown on synthetic data series, synthetic seismic data, and marine seismic data sets. The MALLSSA can be seen as a more robust generalization of the MIMAP.

Methods

Multichannel interpolation by matching pursuit

In this section, the MIMAP is briefly described in matrix form (Vassallo et al. 2010). Suppose that $\mathbf{\Omega}$ is a set of wavenumbers. In this paper, we choose $\mathbf{\Omega} = \{1, 2, \ldots, \eta - 1\}$, where η is the Nyquist wavenumber of the desired regular series. Let $\mathbf{f} = [f(x_\ell)]$ be a column vector of n samples ($\ell = 1, \ldots, n$), and $\mathbf{f_x} = [f_x(x_\ell)]$ be its spatial gradient, where the x_ℓ's may be irregularly spaced. Note that $\mathbf{f}$ and $\mathbf{f_x}$ do not need to have the same size; however, their sizes are assumed to be the same in this contribution. For each wavenumber $\omega_k \in \mathbf{\Omega}$, let $\mathbf{\Phi}_k$ and $\mathbf{\Phi}_{k_x}$ be design matrices of orders $n \times 2$ defined as

$$\mathbf{\Phi}_k = \left[\cos(2\pi\omega_k x_\ell),\ \sin(2\pi\omega_k x_\ell)\right], \tag{1}$$

$$\mathbf{\Phi}_{k_x} = \left[-2\pi\omega_k \sin(2\pi\omega_k x_\ell),\ 2\pi\omega_k \cos(2\pi\omega_k x_\ell)\right]. \tag{2}$$

Also, let $\mathbf{c}_k = [a_k\ b_k]^{\mathrm{T}}$, where T is the transpose, and a_k and b_k are the coefficients of the cosine and sine basis functions being estimated, respectively. For each wavenumber $\omega_k \in \mathbf{\Omega}$, the MIMAP minimizes the following cost function with respect to $\mathbf{c}_k$

$$\Psi_k(\mathbf{c}_k) = (\mathbf{f} - \mathbf{\Phi}_k\,\mathbf{c}_k)^{\mathrm{T}}(\mathbf{f} - \mathbf{\Phi}_k\,\mathbf{c}_k) + \lambda(\mathbf{f_x} - \mathbf{\Phi}_{k_x}\,\mathbf{c}_k)^{\mathrm{T}}(\mathbf{f_x} - \mathbf{\Phi}_{k_x}\,\mathbf{c}_k), \tag{3}$$

where T is the transpose, and so it estimates $\mathbf{c}_k$ as

$$\hat{\mathbf{c}}_k = \left(\mathbf{\Phi}_k^{\mathrm{T}}\mathbf{\Phi}_k + \lambda\,\mathbf{\Phi}_{k_x}^{\mathrm{T}}\mathbf{\Phi}_{k_x}\right)^{-1}\left(\mathbf{\Phi}_k^{\mathrm{T}}\mathbf{f} + \lambda\,\mathbf{\Phi}_{k_x}^{\mathrm{T}}\mathbf{f_x}\right). \tag{4}$$

The derivation of Eq. (4) is shown in 'Appendix' in a more general case (i.e., for the MALLSSA). The weighting parameter λ in Eq. (3) adjusts the relative contributions of data and gradient residuals to the cost function. The proper selection of λ should consider the expected energy difference between the two signals and the signal-to-noise ratio of the respective multicomponent measurements (Vassallo et al. 2010). When $\lambda = 0$, the gradient measurements are discarded, and the MIMAP degenerates to the IMAP.

After substituting $\hat{\mathbf{c}}_k$ in Eq. (3), the optimization problem becomes estimating ω_k that maximizes

$$s_\lambda(\omega_k) = \mathbf{f}^{\mathrm{T}}\mathbf{\Phi}_k\hat{\mathbf{c}}_k + \lambda\,\mathbf{f_x}^{\mathrm{T}}\mathbf{\Phi}_{k_x}\hat{\mathbf{c}}_k, \tag{5}$$

corresponding to the extension of the Lomb spectrum to the case of multicomponent reconstruction (Vassallo et al. 2010, Appendix A). After estimating ω_k and $\hat{\mathbf{c}}_k$, the MIMAP obtains the residual data series ($\mathbf{f} - \mathbf{\Phi}_k\, \hat{\mathbf{c}}_k$) and its residual gradient ($\mathbf{f_x} - \mathbf{\Phi}_{k_\mathbf{x}}\, \hat{\mathbf{c}}_k$) and treats them as the new input data series and gradient; then, it performs the entire process on these residual series. When the energy of the overall residual is sufficiently reduced, the iterations are terminated, and the estimated wavenumbers and sinusoidal coefficients (ω_k's and $\hat{\mathbf{c}}_k$'s) will be used to reconstruct the data series on any desired equally spaced series.

In the IMAP and MIMAP, the wavenumbers are estimated one at a time (out-of-context), ignoring the correlations between the sinusoids of different wavenumbers (Ghaderpour et al. 2018). Thus, the estimated wavenumbers and sinusoidal coefficients are not very accurate, resulting in large number of iterations and less accuracy compared to when these correlations are considered.

Multichannel antileakage least-squares spectral analysis

In this section, it is shown how the correlations between the sinusoidal basis functions can be considered simultaneously with the constituents of known forms and covariance matrices associated with the data series. Suppose that $\mathbf{\Omega}$, $\mathbf{f}$, and $\mathbf{f_x}$ are the same as the ones in the previous section, and $\mathbf{C_f}$ and $\mathbf{C_{f_x}}$ are the (regular and Hermitian) covariance matrices associated with $\mathbf{f}$ and $\mathbf{f_x}$, respectively, given by

$$\mathbf{C_f} = \begin{bmatrix} \sigma^2_{f(x_1)} & \sigma_{f(x_1)f(x_2)} & \cdots & \sigma_{f(x_1)f(x_n)} \\ \sigma_{f(x_2)f(x_1)} & \sigma^2_{f(x_2)} & \cdots & \sigma_{f(x_2)f(x_n)} \\ \vdots & \vdots & & \vdots \\ \sigma_{f(x_n)f(x_1)} & \sigma_{f(x_n)f(x_2)} & \cdots & \sigma^2_{f(x_n)} \end{bmatrix},$$

$$\mathbf{C_{f_x}} = \begin{bmatrix} \sigma^2_{f_x(x_1)} & \sigma_{f_x(x_1)f_x(x_2)} & \cdots & \sigma_{f_x(x_1)f_x(x_n)} \\ \sigma_{f_x(x_2)f_x(x_1)} & \sigma^2_{f_x(x_2)} & \cdots & \sigma_{f_x(x_2)f_x(x_n)} \\ \vdots & \vdots & & \vdots \\ \sigma_{f_x(x_n)f_x(x_1)} & \sigma_{f_x(x_n)f_x(x_2)} & \cdots & \sigma^2_{f_x(x_n)} \end{bmatrix},$$

where $\sigma^2_{f(x_\ell)}$ and $\sigma^2_{f_x(x_\ell)}$ are the variances of random variables $f(x_\ell)$ and $f_x(x_\ell)$, respectively, and $\sigma_{f(x_u)f(x_v)}$ is the covariance between two random variables $f(x_u)$ and $f(x_v)$, and $\sigma_{f_x(x_u)f_x(x_v)}$ is the covariance between two random variables $f_x(x_u)$ and $f_x(x_v)$. Let $\mathbf{P} = \mathbf{C_f}^{-1}$ and $\mathbf{P_x} = \mathbf{C}^{-1}_{\mathbf{f_x}}$ that can identify parameter λ more rigorously in the MIMAP. In many practical applications, weight matrices $\mathbf{P}$ and $\mathbf{P_x}$ are approximately diagonal matrices (the correlations between the data points are negligible), so one may treat them as vectors for computational efficiency (Ghaderpour et al. 2018). In a marine environment, the particle velocity measurements are usually noisier than pressure measurements, especially in the low frequencies (Özbek et al. 2010b), and so depending on the seismic signal energy distribution and the signal-to-noise ratio, the diagonal entries of $\mathbf{P_x}$ are usually much smaller than the ones in $\mathbf{P}$.

For each wavenumber $\omega_k \in \mathbf{\Omega}$, the MALLSSA minimizes the following cost function with respect to $\overline{\mathbf{c}}$

$$\Psi_k(\overline{\mathbf{c}}) = \left(\mathbf{f} - \overline{\mathbf{\Phi}}\,\overline{\mathbf{c}}\right)^{\mathrm{T}}\mathbf{P}\left(\mathbf{f} - \overline{\mathbf{\Phi}}\,\overline{\mathbf{c}}\right) + \left(\mathbf{f_x} - \overline{\mathbf{\Phi}}_\mathbf{x}\,\overline{\mathbf{c}}\right)^{\mathrm{T}}\mathbf{P_x}\left(\mathbf{f_x} - \overline{\mathbf{\Phi}}_\mathbf{x}\,\overline{\mathbf{c}}\right), \tag{6}$$

where T is the (conjugate) transpose, and $\overline{\mathbf{\Phi}}$ and $\overline{\mathbf{\Phi}}_\mathbf{x}$ are the design matrices of orders $n \times (q+2)$ defined as

$$\overline{\mathbf{\Phi}} = \left[\underline{\mathbf{\Phi}},\ \mathbf{\Phi}_k\right],\quad \overline{\mathbf{\Phi}}_\mathbf{x} = \left[\underline{\mathbf{\Phi}}_\mathbf{x},\ \mathbf{\Phi}_{k_\mathbf{x}}\right], \tag{7}$$

such that $\underline{\mathbf{\Phi}} = \left[\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_q\right]$ contains the constituents of known forms, $\underline{\mathbf{\Phi}}_\mathbf{x} = \left[\mathbf{\Phi}_{1_\mathbf{x}}, \ldots, \mathbf{\Phi}_{q_\mathbf{x}}\right]$ contains their gradients, and $\mathbf{\Phi}_k$ and $\mathbf{\Phi}_{k_\mathbf{x}}$ are given by Eqs. (1) and (2), respectively. The constituents of known forms can be the column vector of all ones [$\mathbf{1}$] and/or the position column vector [$\mathbf{x}$] to explicitly account for any linear trend. If there is any prior information on some existing constituents, such as sinusoids of particular wavenumbers and/or any specific wavelets, then one may consider them in $\underline{\mathbf{\Phi}}$ as well as their gradients in $\underline{\mathbf{\Phi}}_\mathbf{x}$ in advance of the analysis. In our examples, sinusoids of particular wavenumbers will be considered as the constituents of known forms in an iterative manner automatically, and we also consider the column vector of all ones (see the last paragraph in this section). Accounting for these constituents results in more accurate estimation of actual constituents in the data series and thus better regularization.

Minimizing the cost function in Eq. (6) with respect to $\overline{\mathbf{c}}$, it is shown in 'Appendix' that

$$\hat{\overline{\mathbf{c}}} = \begin{bmatrix} \mathbf{v} \\ \hat{\mathbf{c}}_k \end{bmatrix} = \left(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}} + \overline{\mathbf{\Phi}}^{\mathrm{T}}_\mathbf{x}\mathbf{P_x}\overline{\mathbf{\Phi}}_\mathbf{x}\right)^{-1}\left(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\mathbf{f} + \overline{\mathbf{\Phi}}^{\mathrm{T}}_\mathbf{x}\mathbf{P_x}\mathbf{f_x}\right), \tag{8}$$

where $\mathbf{v}$ is the estimated coefficients of the constituent of known forms, and $\hat{\mathbf{c}}_k$ is the estimated sinusoidal coefficients corresponding to wavenumber ω_k. Now in the process of estimating the wavenumbers, at each step, a wavenumber $\omega_k \in \mathbf{\Omega}$ will be selected that maximizes the following normalized ratio:

$$s_{\mathrm{multi}}(\omega_k) = \frac{\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}\mathbf{\Phi}_k\hat{\mathbf{c}}_k + \hat{\mathbf{g}}^{\mathrm{T}}_\mathbf{x}\mathbf{P_x}\mathbf{\Phi}_{k_\mathbf{x}}\hat{\mathbf{c}}_k}{\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{g}} + \hat{\mathbf{g}}^{\mathrm{T}}_\mathbf{x}\mathbf{P_x}\hat{\mathbf{g}}_\mathbf{x}}, \tag{9}$$

where $\hat{\mathbf{g}} = \mathbf{f} - \underline{\mathbf{\Phi}}\,\hat{\underline{\mathbf{c}}}$ and $\hat{\mathbf{g}}_\mathbf{x} = \mathbf{f_x} - \underline{\mathbf{\Phi}}_\mathbf{x}\,\hat{\underline{\mathbf{c}}}$, and

$$\hat{\underline{\mathbf{c}}} = \left(\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\underline{\mathbf{\Phi}} + \underline{\mathbf{\Phi}}^{\mathrm{T}}_\mathbf{x}\mathbf{P_x}\underline{\mathbf{\Phi}}_\mathbf{x}\right)^{-1}\left(\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\mathbf{f} + \underline{\mathbf{\Phi}}^{\mathrm{T}}_\mathbf{x}\mathbf{P_x}\mathbf{f_x}\right). \tag{10}$$

Note that Eq. (10) is obtained by minimizing $\Psi(\underline{\mathbf{c}}) = (\mathbf{f} - \underline{\mathbf{\Phi}}\,\underline{\mathbf{c}})^{\mathrm{T}}\mathbf{P}(\mathbf{f} - \underline{\mathbf{\Phi}}\,\underline{\mathbf{c}}) + (\mathbf{f}_{\mathbf{x}} - \underline{\mathbf{\Phi}}_{\mathbf{x}}\,\underline{\mathbf{c}})^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}(\mathbf{f}_{\mathbf{x}} - \underline{\mathbf{\Phi}}_{\mathbf{x}}\,\underline{\mathbf{c}})$ with respect to $\underline{\mathbf{c}}$ in the same way as Eq. (8) is derived ("Appendix").

Next, a wavenumber will be chosen in a small neighborhood that maximizes Eq. (9), and then the sine and cosine basis functions of this wavenumber and their derivatives will be added to the design matrices $\underline{\mathbf{\Phi}}$ and $\underline{\mathbf{\Phi}}_{\mathbf{x}}$ in Eq. (7), respectively (as additional constituents of known forms). In practice, the number of columns in $\underline{\mathbf{\Phi}}$ and $\underline{\mathbf{\Phi}}_{\mathbf{x}}$ is small, and so the algorithm is computationally efficient, especially when the size of data series is small. One may obtain $\hat{\mathbf{c}}_k$ in Eq. (8) more efficiently using similar methods used in Ghaderpour (2018, Appendix B) or via similar techniques used in the fast Fourier transform (FFT).

The sinusoids and their derivatives may be removed from the design matrices to be estimated (re-optimized) more accurately in the next step like the ALLSSA. This process continues until $s_{\mathrm{multi}}(\omega_k)$ given by Eq. (9) is no longer statistically significant at a certain confidence level ("Appendix"), so $\hat{\underline{\mathbf{c}}}$ in Eq. (10) will comprise the estimated coefficients of the final constituents in the first round of iterations, and it will produce a spectrum called the multichannel antileakage least-squares spectrum (MALLSS). When the data series is wide-sense stationary, one round of iterations will provide sufficient wavenumbers to regularize the data series. However, one may repeat the entire process on the new residuals in the next round of iterations and continue until the L2 norm of residuals goes below a threshold (the residuals become random noise).

The weight matrices $\mathbf{P}$ (associated with $\mathbf{f}$) and $\mathbf{P}_{\mathbf{x}}$ (associated with $\mathbf{f}_{\mathbf{x}}$) balance the relative importance between the data series and its gradient (the noisier the gradient is, the less it will be considered). After each round of iterations, these matrices may be updated using the covariance law (Vaníček and Krakiwsky 1986; Ghaderpour 2018). Note that when $\mathbf{P}_{\mathbf{x}} = \mathbf{0}$, the MALLSS degenerates to the ALLSSA spectrum, following the beta distribution (Ghaderpour et al. 2018). As mentioned above, a few of the constituents of known forms can be preselected, such as the column vector of all ones or the distance vector. Other basis functions can be considered, such as the sinusoids, whose wavenumbers will be estimated in an iterative manner using the same partitioning of the wavenumbers described in Ghaderpour et al. (2018).

When a data series is coarsely sampled, several sinusoids of various wavenumbers may identically fit the constituents of the data series; however, the coefficients of their derivatives are different, resulting in selecting the correct wavenumbers when minimizing Eq. (6). In other words, the derivative measurements can alter the sampling criteria (Vassallo et al. 2010). Note that the MALLSSA is slightly different from the MIMAP in that the wavenumbers are simultaneously being estimated (re-optimized) in an iterative manner, resulting in higher accuracy and a smaller number of iterations. If one does not consider $\underline{\mathbf{\Phi}}$ and $\underline{\mathbf{\Phi}}_{\mathbf{x}}$ in Eq. (7), containing the constituents of known forms, and ignores the weight matrices $\mathbf{P}$ and $\mathbf{P}_{\mathbf{x}}$ (simply replace them by parameter λ), then the MALLSSA is in fact the MIMAP.

In seismic data regularization, it is customary to transform each trace from the time domain to the frequency domain using the FFT. Then for each frequency, generate a data series whose data points, located at the trace locations, are the Fourier coefficients of that frequency (a temporal frequency slice) (Spitz 1991; Abma and Claerbout 1995; Xu et al. 2005; Vassallo et al. 2010). In the seismic data examples in this contribution, the frequency slices (data series) for both pressure and gradient data are obtained, and then the MALLSSA is applied to regularize the frequency slices. Then, the regularized pressure series from the $f-x$ domain are transformed back to the $t-x$ domain using the inverse FFT. The real and imaginary parts of a frequency slice are simultaneously regularized in the MALLSSA. An alternative approach is to regularize the real and imaginary parts of a frequency slice independently, so a different set of wavenumbers may be estimated for each part to regularize the frequency slice (some information may be lost, however).

The MALLSSA like the MIMAP or almost any the regularization method in the presence of aliasing has the potential overlap of two or more spectral replicas at the same wavenumber. The perfect overlap of aliased events happens only in the ideal case of perfectly regular sampling (Vassallo et al. 2010). However, in practical applications, this is not a significant issue because the seismic data are often irregularly spaced. The overlapping effect of replicas can also be reduced in practice by appropriate windowing of data series prior to regularization.

In the following synthetic and seismic data examples, $\mathbf{\Phi}_1$ in $\underline{\mathbf{\Phi}}$ is selected as the column vector of ones to reconstruct the horizontal linear events more accurately. This selection is crucial in the regularization of seismic data beyond aliasing because it also considers the trace averaging in the least-squares sense. The gradient of this vector is zero, and it cannot be useful for horizontal events whose spectra are vertical in the $f-k$ spectrum. For irregularly sampled data series, the sinusoidal functions no longer have average value zero, causing an error in determining the zero point of the signals. Simultaneous consideration of the column vector of ones with the sinusoids solves this problem, and so it improves the accuracy of wavenumber estimation (Ferraz-Mello 1981; Foster 1996; Ghaderpour and Pagiatakis 2017).

Results and discussion

Synthetic data series regularization beyond aliasing

Consider the following data series and its gradient series

$$f(x_\ell) = 5\sin(25.6x_\ell) + 2.5\sin(128x_\ell + 1) + \sqrt{3}\sin(140x_\ell) + 4\cos(350x_\ell), \quad (11)$$

$$f_x(x_\ell) = 5(25.6)\cos(25.6x_\ell) + 2.5(128)\cos(128x_\ell + 1) + \sqrt{3}(140)\cos(140x_\ell) - 4(350)\sin(350x_\ell), \quad (12)$$

where $x_\ell = \ell/32$, $\ell = 1, \ldots, 32$. The data series is shown by black stars in Fig. 1b. The goal is to regularize the data series on a series with regular spacing 1/128 whose Nyquist wavenumber is 64, so we choose the initial set of wavenumbers for the LSSA, MIMAP, and MALLSSA as $\mathbf{\Omega} = \{1, 2, \ldots, 63\}$. Set $\mathbf{\Omega}$ does not have to contain only integers, and it can be any set of real numbers. The denser the wavenumbers are in $\mathbf{\Omega}$, the more accurate the spectral peaks in the LSSA will be estimated; however, this will increase the computational cost. We also use λ for both MIMAP and MALLSSA in this example that is approximately equal to 0.001. The actual wavenumbers of the data series are real numbers: $25.6/(2\pi) = 4.0743665, 128/(2\pi) = 20.3718327$, $140/(2\pi) = 22.2816920, 350/(2\pi) = 55.7042301$. Since the distance between every two consecutive samples is 1/32 and the goal is to reconstruct the data series on a series with regular spacing 1/128, the data series presents three Nyquist wavenumbers at 16, 32, and 48 (cf., Fig. 1a). The actual locations of two wavenumbers are between 16 and 32, and one is after 48.

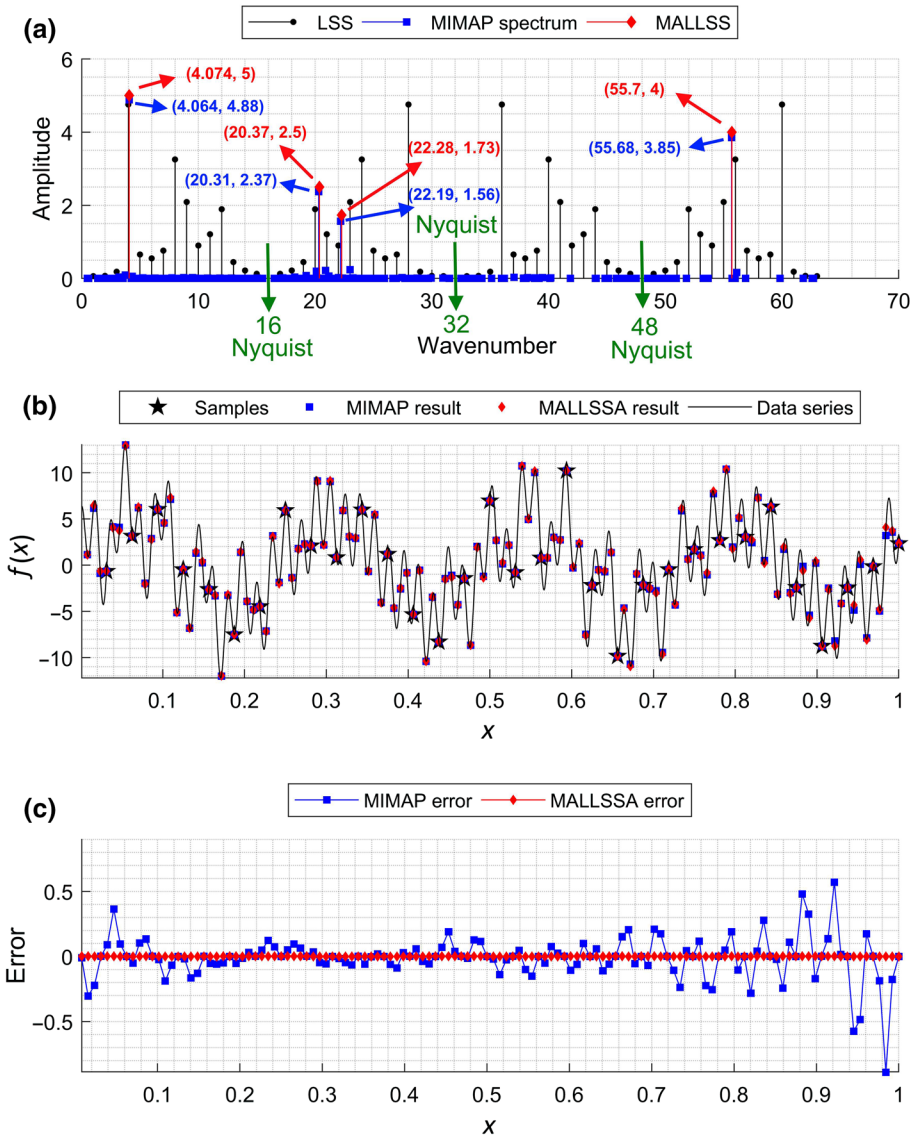


Fig. 1 A coarsely sampled data series (black stars in panel **b**) presenting severe aliasing (black circles in panel **a**) and its regularization on series with spacing 1/128 using the MIMAP (blue squares in panel **b** obtained from wavenumbers shown by blue squares in panel **a**) and MALLSSA (red diamonds in panel **b** obtained from the four wavenumbers shown by red diamonds in panel **a**), and **c** the difference between the ideal data series and the regularized data series using the MIMAP (blue squares) and MALLSSA (red diamonds). The MIMAP spectrum in panel **a** shows spectral leakages and has many wavenumbers in the spectrum that increase the computational cost and reduce the accuracy of the regularization (see panel **c**)

The LSSA amplitude spectrum is shown in Fig. 1a (black peaks). The amplitude corresponding to a wavenumber is the square root of the sum of squares of estimated cosine and sine coefficients of that wavenumber. The estimated wavenumbers using the LSSA show severe aliasing. For example, in the LSSA amplitude spectrum, the sinusoids of wavenumbers 4, 28, 36, and 60 fit the data series identically, and so the true signal that has wavenumber close to 4 is indistinguishable from wavenumbers 28, 36, and 60. Since the amplitudes of the gradient series corresponding to these wavenumbers are different, the MALLSSA is able to estimate the true wavenumber accurately that is 4.0743. The estimated wavenumbers using the MALLSSA are shown by red in Fig. 1a that are accurately estimated.

The MIMAP spectrum is also illustrated by blue in Fig. 1a, showing spectral leakages and has many wavenumbers in the spectrum that increase the computational cost (about 10 times slower than the MALLSSA in this example) and reduce the accuracy of the regularization. The L2 norm of the ideal data series with 1/128 sample spacing is 56.2436, and the L2 norm of the MIMAP and MALLSSA residuals (the difference between the interpolated and actual data series with 1/128 sample spacing) is 1.863 and 0.006, respectively. The differences between the ideal and the regularized data series using the MIMAP and MALLSSA are illustrated in Fig. 1c. Since the coefficients of the components of the gradient series depend on the wavenumber, minimization of Eq. (6) resulted in picking the correct wavenumbers in the spectrum.

To understand the MALLSSA algorithm better, we show the iteration results of the MALLSSA in Table 1. In the first iteration, wavenumber 55.6819 is estimated that is approximately 0.0223 different from its actual value 55.7042. This is a shortcoming of the MIMAP method (out-of-context) caused by the presence of other constituents in the data series. In the second, third, and fourth iterations, the other three wavenumbers are estimated simultaneously. By removing their corresponding components from the data and gradient series simultaneously, the second wavenumber is re-optimized in the fifth iteration, and so on (see the highlighted numbers in Table 1). One can see that all the wavenumbers are accurately estimated in the last iteration. In this

Table 1 The result of wavenumber estimation using the MALLSSA algorithm after each iteration

Iteration number	First wavenumber	Second wavenumber	Third wavenumber	Fourth wavenumber	L2 norm of residual $\hat{\mathbf{g}}$	L2 norm of residual $\hat{\mathbf{g}}_x$
1				55.6819	23.3064	1596.2916
2		20.3075		55.6819	21.3281	1019.4372
3		20.3075	22.1898	55.6819	20.0021	513.6049
4	4.0631	20.3075	22.1898	55.6819	1.5029	202.9275
5	4.0631	**20.3632**	22.1898	55.6819	1.0830	162.6327
6	4.0631	20.3632	22.1898	**55.6931**	1.0017	123.3733
7	4.0631	20.3632	**22.2438**	55.6931	0.5938	79.9480
8	4.0631	20.3632	22.2438	**55.6993**	0.5324	53.5090
9	4.0631	20.3632	**22.2651**	55.6993	0.4580	38.5508
10	4.0631	20.3632	22.2651	**55.7019**	0.4368	29.6227
11	4.0631	**20.3711**	22.2651	55.7019	0.4199	24.1281
12	4.0631	20.3711	**22.2736**	55.7019	0.4157	18.7725
13	**4.0737**	20.3711	22.2736	55.7019	0.0952	16.5123
14	4.0737	20.3711	22.2736	**55.7032**	0.0829	10.6737
15	4.0737	20.3711	**22.2781**	55.7032	0.0471	7.3391
16	4.0737	20.3711	22.2781	**55.7038**	0.0417	4.9594
17	4.0737	20.3711	**22.2802**	55.7038	0.0305	3.3128
18	4.0737	20.3711	22.2802	**55.7040**	0.0287	2.5437
19	4.0737	**20.3718**	22.2802	55.7040	0.0266	2.0808
20	4.0737	20.3718	**22.2809**	55.7040	0.0254	1.7029
21	4.0737	20.3718	22.2809	**55.7041**	0.0245	1.2301
22	**4.0743**	20.3718	22.2809	55.7041	0.0081	1.0792
23	4.0743	20.3718	**22.2812**	55.7041	0.0058	0.9204
24	4.0743	20.3718	22.2812	**55.7042**	0.0054	0.7126
25	4.0743	20.3718	**22.2816**	55.7042	0.0026	0.2233

Note that wavenumbers in bold are estimated more accurately after iteration

example, the MALLSSA had 25 iterations with maximum 4 wavenumbers being estimated simultaneously. However, the MIMAP had 526 iterations to estimate many wavenumbers one at a time, making the computational speed slower than the MALLSSA and with less regularization accuracy.

Synthetic seismic data regularization

The wavelet used in the synthetic examples is an Ormsby wavelet, a common type of synthetic wavelet in reflection seismology. The wavelet is defined by the sinc function as

$$A(t) = \frac{f_4^2 \operatorname{sinc}^2(f_4 t) - f_3^2 \operatorname{sinc}^2(f_3 t)}{f_4 - f_3} - \frac{f_2^2 \operatorname{sinc}^2(f_2 t) - f_1^2 \operatorname{sinc}^2(f_1 t)}{f_2 - f_1}. \tag{13}$$

Ormsby wavelets have several sidelobes (see Fig. 2a), unlike Ricker wavelets which only have two, one on either side. We choose $f_1 = 5\pi$, $f_2 = 10\pi$, $f_3 = 20\pi$, and $f_4 = 30\pi$ Hz, and so the wavelet defines a trapezoidal shape in the frequency spectrum with low-cut, low-pass, high-pass, and high-cut frequencies f_1, f_2, f_3, and f_4, respectively (see Fig. 2b). The wavelet shown in Fig. 2a may be considered as a pressure wave with the unit of kilopascal (kPa).

In real marine acquisition data sets, noise is expected to be higher at low frequencies than at high frequencies (Vassallo et al. 2010). Therefore, in the synthetic examples, we add special type of random noise to both pressure and gradient data sets as described below. The MATLAB command filtfilt (b, a, ns) is used to generate $1/f$ noise, where a and b are transfer function coefficients, ns is a vector containing normally distributed random numbers (can be generated by MATLAB command 'randn'), and f is frequency. Figure 3a shows such noise when $a = 1$ and b is generated using the MATLAB command 'firls' that creates a series containing the coefficients of the finite-duration impulse response filter closely matching the $1/f$ pass-band (red graph in Fig. 3). To see the performance of the MALLSSA in higher frequencies, random noise (Fig. 3b) is also considerably noisy in higher frequencies. The mean of random noise is approximately zero, and its relative amplitude is about 30% and 15% of the signals used in the following synthetic stationary and non-stationary seismic data examples, respectively.

Simple synthetic stationary seismic data regularization

A synthetic seismic data (pressure) containing four linear events with different amplitudes and dips are shown in Fig. 4a. The events are generated using the Ormsby wavelet illustrated in Fig. 2, and the time sampling rate is 1000 samples per second. Suppose that the trace spacing is 10 m, toward the crossline direction. Assume that the unit distance is 1 km in this and the next example, so the wavenumbers are the number of cycles per kilometer (c/km). The $f-k$ spectrum of this data is shown in Fig. 4b. The vertical peaks at zero wavenumber in the $f-k$ spectrum correspond to the horizontal event in Fig. 4a. The spatial gradient data of the pressure data are shown in Fig. 4c. The $f-k$ spectrum of the spatial gradient data is also shown in Fig. 4d. From Fig. 4d, one can observe that the crossline gradient amplitude is very low at low wavenumbers, and it linearly grows toward higher wavenumbers. The horizontal linear event in the seismic data disappeared in its crossline gradients as seen in Fig. 4c and d.

In this example, 75 traces are removed such that the remaining traces are equally spaced with 40 m spacing, and random noise as shown in Fig. 3 is introduced to each trace and illustrated with its $f-k$ spectrum in Fig. 5a and b, respectively. The crossline gradients are kept in the samples

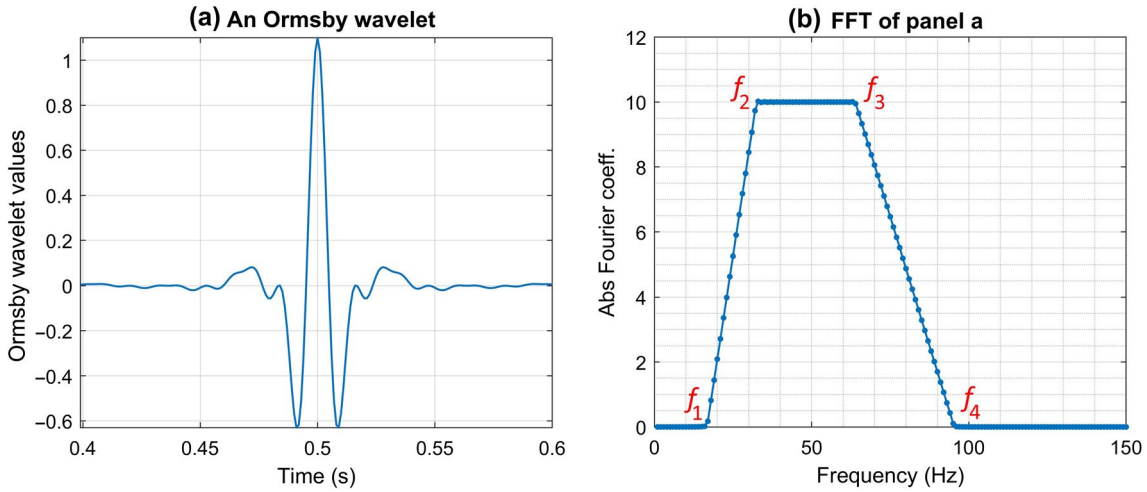


Fig. 2 **a** An Ormsby wavelet given by Eq. (13) multiplied by factor 0.01 with low-cut, low-pass, high-pass, and high-cut frequencies $f_1 = 5\pi$, $f_2 = 10\pi$, $f_3 = 20\pi$, and $f_4 = 30\pi$ Hz, respectively, and **b** the spectrum of panel **a**. The time sampling rate is 1000 samples per second

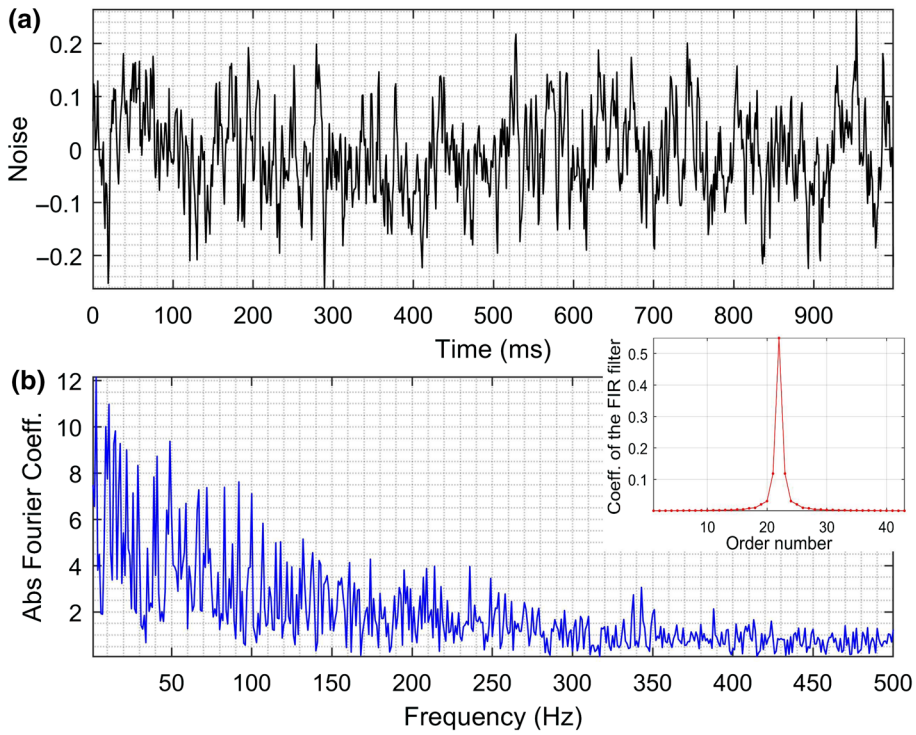


Fig. 3 **a** Random noise (1/*f* noise, where *f* is frequency), and **b** its Fourier spectrum. It is noisier at lower frequencies than higher frequencies. The small panel on the right in red is generated by implementing the MATLAB command 'firls' to create random noise

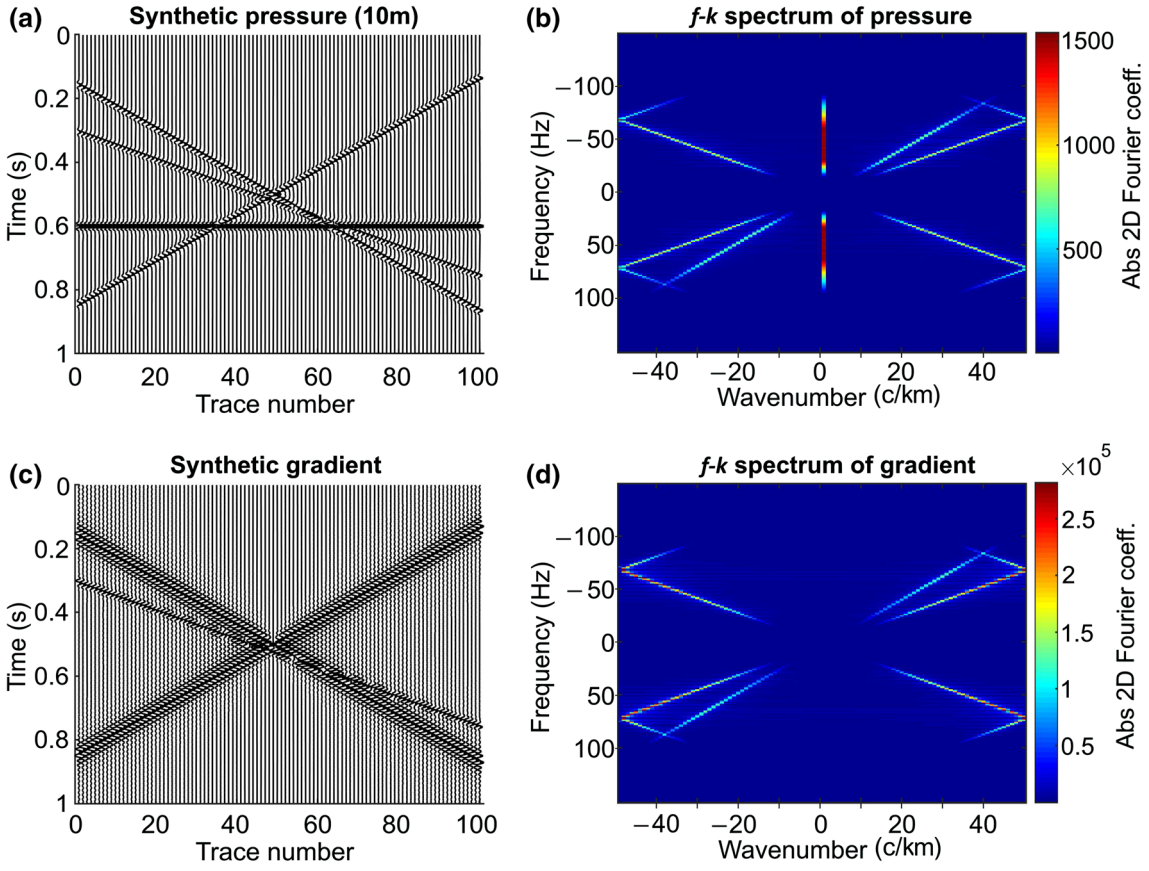


Fig. 4 Simple synthetic seismic data set. **a** Pressure (10 m trace spacing), **b** the $f-k$ spectrum of panel **a**, **c** gradient, and **d** the $f-k$ spectrum of panel **c**

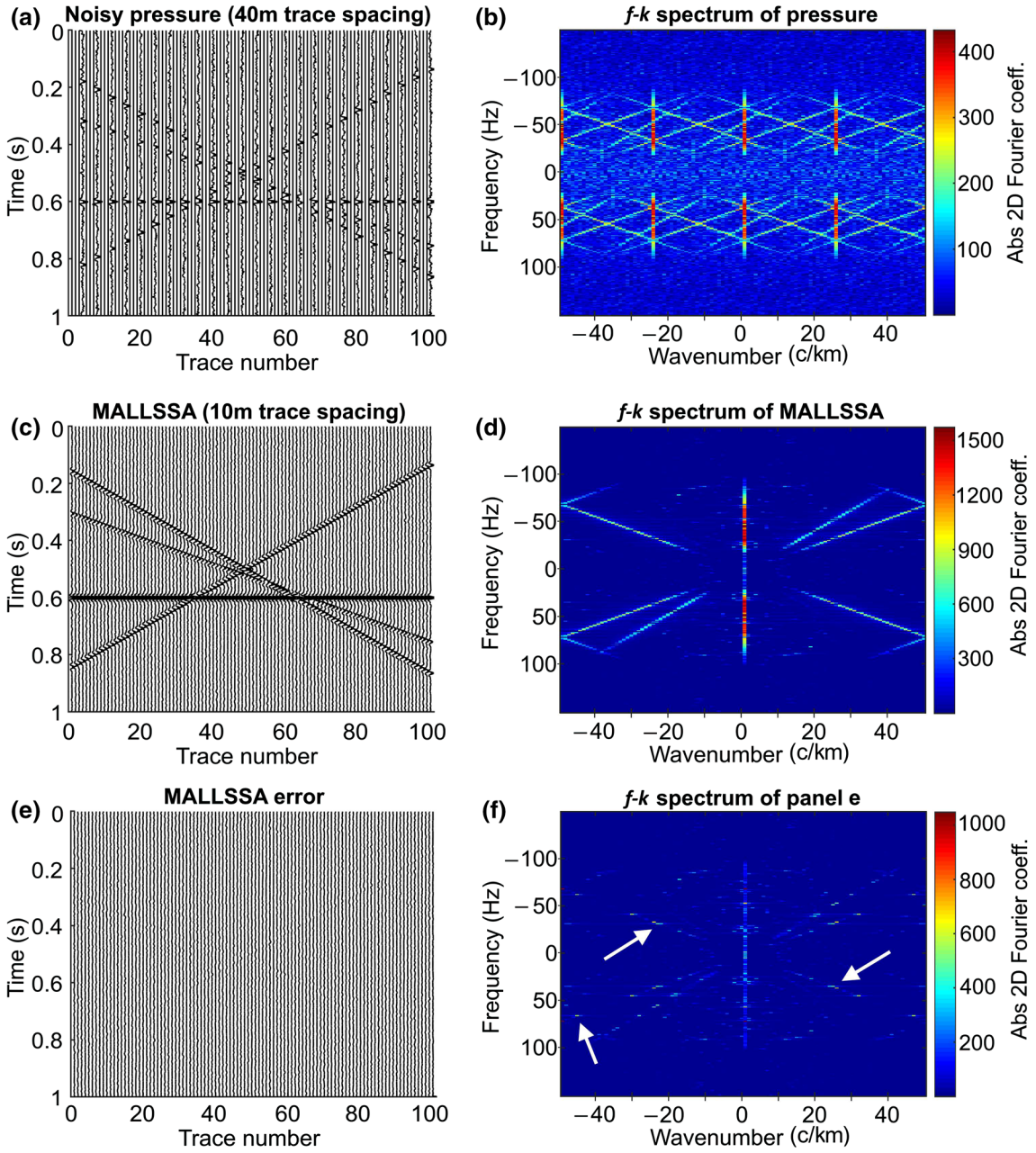


Fig. 5 **a** The noisy seismic data with 40 m trace spacing, **b** the $f-k$ spectrum of panel **a**, **c** the MALLSSA result with 10 m trace spacing, **d** the $f-k$ spectrum of panel **c**, **e** the difference between the original noise-free pressure shown in Fig. 4a and panel **c**, and **f** the $f-k$ spectrum of panel **e**. The linear events are well constructed, and random noise and potential overlaps of spectral replicas at the same wavenumbers mainly cause the small error (cf., arrows in panel **f**)

positions with added independent noise as shown in Fig. 3 and are not shown here. Also, the removed traces are set to zero only to generate the $f-k$ spectra. Figure 5b shows that the events are spatially aliased.

The MALLSSA regularization result is obtained without windowing after one round of iterations, and its $f-k$ spectrum is shown in Fig. 5c and d, respectively. The MALLSSA regularization result clearly constructed all the events in 10 m trace spacing, and its $f-k$ spectrum is approximately the same as the noise-free spectrum shown in Fig. 4b. The difference between the original noise-free data (cf., Fig. 4a) and the MALLSSA result with its $f-k$ spectrum is illustrated in Fig. 5e and f, respectively. The presence of error is mainly due to potential overlap of spectral replicas at the same wavenumbers and also the noisy pressure and gradient data (see arrows in Fig. 5f). Note that for each data series

in the regularization process, a few wavenumbers were estimated simultaneously in the first round of iterations (the only round) at 99% confidence level.

Simple synthetic non-stationary seismic data regularization

A synthetic seismic data (pressure) containing three curved and two linear events with different amplitudes and dips are shown in Fig. 6a. The events are generated using the same Ormsby wavelet as the previous example, and the time sampling rate is 1000 samples per second with 10 m trace spacing. Also, random noise is introduced to each trace. The $f-k$ spectrum of this data is shown in Fig. 6b. In the $f-k$ spectrum, the vertical peaks at zero wavenumber correspond to the horizontal event, and the slanted peaks correspond to the other linear event in Fig. 6a. The noisy gradient data of the pressure data are shown in Fig. 6c. The $f-k$ spectrum of the spatial gradient data is also shown in Fig. 6d. All the

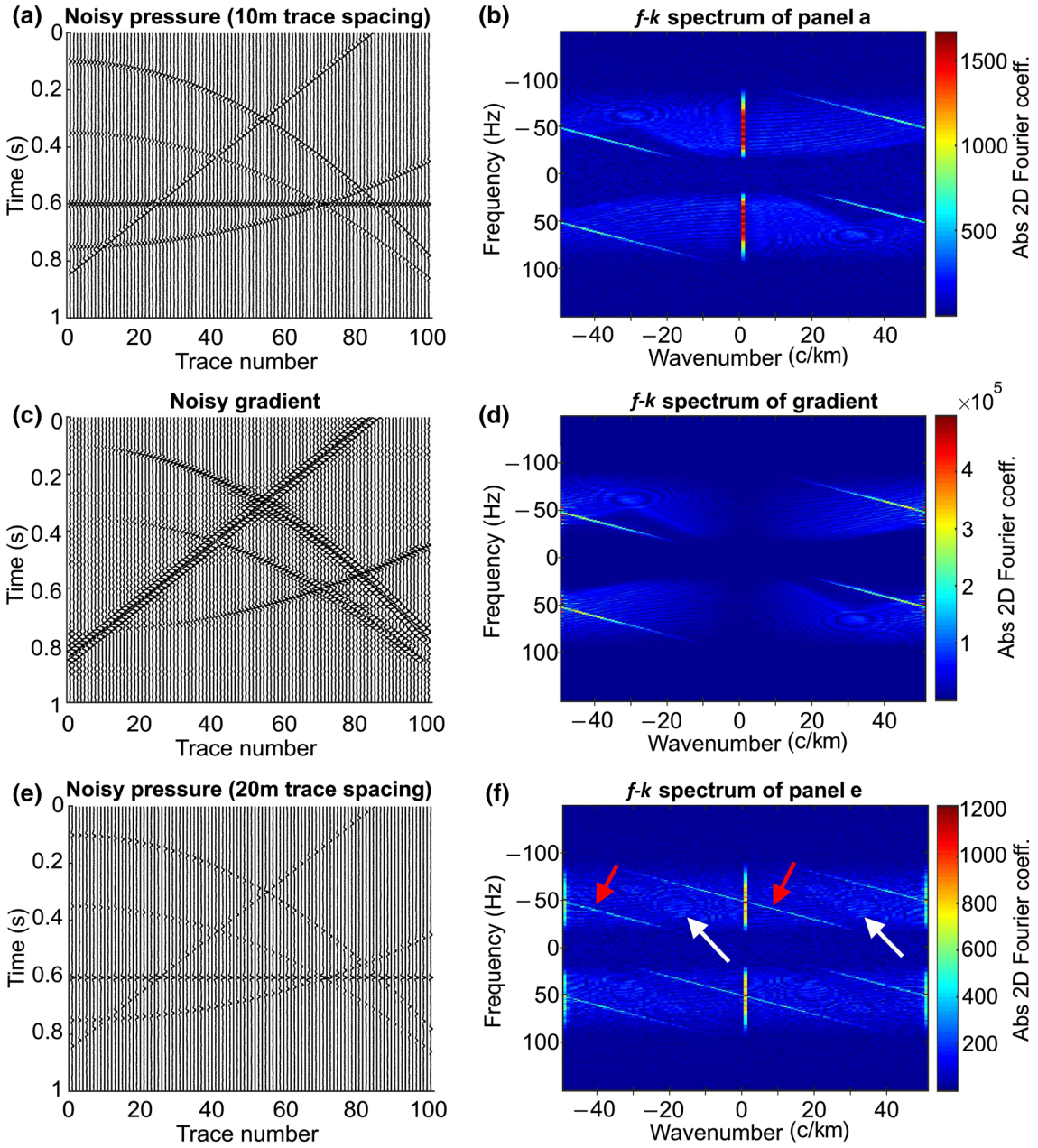


Fig. 6 Simple synthetic seismic data set. **a** Noisy pressure, **b** the $f-k$ spectrum of panel **a**, **c** noisy gradient, **d** the $f-k$ spectrum of panel **c**, **e** the noisy pressure with 20 m trace spacing, and **f** the $f-k$ spectrum of panel **e**. The vertical peaks at zero wavenumber (reddish peaks) in panel **b**, corresponding to the horizontal event, disappeared in panel **d**. Panel **f** shows that all the events are spatially aliased (the red and white arrows in panel **f** show some of the aliasing effects of the linear and curved events, respectively)

traces of even numbers are removed so that the trace spacing is 20 m (cf., Fig. 6e). The $f-k$ spectrum of the result is illustrated in Fig. 6f, showing that all the events are spatially aliased (e.g., the red and white arrows show some of the aliasing effects of the linear and curved events, respectively). The crossline gradients are kept in the samples positions in the analysis and are not shown here.

The linear events can be well constructed by the first round of iterations in the MALLSSA (cf., Fig. 7a, b), yet the curved events cannot be properly constructed and produce artifacts in the traces along with random noise. Arrow in Fig. 7a shows some artifacts from the linear event that is due to the non-stationary behavior of the event. However, after a few rounds of iterations the curved events can also be well constructed and interpolated (cf., Fig. 7c and d). The difference between the original noise-free pressure and the regularization result (Fig. 7c) is illustrated in Fig. 7e. The $f-k$ spectrum of Fig. 7e is also illustrated in Fig. 7f. Note that no

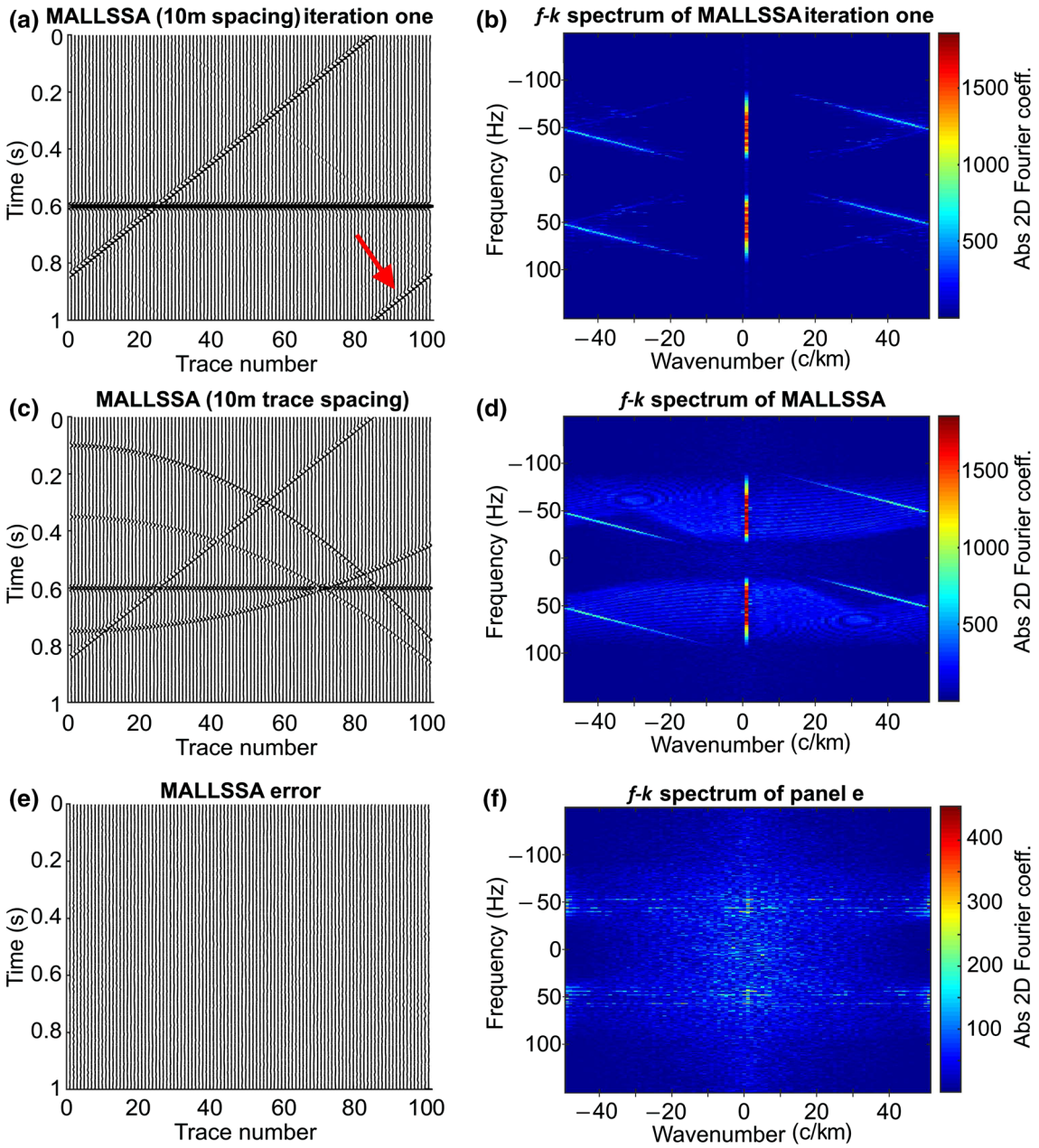


Fig. 7 The MALLSSA results. **a** After first iteration with 10 m trace spacing (the linear events are constructed), **b** the $f-k$ spectrum of panel **a**, **c** after a few more iterations with 10 m trace spacing, **d** the $f-k$ spectrum of panel **c**, **e** the difference between the original noise-free pressure and panel **c**, and **f** the $f-k$ spectrum of panel **e**. Arrow in panel **a** shows some artifacts from the linear event that are mitigated after a few more iterations (compare with panel **c**, where all the events are well constructed)

windowing technique is used in this example. In the regularization process, for each data series, a few wavenumbers were estimated simultaneously at each round of iterations, and in total only a few rounds of iterations were needed to regularize each data series at 99% confidence level.

A 2D marine seismic data regularization

The field data example is a marine 2D shot gather from a deep water of Gulf of Mexico (Fig. 8a). This marine seismic data are the same as the one used by Fomel (2002); Chen et al. (2015); Ghaderpour et al. (2018). The time sampling rate is 250 samples per second. The $f-k$ spectrum of this data is shown in Fig. 8b. The spatial gradient of the marine data is estimated using the ALLSSA with added independent noise, and it is shown in Fig. 8c. The $f-k$ spectrum of the spatial gradient data is also shown in Fig. 8d. The normalized wavenumber axis in the $f-k$ spectrums is obtained by considering the offset of the far offset trace equal to one. The traces of even numbers are removed from the marine seismic data and its gradient data. The results of marine seismic data and its $f-k$ spectrum are illustrated in Fig. 9a and b, respectively. The effect of aliased events can be clearly seen from the $f-k$ spectrum.

For better performance, the marine seismic data are divided into two spatial non-overlapping windows of the same size. The MALLSSA regularization result and its $f-k$ spectrum are shown in Fig. 9c and d, respectively. For better comparison between the original marine seismic data and its regularization result, their difference is calculated and illustrated along with its $f-k$ spectrum in Fig. 9e and f, respectively. We used a mask to set the values (minor artifacts) of the northeast part of the images shown in Fig. 9c and e to zero after the interpolation. In the iteration process, a few wavenumbers were estimated simultaneously at each round of iteration, and the process stopped after a few rounds of iterations by applying a threshold at 99% confidence level. The final regularization result is pleasant and acceptable for practical applications.

In this example, the traces were equally spaced and removing the traces of even numbers makes the remaining traces also equally spaced with a Nyquist wavenumber that is half of the original Nyquist wavenumber. When the distance between traces varies, the ALLSSA and MALLSSA perform better in regularization because these methods consider the correlations among the sinusoids of various wavenumbers simultaneously.

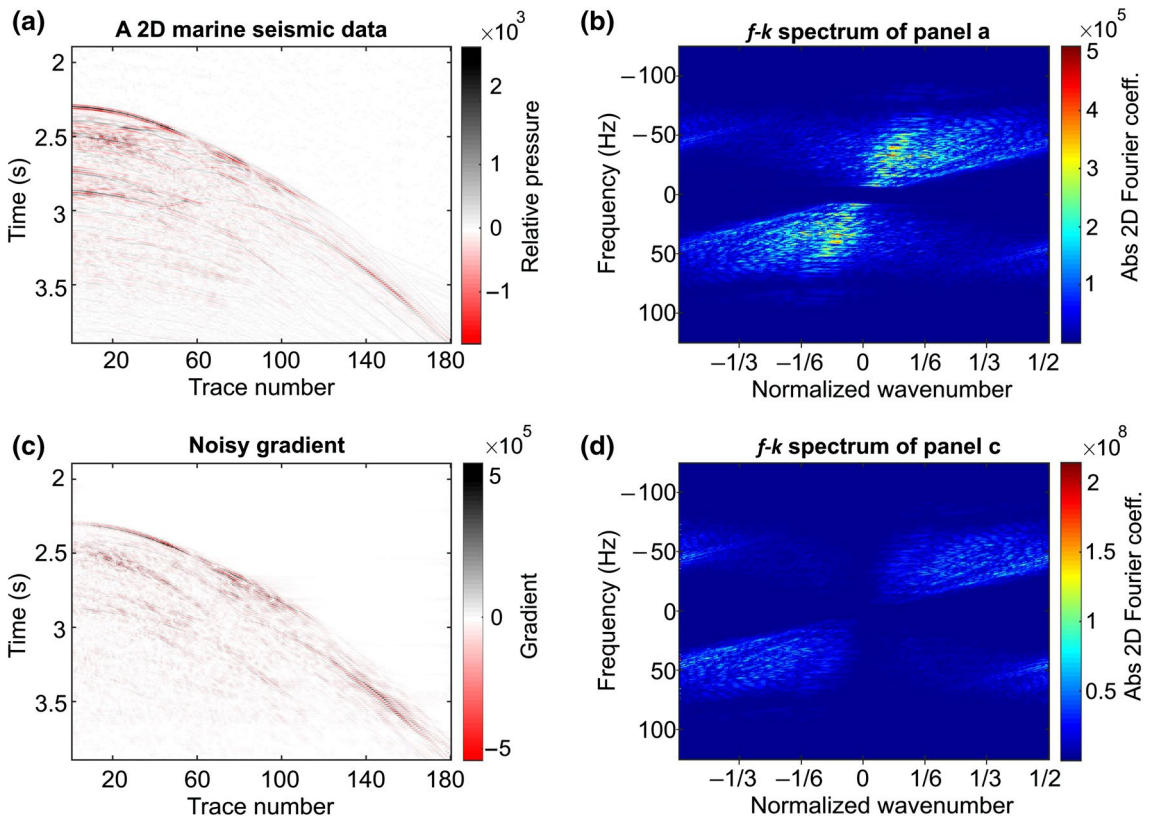


Fig. 8 **a** A 2D marine seismic data obtained from a deep water of Gulf of Mexico, **b** the $f-k$ spectrum of panel **a**, **c** the estimated noisy gradient data, and **d** the $f-k$ spectrum of panel **c**

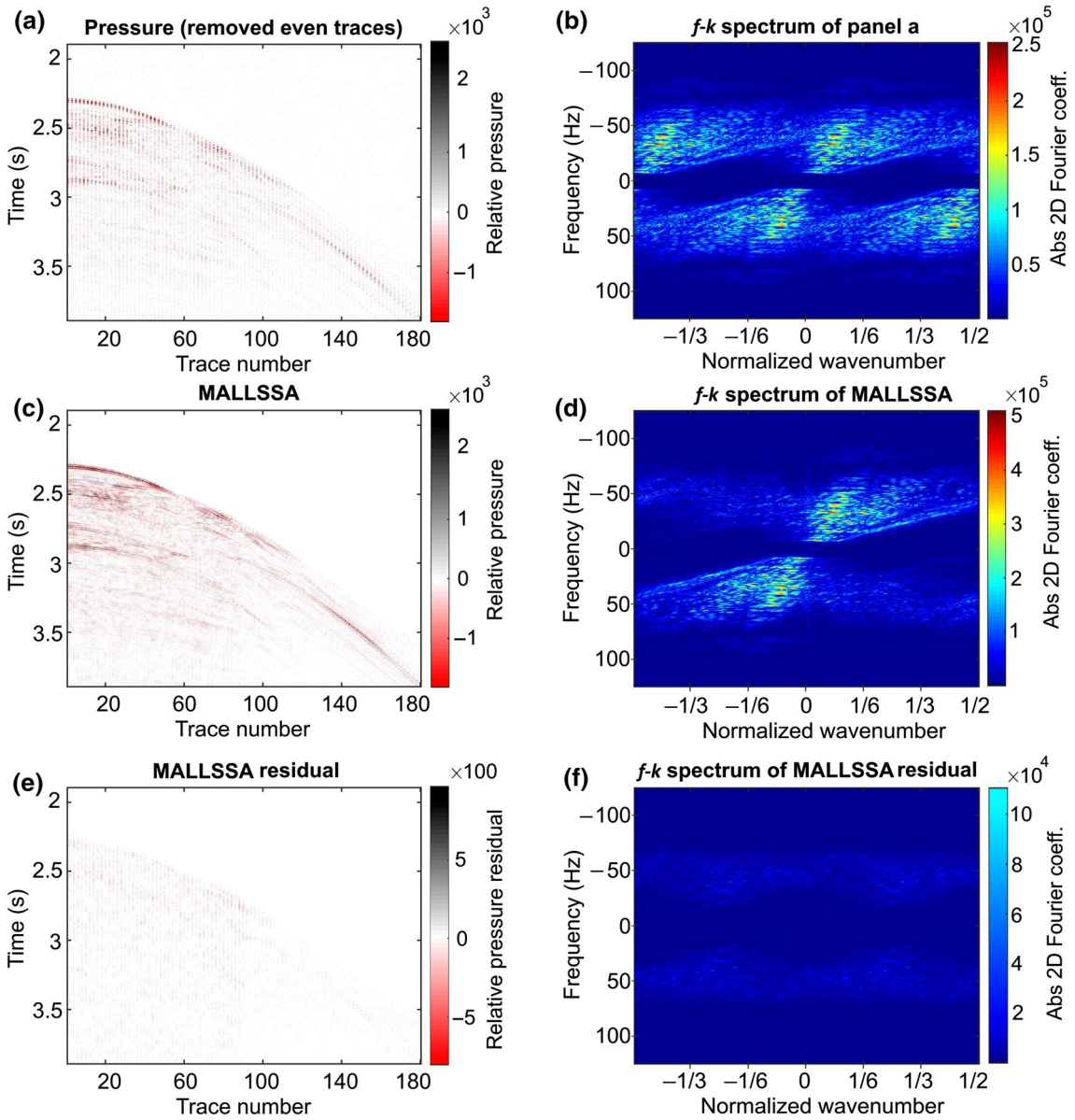


Fig. 9 **a** The traces of even numbers in Fig. 8a are removed, **b** the $f-k$ spectrum of panel **a**, **c** the MALLSSA regularization result, **d** the $f-k$ spectrum of panel **c**, **e** the difference (residual) between panel **c** and Fig. 8a, and **f** the $f-k$ spectrum of panel **e**

Conclusions

The ALLSSA is a robust method of regularizing seismic data presenting mild aliasing in the $f-k$ spectrum. However, when the seismic data are coarsely sampled, additional information is needed to aid regularization. This is crucial in regularization because incorrect selection of the spectral peaks results in inaccurate regularization. In this paper, we developed a multichannel method (MALLSSA) that incorporates the spatial gradients of seismic data into the ALLSSA to regularize data series beyond aliasing.

Unlike the MIMAP, the MALLSSA estimates the wavenumbers more accurately by re-optimizing them simultaneously at each round of iterations, reducing the computational cost and increasing the regularization accuracy. A limitation of the MALLSSA is the potential overlap of two or more spectral replicas at the same wavenumber, similar to the MIMAP (Vassallo et al. 2010). In practice, this effect can be reduced by an appropriate windowing technique.

The MALLSSA simultaneously considers the constituents of known forms and the covariance matrices associated with data series. The MALLSSA is designed for wide-sense stationary data series; however, appropriate windowing techniques similar to the ones proposed by Ghaderpour and Pagiatakis

(2017) and Ghaderpour et al. (2018) can be used to adapt it to seismic data with more complex geologic structure.

In the MIMAP, the wavenumbers are estimated one at a time as iterations progress, while in the MALLSSA, multiple wavenumbers are estimated simultaneously in a single round of iterations. In practice, a few wavenumbers will be estimated simultaneously in a single round of iterations within a spatial window; however, the total number of iterations decreases significantly due to the accuracy of the estimated wavenumbers, making the MALLSSA generally faster and more accurate than the MIMAP. Furthermore, by applying a confidence level in the MALLSSA, one can estimate the statistically significant spectral components at that level to attenuate random noise like the ALLSSA. The MALLSSA may be naturally extended to regularize 3D and 5D seismic data sets, and analyses of such data sets are subject to future work.

Acknowledgements This paper constitutes Chapter 5 of the author's second Ph.D. thesis supervised by Spiros D. Pagiatakis. The author would like to thank the anonymous reviewers for their time and great comments and Drs. Ian Moore, Spiros D. Pagiatakis, Amir Asif, and Jianguo Wang for useful discussions. This research has been financially supported by Pacific Institute for the Mathematical Sciences (PIMS), Natural Sciences and Engineering Research Council (NSERC) of Canada and the postdoctoral program at the Department of Mathematics and Statistics, University of Calgary.

Compliance with ethical standards

Conflicts of Interest The author states that there is no conflict of interest.

Appendix: Derivation of estimated multichannel coefficients

A similar methodology as in Wells and Krakiwsky (1971) and Vaníček and Krakiwsky (1986) is used to obtain Eq. (8) and to show that it minimizes Eq. (6). Simplify Eq. (6) using elementary matrix operations as follows (Horn and Johnson 2012):

$$\begin{aligned}\Psi_k(\bar{\mathbf{c}}) =& \left(\mathbf{f}-\overline{\mathbf{\Phi}}\,\bar{\mathbf{c}}\right)^{\mathrm{T}}\mathbf{P}\left(\mathbf{f}-\overline{\mathbf{\Phi}}\,\bar{\mathbf{c}}\right)+\left(\mathbf{f}_{\mathbf{x}}-\overline{\mathbf{\Phi}}_{\mathbf{x}}\,\bar{\mathbf{c}}\right)^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\left(\mathbf{f}_{\mathbf{x}}-\overline{\mathbf{\Phi}}_{\mathbf{x}}\,\bar{\mathbf{c}}\right)\\ =&\left(\mathbf{f}^{\mathrm{T}}-\bar{\mathbf{c}}^{\mathrm{T}}\overline{\mathbf{\Phi}}^{\mathrm{T}}\right)\mathbf{P}\left(\mathbf{f}-\overline{\mathbf{\Phi}}\,\bar{\mathbf{c}}\right)+\left(\mathbf{f}_{\mathbf{x}}^{\mathrm{T}}-\bar{\mathbf{c}}^{\mathrm{T}}\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\right)\mathbf{P}_{\mathbf{x}}\left(\mathbf{f}_{\mathbf{x}}-\overline{\mathbf{\Phi}}_{\mathbf{x}}\,\bar{\mathbf{c}}\right)\\ =&\mathbf{f}^{\mathrm{T}}\mathbf{P}\mathbf{f}-\mathbf{f}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}\,\bar{\mathbf{c}}-\bar{\mathbf{c}}^{\mathrm{T}}\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\mathbf{f}+\bar{\mathbf{c}}^{\mathrm{T}}\left(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}\right)\bar{\mathbf{c}}\\ &+\mathbf{f}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\mathbf{f}_{\mathbf{x}}-\mathbf{f}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\overline{\mathbf{\Phi}}_{\mathbf{x}}\,\bar{\mathbf{c}}-\bar{\mathbf{c}}^{\mathrm{T}}\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\mathbf{f}_{\mathbf{x}}+\bar{\mathbf{c}}^{\mathrm{T}}\left(\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\overline{\mathbf{\Phi}}_{\mathbf{x}}\right)\bar{\mathbf{c}}.\end{aligned} \tag{14}$$

To minimize Eq. (14) with respect to $\bar{\mathbf{c}}$, one may calculate the derivative of $\Psi_k(\bar{\mathbf{c}})$ with respect to $\bar{\mathbf{c}}$ and set it equal to zero as follows:

$$\begin{aligned}\frac{\partial\Psi_k(\bar{\mathbf{c}})}{\partial\bar{\mathbf{c}}} =& -\mathbf{f}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}-\mathbf{f}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}+2\,\bar{\mathbf{c}}^{\mathrm{T}}\left(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}\right)\\ &-\mathbf{f}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\overline{\mathbf{\Phi}}_{\mathbf{x}}-\mathbf{f}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\overline{\mathbf{\Phi}}_{\mathbf{x}}+2\,\bar{\mathbf{c}}^{\mathrm{T}}\left(\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\overline{\mathbf{\Phi}}_{\mathbf{x}}\right)=0.\end{aligned} \tag{15}$$

Rearranging and simplifying the terms in Eq. (15), one obtains

$$\frac{\partial\Psi_k(\bar{\mathbf{c}})}{\partial\bar{\mathbf{c}}}=-2\left(\mathbf{f}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}+\mathbf{f}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\overline{\mathbf{\Phi}}_{\mathbf{x}}\right)+2\,\bar{\mathbf{c}}^{\mathrm{T}}\left(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}+\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\overline{\mathbf{\Phi}}_{\mathbf{x}}\right)=0. \tag{16}$$

Taking the transpose of both sides of Eq. (16) and solve for $\bar{\mathbf{c}}$ yields Eq. (8):

$$\hat{\bar{\mathbf{c}}}=\begin{bmatrix}\mathbf{v}\\ \hat{\mathbf{c}}_k\end{bmatrix}=\left(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}+\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\overline{\mathbf{\Phi}}_{\mathbf{x}}\right)^{-1}\left(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\mathbf{f}+\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\mathbf{f}_{\mathbf{x}}\right), \tag{17}$$

where $\hat{\bar{\mathbf{c}}}$ is used to indicate that $\bar{\mathbf{c}}$, minimizing Eq. (6), is an estimation. Using the second derivative test in Calculus, one may verify that $\hat{\bar{\mathbf{c}}}$ in fact minimizes Eq. (6):

$$\frac{\partial}{\partial\bar{\mathbf{c}}}\left\{\frac{\partial\Psi_k(\bar{\mathbf{c}})}{\partial\bar{\mathbf{c}}}\right\}=2\left(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}+\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\overline{\mathbf{\Phi}}_{\mathbf{x}}\right), \tag{18}$$

that is positive definite because for any $\mathbf{z}\neq\mathbf{0}$,

$$\mathbf{z}^{\mathrm{T}}\left(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}+\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\overline{\mathbf{\Phi}}_{\mathbf{x}}\right)\mathbf{z}=\mathbf{z}^{\mathrm{T}}\left(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}\right)\mathbf{z}+\mathbf{z}^{\mathrm{T}}\left(\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\overline{\mathbf{\Phi}}_{\mathbf{x}}\right)\mathbf{z}>0. \tag{19}$$

Note that $\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}$ and $\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\overline{\mathbf{\Phi}}_{\mathbf{x}}$ are positive definite, and $\overline{\mathbf{\Phi}}$ has the same column dimension as $\overline{\mathbf{\Phi}}_{\mathbf{x}}$.

Suppose that data series $\mathbf{f}$ (dimension n_1) and $\mathbf{f}_{\mathbf{x}}$ (dimension n_2) have been derived from two statistically independent populations of random variables following the multidimensional normal distributions $\mathcal{N}(\mathbf{0},\mathbf{C}_{\mathbf{f}})$ and $\mathcal{N}(\mathbf{0},\mathbf{C}_{\mathbf{f}_{\mathbf{x}}})$, respectively. The probability distribution function of the MALLSS given by Eq. (9) can be derived using similar techniques in Pagiatakis (1999). Let $Q_s=\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}\mathbf{\Phi}_k\hat{\mathbf{c}}_k$, $Q'_s=\hat{\mathbf{g}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\mathbf{\Phi}_{k_{\mathbf{x}}}\hat{\mathbf{c}}_k$, $Q_s+Q_n=\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{g}}$, $Q'_s+Q'_n=\hat{\mathbf{g}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_{\mathbf{x}}\hat{\mathbf{g}}_{\mathbf{x}}$, and q be the number of constituents of known forms in $\mathbf{f}$ and $\mathbf{f}_{\mathbf{x}}$. Eq. (9) can be written as:

$$s_{\mathrm{multi}}(\omega_k)=\frac{Q_s+Q'_s}{Q_s+Q_n+Q'_s+Q'_n}=\left(1+\frac{Q_n+Q'_n}{Q_s+Q'_s}\right)^{-1}. \tag{20}$$

Therefore, $Q_s\sim\chi^2_2$, $Q'_s\sim\chi^2_2$, $Q_n\sim\chi^2_{n_1-q-2}$, and $Q_n\sim\chi^2_{n_2-q-2}$, where χ^2_r is the chi-squared distribution with r degrees of freedom, and '$\sim$' means follows. According to the assumption, Q_s, Q'_s, Q_n, and Q'_n are statistically independent random variables, and so

$$s_{\text{multi}}(\omega_k) = \left(1 + \frac{Q_n + Q_n'}{Q_s + Q_s'}\right)^{-1} \sim \left(1 + \frac{\chi^2_{\mathfrak{R}}}{\chi^2_4}\right)^{-1} = \left(1 + \frac{\mathfrak{R}}{4} F_{\mathfrak{R},4}\right)^{-1} = \beta_{2,\mathfrak{R}/2}, \tag{21}$$

where $\mathfrak{R} = n_1 + n_2 - 2q - 4$ and $\beta_{2,\mathfrak{R}/2}$ is the β-distribution with shape parameters 2 and $\mathfrak{R}/2$ (Hogg et al. 2013, Chapter 3). If $n = n_1 = n_2$, then $s_{\text{multi}}(\omega_k) \sim \beta_{2,n-q-2}$. From the right tail of the beta distribution, one may obtain a critical value at certain confidence level (usually 95% or 99%) to identify statistically significant spectral peaks in the multichannel least-squares spectrum for each round of iterations in the MALLSSA (Ghaderpour 2018, Appendix C).

References

Abma R, Claerbout J (1995) Lateral prediction for noise attenuation by $t - x$ and $f - x$ techniques. Geophysics 60(6):1887–1896

Abma R, Kabir N (2006) 3D interpolation of irregular data with a POCS algorithm. Geophysics 71(6):E91–E97

Chen Y, Zhang L, Mo L (2015) Seismic data interpolation using nonlinear shaping regularization. J Seism Explor 24(5):327–342

Chen Y, Zhang D, Jin Z, Chen X, Zu S, Huang W, Gan S (2016) Simultaneous denoising and reconstruction of 5D seismic data via damped rank-reduction method. Geophys J Int 206(3):1695–1717

Crawley S (2000) Seismic trace interpolation with nonstationary prediction-error filters. PhD thesis, Stanford University, USA

Craymer MR (1998) The least-squares spectrum, its inverse transform and autocorrelation function: theory and some application in geodesy. PhD thesis, University of Toronto, Canada

Dragoset B, Verschuur E, Moore I, Bisley R (2010) A perspective on 3D surface-related multiple elimination. Geophysics 75(5):75A245–75A261

Ferraz-Mello S (1981) Estimation of periods from unequally spaced observations. Astron J 86(4):619–624

Fomel S (2002) Applications of plane-wave destruction filters. Geophysics 67(6):1946–1960

Fomel S (2007) Shaping regularization in geophysical-estimation problems. Geophysics 72(2):R29–R36

Foster G (1996) Wavelet for period analysis of unevenly sampled time series. Astron J 112(4):1709–1729

Gao JJ, Stanton A, Naghizadeh M, Sacchi MD, Chen X (2012) Convergence improvement and noise attenuation considerations for beyond alias projection onto convex sets reconstruction. Geophys Prospect 61(s1):138–151

Ghaderpour E (2018) Least-squares wavelet analysis and its applications in geodesy and geophysics. PhD thesis, York University, Canada

Ghaderpour E, Pagiatakis SD (2017) Least-squares wavelet analysis of unequally spaced and non-stationary time series and its applications. Math Geosci 49(7):819–844

Ghaderpour E, Liao W, Lamoureux MP (2018) Antileakage least-squares spectral analysis for seismic data regularization and random noise attenuation. Geophysics 83(3):V157–V170

Guo Z, Zheng Y, Liao W (2015) High fidelity seismic trace interpolation. In: SEG, expanded abstracts, pp 3915–3919. https://doi.org/10.1190/segam2015-5923716.1

Hogg RV, McKean J, Craig AT (2013) Introduction to mathematical statistics, 7th edn. Pearson, Boston, p 694

Horn RA, Johnson CR (2012) Matrix analysis, 2nd edn. Cambridge University Press, Cambridge

Liu G, Chen X (2017) Seismic data interpolation using frequency-domain complex nonstationary autoregression. Geophys Prospect 66(3):478–497

Liu B, Sacchi MD (2004) Minimum weighted norm interpolation of seismic records. Geophysics 69(6):1560–1568

Özbek A, Vassallo M, Özdemir K, Molteni D, Alp YK (2010a) Anti-alias optimal interpolation with priors. In: SEG technical program expanded abstracts, pp 3401–3405. https://doi.org/10.1190/1.3513555

Özbek A, Vassallo M, Özdemir K, van Manen DJ, Eggenberger K (2010b) Crossline wavefield reconstruction from multicomponent streamer data: part 2-joint interpolation and 3D up/down separation by generalized matching pursuit. Geophysics 75(6):WB69–WB85

Özbek A, Vassallo M, Eggenberger K, van Manen DJ, Özdemir K, Curtis T (2012) On the role of priors in generalized matching pursuit to reconstruct wavefields from multicomponent streamer data. In: 74th EAGE conference and exhibition incorporating SPE EUROPEC, Copenhagen, Denmark. https://doi.org/10.3997/2214-4609.20148377

Pagiatakis S (1999) Stochastic significance of peaks in the least-squares spectrum. J Geod 73(2):67–78

Robertsson JO, Moore I, Vassallo M, Özdemir AK, van Manen DJ, Özbek A (2008) On the use of multicomponent streamer recordings for reconstruction of pressure wavefields in the crossline direction. Geophysics 73(5):A45–A49

Schonewille M, Klaedtke A, Vigner A (2009) Anti-alias anti-leakage Fourier transform. In: SEG technical program expanded abstracts, pp 3249–3253. https://doi.org/10.1190/1.3255533

Spitz S (1991) Seismic trace interpolation in the $f - x$ domain. Geophysics 56(6):785–794

Vaníček P (1969) Approximate spectral analysis by least-squares fit. Astrophys Space Sci 4:387–391

Vaníček P, Krakiwsky EJ (1986) Geodesy the concepts. University of New Brunswick, Amsterdam

Vassallo M, Özbek A, Özdemir AK, Eggenberger K (2010) Crossline wavefield reconstruction from multicomponent streamer data: part 1-multichannel interpolation by matching pursuit (MIMAP) using pressure and its crossline gradient. Geophysics 75(6):WB53–WB67

Wang Y (2002) Seismic trace interpolation in the $f - x - y$ domain. Geophysics 67(4):1232–1239

Wang Y, Cao J, Yang C (2011) Recovery of seismic wavefields based on compressive sensing by an l1-norm constrained trust region method and the piecewise random subsampling. Geophys J Int 187(1):199–213

Wang B, Chen X, Li J, Cao J (2016) An improved weighted projection onto convex sets method for seismic data interpolation and denoising. IEEE J Sel Top Appl Earth Obs Remote Sens 9(1):228–235

Weglein AB, Gasparotto FA, Carvalho PM, Stolt RH (1997) An inverse-scattering series method for attenuating multiples in seismic reflection data. Geophysics 62(6):1975–1989

Wells DE, Krakiwsky EJ (1971) The method of least-squares. Department of Surveying Engineering, University of New Brunswick, Canada

Xu S, Zhang Y, Pham D, Lambaré G (2005) Antileakage Fourier transform for seismic data regularization. Geophysics 70(4):V87–V95

Xu S, Zhang Y, Lambaré G (2010) Antileakage Fourier transform for seismic data regularization in higher dimensions. Geophysics 75(6):WB113–WB120

Yang P, Gao J, Chen W (2012) Curvelet-based POCS interpolation of nonuniformly sampled seismic records. J Appl Geophys 79:90–99

Acta Geophysica (2019) 67:1365–1377
https://doi.org/10.1007/s11600-019-00325-y

REVIEW ARTICLE - APPLIED GEOPHYSICS

Magnetotelluric inversion of one- and two-dimensional synthetic data based on hybrid genetic algorithms

Joelson da Conceição Batista[1] · **Edson Emanoel Starteri Sampaio**[1]

Received: 15 July 2018 / Accepted: 4 July 2019 / Published online: 10 July 2019

Abstract
We applied the technique of the genetic algorithms and a local methodology integrating the Gauss–Newton and Conjugate Gradient (GNCG) techniques to test one-dimensional inverse modeling of synthetic magnetotelluric data. The result of this modeling applied to a homogeneous and isotropic five-layer model led to the development a hybrid algorithm (GAGNCG), combining the aforementioned techniques, for inverse modeling of one-dimensional magnetotelluric data. The GAGNCG modeling of the synthetic data performs more efficiently than the local methodology in terms of both procedure and results. This showed that the hybridization procedure maximized the advantages of using the global search methodology and minimized the disadvantages of the local technique. Based on these results, we developed another hybrid methodology (GA2D), built from some characteristics of the genetic algorithm and the simulated annealing method, for the inverse modeling of two-dimensional magnetotelluric data. The results were satisfactory, and the GA2D algorithm was a good starting point for the inverse modeling of two-dimensional data.

Keywords Magnetotelluric · Inversion · Hybrid genetic algorithms

Introduction

The magnetotelluric (MT) method is based on the fundamental works of Tikhonov (1950) and Cagniard (1953). Traditionally, MT surveys focus primarily on the detection of stratified media, conductive zones in the crust and upper mantle, and recognition of deep faults. Its interpretation may provide information on porosity, permeability, graphitization, fluid regimen, mineralization of groundwater, rheological properties, and thermodynamic geodynamical processes within the Earth.

The sources of the MT electromagnetic (EM) signal are natural, and its frequency bandwidth is: 10^{-4} Hz $\leq f \leq 10^{+4}$ Hz. The amplitude, phase, and the directional relationship between the magnetic field **H** and the electric field **E** on the surface depend on the distribution of the electrical conductivity of the subsurface.

✉ Joelson da Conceição Batista
jdbatista@ufba.br

[1] Institute of Geosciences of the Federal University of Bahia, # 265-E, University Campus of Ondina, Salvador, Bahia, Brazil

Interpretation of MT is based on the complex task of establishing, via inverse numerical modeling, a geophysical model that satisfies two requirements: (1) meets the set of acquired data according to previously established criteria; and (2) conforms to geological interpretation according to known information. This requires the use of numerical techniques to solve the nonlinearity of the inverse problem, to estimate the spatial distribution of ground resistivity.

Important local and global algorithms in MT modeling are, respectively, Gauss–Newton (GN) and Conjugate Gradient (CG) and Genetic Algorithms (GA) and Simulated Annealing (SA). Optimization of these techniques has been used in inversion of 1D geophysical data by Whittall and Oldenburg (1992), Tarantola and Valette (1982), Sen et al. (1993), Chunduru et al. (1995), Rothman (1985) and Sen and Stoffa (1995). GA emerged in the 1960s from the ideas of evolutionary computing. Holland (1975) and Goldberg (1989) contain a description of it, and Sen and Stoffa (1995) report two applications. Some classic references in 2D MT inverse modeling are Rodi and Mackie (2001), de Groot-Hedlin and Constable (1990) and Siripunvaraporn and Egbert (2000), in 3D are Siripunvaraporn and Egbert (2009) and Kelbert et al. (2014). In both cases, the model

is a fundamental tool and one of the approaches taken in its solution is the finite element (FE) method.

The motivation and objective of this paper is to evaluate the efficiency of the integration of local with global algorithms vis–vis of inverse modeling based on only one of them. Few papers report hybridization. Among them, Porsani et al. (1993) integrate the genetic algorithm with a linearized inversion scheme to develop a new approach to seismic waveform inversion, Santos et al. (2005) propose and evaluate an inversion procedure, which consists of a robust two-step inversion of complex magnetotelluric apparent resistivity data, and Ferreira et al. (2003) propose a hybrid genetic–linear algorithm for 2D inversion of sets of vertical electrical sounding. This work presents a new implementation for the inversion of MT data of complex resistivity for the interpretation of two-dimensional structures, developed from the association between GA, SA, and GN techniques.

Two algorithms were developed for the 1D inversion: one based on the gradient method of local optimization, the program GNCG, and the other based on combining these methods with genetic algorithms, the GAGNCG. After analyzing the results of the 1D algorithms and evaluating the efficiency of the two algorithms, we have decided to implement a 2D inversion program which includes the genetic characteristics modified by features of the simulated annealing—the program GA2D—and applied it to 2D modeling of synthetic and real data. We have used the Runge–Kutta standard approach of the FE instead of the adaptive mesh employed by Travis and Chave (1989) in which the nodes are part of the solution.

We analyze the GA2D effectiveness for the inversion of complex MT synthetic and field data, and it proved to be quite efficient. Although slower than the algorithms traditionally employed in the inversion of MT data, it does not require any a priori knowledge, which can be of great validity when considering an investigation of little known areas.

Bases of the MT geophysical method

The homogeneous Helmholtz wave equations for the EM field components: electric field, **E**, and magnetic field, **H**, are given by Ward and Hohmann (1988):

$$\nabla^2\mathbf{E} + k^2\mathbf{E} = 0 \tag{1}$$

$$\nabla^2\mathbf{H} + k^2\mathbf{H} = 0, \tag{2}$$

and k is the wave number, $k^2 = \mu_0\varepsilon_0\omega^2 - i\mu_0\sigma\omega$. For a good conductor $\sigma >> \epsilon\omega$ and so $k^2 \approx -i\mu_0\sigma\omega$.

Equations 1 and 2 are derived from Maxwell's equations in homogeneous media. The parameters μ_0 and ε_0 are known, respectively, as the free-space magnetic permeability and dielectric permittivity and σ is the electrical conductivity

For a homogeneous half-space, 1D case, and the electric field, **E**, and magnetic, **H**, fields along the northern (x) and east (y) directions, respectively, Maxwell's equations are given by

$$\frac{\mathrm{d}E_x}{\mathrm{d}z} = i\omega\mu_0 H_y \tag{3}$$

and

$$\frac{\mathrm{d}H_y}{\mathrm{d}z} = -\sigma E_x \tag{4}$$

in the absence of displacement currents. In case the directions of E and H are interchanged, we have:

$$\frac{\mathrm{d}E_y}{\mathrm{d}z} = -i\omega\mu_0 H_x \tag{5}$$

and

$$\frac{\mathrm{d}H_x}{\mathrm{d}z} = \sigma E_y. \tag{6}$$

For propagation in a 1D half-space, the complex impedance $Z(\omega)$ is scalar and given by:

$$Z(\omega) = \frac{E_x(\omega)}{H_y(\omega)} = -\frac{E_y(\omega)}{H_x(\omega)}. \tag{7}$$

because there is no preferred orientation between the two EM field components.

However, the complex impedance becomes a tensor in the 2D and 3D cases. So, $\tilde{\mathbf{Z}}(\omega)$ is, in general, a 3×3 matrix and

$$\mathbf{E}(\omega) = \tilde{\mathbf{Z}}(\omega)\mathbf{H}(\omega) \tag{8}$$

or

$$\mathbf{H}(\omega) = \tilde{\mathbf{Y}}(\omega)\mathbf{E}(\omega), \tag{9}$$

such that $\tilde{\mathbf{Y}}(\omega)$ represents the inverse matrix of $\tilde{\mathbf{Z}}(\omega)$.

We have two separated conditions for the 2D case: TM (E_x, E_z, H_y) and TE (E_y, H_x, H_z). For the TM condition, $E_x = Z_{xy}H_y$ and $E_z = Z_{zy}H_y$. For the TE condition, $H_x = Y_{xy}E_y$ and $H_z = Y_{zy}E_y$.

Instead of the impedance tensor, we usually employ two derived functions: the modulus ρ_a and the phase ϕ of the complex impedance in MT modeling and interpretation. They are defined for the 2D case as:

$$\rho_{a,ij} = \frac{1}{\omega\mu_0}\left|\frac{E_i(\omega)}{H_j(\omega)}\right|^2, \tag{10}$$

$$\phi_{ij} = \arctan\left(\frac{E_i(\omega)}{H_j(\omega)}\right), \tag{11}$$

$i = x,\ y, j = x,\ y$. Another important concept is the depth of research, or skin depth (δ). It is the depth at which the amplitude of the electromagnetic wave decays to $1 / e$ ($\approx$ 37%) of its value at the subsurface. A valid formula of δ for a plane wave is:

$$\delta = 0.503\sqrt{\frac{\rho}{f}}\,\text{km}. \tag{12}$$

The frequency f is given in hertz and the resistivity ρ is given in Ωm. Further information is detailed by Stratton (1941).

Fundamentals of the algorithms of the FE method

The finite element (FE) method was first developed in the 1950s with the works of Galerkin, Ritz, Courant and Hilbert. It is based on the construction of approximate solutions of differential equations to problems restricted to limited spaces. The use of the method in solving differential EM field equations involves dividing the total area into a finite number of elements. In each element, an approximation is obtained for the EM fields, performing arithmetic operations between the basic functions and the fields in each corner of the element.

We can separate the electric and magnetic fields in primary (E_p, H_p), representing the homogeneous parts of the medium, and secondary (E_s, H_s), created by the presence of nonhomogeneous regions of the model. The differential equations of the magnetic and electric field for this case are given by Batista and Porsani (1991):

$$\begin{aligned}\nabla^2\mathbf{E_s} + \nabla\left(\mathbf{E_s}\cdot\frac{\nabla\mathcal{Y}}{\mathcal{Y}}\right) - \mathcal{Z}\mathcal{Y}\mathbf{E_s} = \mathcal{Z}\Delta\mathcal{Y}\mathbf{E_p}\\ - \nabla\left(\mathbf{E_p}\cdot\frac{\nabla(\Delta\mathcal{Y})}{\mathcal{Y}}\right),\end{aligned} \tag{13}$$

$$\begin{aligned}\nabla^2\mathbf{H_s} + \mathcal{Y}(\nabla\times\mathbf{H_s})\times\nabla\left(\frac{1}{\mathcal{Y}}\right) - \mathcal{Y}\mathcal{Z}\mathbf{H_s} = \mathbf{Z}\Delta\mathcal{Y}\mathbf{H_p}\\ - \mathcal{Y}\nabla\left(\frac{\Delta\mathcal{Y}}{\mathcal{Y}}\right)\times\mathbf{E_p},\end{aligned} \tag{14}$$

where the parameters $\mathcal{Z} = i\omega\mu$ e $\mathcal{Y} = \sigma + i\omega\epsilon$, are, respectively, the impedance per unit of length and admittance per unit length of medium.

A simplified way to model fields using more complex 2D models must solve the differential Equations 13 and 14 subject to the appropriate boundary conditions. These equations can be written as follows:

$$\mathbf{Lu} = \mathbf{f}, \tag{15}$$

$\mathbf{L}$ is a differential operator, $\mathbf{u}$ is the electric or magnetic field to be determined, and $\mathbf{f}$ is the source.

For every element, we obtain an approximation of the field summing the product of the base functions with the value of the field at each node. The solution of Eq. 15 is achieved by the FE method which approximates the exact solution of this equation to a collection of basic functions ϕ_i defined in any area. Each basic function usually depends only on the spatial relationship of each domain to determine the approximate value of the solution.

The approximate solution of the field $\mathbf{u}$ is determined by the product of the sum of the basic function and the field in each node element,

$$u(r) = \sum_{i=1}^{n} \boldsymbol{\Phi}_i(r)u_i, \tag{16}$$

u_i is the field value in each node, and n is the total number of nodes of the element. So,

$$\sum_{i=1}^{n} L\boldsymbol{\Phi}_i\, u_i \ - f = \epsilon(r), \tag{17}$$

$\epsilon(r)$ is the approximation error. The minimization of the error function of Eq. 17 is accomplished by choosing a weight function $w_j(r)$ such that the inner product of this chosen function with $\epsilon(r)$ is equal to zero:

$$\langle w_j, \epsilon\rangle = \sum_{i=1}^{n}\langle w_j, L\boldsymbol{\Phi}_i u_i - f\rangle = 0. \tag{18}$$

The method of choice of $w_j(r)$ is called the Residual Galerkin method (Becker et al. 1981). Development of Eq. 18 with the Galerkin residual method results in the following matrix system of equations:

$$\begin{pmatrix} \int_r w_1 L\boldsymbol{\Phi}_1 \mathrm{d}r & \int_r w_1 L\boldsymbol{\Phi}_2 \mathrm{d}r & \cdots & \int_r w_1 L\boldsymbol{\Phi}_n \mathrm{d}r \\ \int_r w_2 L\boldsymbol{\Phi}_1 \mathrm{d}r & \int_r w_2 L\boldsymbol{\Phi}_2 \mathrm{d}r & \cdots & \int_r w_2 L\boldsymbol{\Phi}_n \mathrm{d}r \\ \vdots & \vdots & \cdots & \vdots \\ \int_r w_n L\boldsymbol{\Phi}_1 \mathrm{d}r & \int_r w_n L\boldsymbol{\Phi}_2 \mathrm{d}r & \cdots & \int_r w_n L\boldsymbol{\Phi}_n \mathrm{d}r \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix}, \tag{19}$$

which may take the form

$$\mathbf{Ku} = \mathbf{q}. \tag{20}$$

The elements $k_{j,i}$ of the matrix are

$$k_{j,i} = \int_r w_j L\boldsymbol{\Phi}_i \mathrm{d}r, \tag{21}$$

the right side of the system is

$$\mathbf{q} = (\mathbf{q_1}, \ldots, \mathbf{q_n})^\mathbf{T}, \tag{22}$$

and the solution vector is

$$\mathbf{u} = (\mathbf{u_1}, \ldots, \mathbf{u_n})^{\mathbf{T}}. \tag{23}$$

The system of Eq. 20 contains the solution at every element. The approximate solutions in the entire grid result from the addition of the contributions of the equations furnished by those elements. The elements are neighbors in the grid such that they share some of their nodes. Addition between equations allows interaction among all the elements of the grid and originates the expanded system of equations.

$$\mathbf{Ax} = \mathbf{b}. \tag{24}$$

The dimension of the matrix $\mathbf{A}$, is $n \times n$, n is the total number of nodes of the grid, and the dimension of the vectors $\mathbf{x}$ and $\mathbf{b}$ is $n \times 1$. Due to the disposition of the elements in the grid, the result of the combination between two nodes is always equal, regardless of the sense of the combination, what makes the matrix to be symmetrical. Since all the nodes are combined with each other, and the combination between two nodes of different elements equals zero, matrix $\mathbf{A}$ is square and sparse. We use FE and boundary conditions suitable for calculations of the E and H fields, TE and TM modes, as a direct methodology for the GA2D algorithm.

The inverse problem in MT

The analysis of MT data by an inverse modeling algorithm consists of estimating the true resistivity, thickness or depth values of the analyzed medium using the complex impedance or as in the present research, using simultaneously the values of apparent resistivity and phase data. For the one-dimensional case, the horizontally layered media are well known (Ward and Hohmann 1988). The free parameters of this model are the resistivity $\rho_i (1 \le i \le N)$ and the thickness $h_i (1 \le i \le N-1)$ of each layer, which are represented by the vector $\mathbf{m}$. In this case, we can evaluate the complex impedance in the form $Z(\mathbf{m}, f_j)$, where $f_j (1 \le j \le M)$ express the dependency with the frequency ($\omega_j = 2\pi f_j$). The two-dimensional model accounts for both lateral and vertical variations of resistivity. In this case, the complex impedance also depends on the position x where the measures are performed, and can be represented by $Z(\mathbf{m}, f_j, x)$.

Gauss–Newton method

Let $\zeta(f_j, x)$ represent a vector with the values of apparent resistivity and phase—module and phase of the complex impedance—observed at positions x and frequencies f_j, and $\zeta(\mathbf{m}, f_j, x)$ be the corresponding calculated values for model $\mathbf{m} = (m_1, \ldots, m_{2N-1})$. From the iterative scheme to minimize the objective function proposed by Gersztenkorn et al. (1986):

$$E(\mathbf{m}) = \sum_{j=1}^{N} \left| \zeta(f_j, x) - \zeta(\mathbf{m}, f_j, x) \right|^p, \tag{25}$$

where the parameter $1 \le p \le 2$. Note that $E(\mathbf{m})$ is the L_p norm of the error of the theoretical ζ to the power p. According to the basic Gauss–Newton method, we linearize $\zeta(\mathbf{m}, f_j, x)$ by Taylor's series about an estimate free parameter vector $\mathbf{m_k}$:

$$\zeta(\mathbf{m}, f_j, x) \approx \overline{\zeta}(\mathbf{m}, f_j, x) = \zeta(\mathbf{m_k}, f_j, x) - \sum_{i=1}^{N} \frac{\partial \zeta(\mathbf{m_k}, f_j, x)}{\partial \mathbf{m}} |_{\mathbf{m_k}} (\mathbf{m} - \mathbf{m_k}). \tag{26}$$

Substituting 26 into 25, we find a quadratic function of $\mathbf{m}$, whose minimum satisfies

$$(G_k^T R_k G_k) \Delta m = G_k^T R_k \Delta d_k, \tag{27}$$

where $d_{k,j} = \zeta(f_j, x) - \zeta(\mathbf{m_k}, f_j, x)$, $r_{k,j} = |\zeta(f_j, x) - \overline{\zeta}(\mathbf{m}, f_j, x)|^{p-2} (1 \le j \le M)$, $\Delta \mathbf{m} = (\mathbf{m_{k+1}} - \mathbf{m_k})$ and

$$\mathbf{G_k} = \begin{vmatrix} \frac{\partial \zeta(\mathbf{m_k}, f_1, x_1)}{\partial m_1} & \cdots & \frac{\partial \zeta(\mathbf{m_k}, f_1, x_1)}{\partial m_1 \partial m_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial \zeta(\mathbf{m_k}, f_M, x_M)}{\partial m_1} & \cdots & \frac{\partial \zeta(\mathbf{m_k}, f_M, x_M)}{\partial m_N} \end{vmatrix},$$
$$\Delta \mathbf{d_k} = \begin{vmatrix} d_{k,1} \\ \vdots \\ d_{k,M} \end{vmatrix}, \mathbf{R_k} = \begin{vmatrix} r_{k,1} & 0 \\ & \ddots \\ 0 & r_{k,M} \end{vmatrix}. \tag{28}$$

By using a regularization factor λ (Menke 1989), we compute the new solution $\mathbf{m_{k+1}}$ as,

$$m_{k+1} = m_k + (G_k^T R_k G_k + I)^{-1} G_k^T R_k \Delta d_k. \tag{29}$$

The conditions $p = 2$ and $\lambda = 0$ correspond to the plain least squares method. The row i of the sensitivity matrix G is k-weighted by the ith diagonal component of the matrix R_k, which is a function of the deviation between the observed ζ values and those computed from the current model m_k. The components of the sensitivity matrix are approximated by forward differences (Mcgillivray and Oldenburg 1990). We employ a conjugated gradient method to evaluate m_{k+1} from (14) and a harmonic measure of fitness (Porsani et al. 2000),

$$\phi(\mathbf{m_k}) = \frac{2 \sum_{i=1}^{N} [\zeta(f_j, x) \zeta(\mathbf{m}, f_j, x)]}{\sum_{i=1}^{N} [\zeta(\mathbf{m}, f_j, x)^2 + \zeta(f_j, x)^2]}, \tag{30}$$

the ratio ϕ varies within $[-1, 1]$ and approaches 1 as $\zeta(\mathbf{m}, f_j, x)$ approaches $\zeta(f_j, x)$.

We use the Gauss–Newton and conjugate gradient techniques together for solving the normal equations, as it may effectively solve nonlinear inverse problems. The GNCG algorithm uses the combination of these two methodologies.

Genetic algorithms

The genetic algorithms (GA) emerged in the 1960s from the ideas of evolutionary computing. Holland (1975) and Goldberg (1989) contain a description of the GA, and Ferreira et al. (2003) and Sen and Stoffa (1995) report two applications in geophysics. The GA employ the concepts of survival of the fittest, crossover, and mutation to generate a set of free parameter vectors that progressively approach field data. These methods fit into the class of global, probabilistic optimization methods, based on the principle of natural selection and genetics.

For inverse modeling of MT soundings, we start by selecting a set (or population) of parameter models. For case 1D, we defined the resistivity and number of layers of the model, and for case 2D we defined distinct regions and their respective resistivities. The second step is the search range setting for each model parameter (guide function):

$$\mathbf{m}_{k,j}^{\min} \leq \mathbf{m}_{k,j} \leq \mathbf{m}_{k,j}^{\max}, \quad 1 \leq j \leq n \ \text{ and } \ k = 0, \tag{31}$$

where $\mathbf{m}_{k,j}$ is the parameter vector and n is the number of parameters.

In the algorithms implemented in the GA we used, all the parameters of the models were encoded in binary form. Unlike the GAGNCG, the GA2D algorithm starts with a set of models generated from the interpretation of the inflection points of the apparent resistivity and phase data, derived from the complex impedance of the synthetic or real MT data, in a process similar to that of the interpretation of the electric probing curve vertical. The models are chosen for reproduction with a probability proportional to their fitness value, and pairs of models are selected at random and exchange parts of their binary chain.

The crossover points are selected at random, and all the bits to the right side are interchanged with a crossover probability, generating new models. To assure genetic variability in the population, a mutation process is adopted by changing at random a bit inside the binary chain based on a mutation probability. In this case, we chose to use an inverse dependence between the crossover probabilities (P_c) and mutation (P_m). We have made P_m high at the beginning of the iterations, decreasing with the evolution of the algorithm, with P_c having an inverse behavior. This will allow the substance to come out of possible global minima. The new set of models are accepted with an update probability by comparing them with the models in the previous generation. The process of selection, crossover, and mutation is applied until the fitness values converge, i.e., until the mean fitness approaches the highest fitness value in the population. We combine the genetic with the Gauss–Newton methods, generating a hybrid algorithm GAGNCG (Porsani et al. 1993). An implementation scheme of this algorithm is shown in Fig. 1.

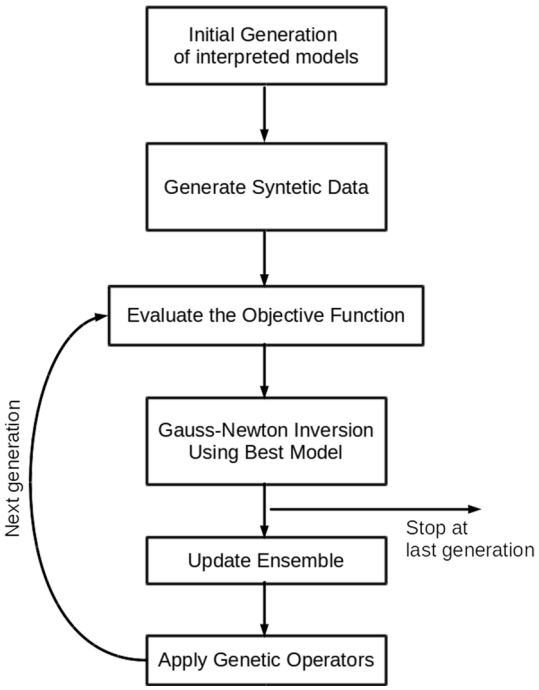


Fig. 1 Flow chart for the GAGNCG algorithm

Simulated annealing

Simulated Annealing (SA) is an optimization method that makes an analogy to the annealing process of metallurgy. Metropolis et al. (1953) introduced a simple numerical method that represents the state of equilibrium of a set of atoms at a given temperature. Let ΔE be the energy of a system of atoms at a temperature T. At each step of the algorithm, a random offset is given to an atom, which implies a new energy system, ΔE. If $\Delta E \leq 0$, a new displacement is accepted; otherwise ($\Delta E > 0$), the probability setup is accepted and will be given by the equation

$$P(\Delta E) = e^{\frac{-\Delta E}{K_{rmb} T}}, \tag{32}$$

K_b is the Boltzmann constant, and T is the absolute temperature in kelvin.

A uniformly distributed random number, r, must be generated in the range [0, 1]. If $r \leq P(\Delta E)$, the new setting is accepted. If $r > P(\Delta E)$, the previous configuration is used to start a new step. The choice of the probability function $P(\Delta E)$, as described above, is due to the fact that the system evolves according to a Boltzmann distribution.

The algorithm parameters are: the cost function, which is the system energy; the design variables that describe the actual configuration; and the temperature, which is a control parameter (Corana et al. 1987). If T has a much higher magnitude than the function of the standard deviation in the range, almost all points are accepted, whereas if T is equal

to zero the method becomes a random search of a minimum. So, it is adopted: T_i as the standard deviation of the value of the objective function in the range studied and T_f with the desired order of magnitude for the optimum accuracy. The SA approach is probabilistic: The SA requirement is not derived from information and is not affected by discontinuities and nonlinearities. We combine the genetic with the SA methods, generating a hybrid algorithm GA2D for bidimensional inverse of MT data.

Numerical results of the 1D inverse modeling

The basis of the GNCG algorithm is the combination of the Gauss–Newton method with the Conjugate Gradient technique. The GAGNCG uses the same methodology of the GNCG with the inclusion of a GA algorithm module (hybridization). Both GNCG and GAGNCG solve a system of equations relating the amplitude and the phase of a complex impedance. GAGNCG differs from GNCG in three aspects for estimating the parameters of the models: (1) generation of 100 initial models with each parameter within a range of minimum and maximum values determined by the aforementioned guide function; (2) insertion of the interpolation function between the mutation and crossover probabilities of the GA; and (3) simultaneous test and modification of the models of the population at every iteration. This second aspect required the definition of probability values equal to 0.7 for the initial crossover and 0.2 for the final mutation. So, the crossover probability decreases and the mutation probability increases as the number of generations increases.

A 1D modeling assumes a stratified model of the earth. So, it is necessary to estimate the resistivity (ρ) and thickness (ϵ) values of each layer of the model. The synthetic model that we have chosen to analyze represents a 5-layer medium with the following parameters from Ramos and Sampaio (1993): $\rho_1 = 200\,\Omega$m, $\rho_2 = 20\,\Omega$m, $\rho_3 = 200\,\Omega$m, $\rho_4 = 15\,\Omega$m, $\rho_5 = 1000\,\Omega$m, $\epsilon_1 = 10$ m, $\epsilon_2 = 200$ m, $\epsilon_3 = 1000$ m, and $\epsilon_4 = 4000$ m. Figure 2 displays the result of the inversion of the synthetic data employing, respectively, the modulus and the phase of the complex apparent resistivity. Because we know the true model for the synthetic data, we represent not only the adjustment of the data at each iteration, but also the fit between the current and the true model. We adopt different strategies to select the range of values of ρ and ϵ. In the present case, we have used a fixed range of 50% of the values of ρ and ϵ for creating the input parameters for the initial model of the GNCG and GAGNCG inversions. Tables 1 and 2 show the resulting model for each performed setting and their initial parameters. The fit of both inverted data, using the modulus and phase of the complex impedance, is excellent, giving an error of less than 1% for the GAGNCG inversion (Fig. 2). Figure 2 also contains the evolution of the data fitness maximum, medium and minimum of the GAGNCG inversion, showing the convergence of the parameters for each iteration.

Figure 3 shows that the GAGNCG final model, resulting from the modulus and phase of the complex impedance inversion, agrees with the synthetic model excellently. However, the final GNCG model presents differences between the true and computed values of the thickness of the two deeper layers larger than 1% and 10%, respectively (see Table 1). Those differences reflect the feature of local search of the

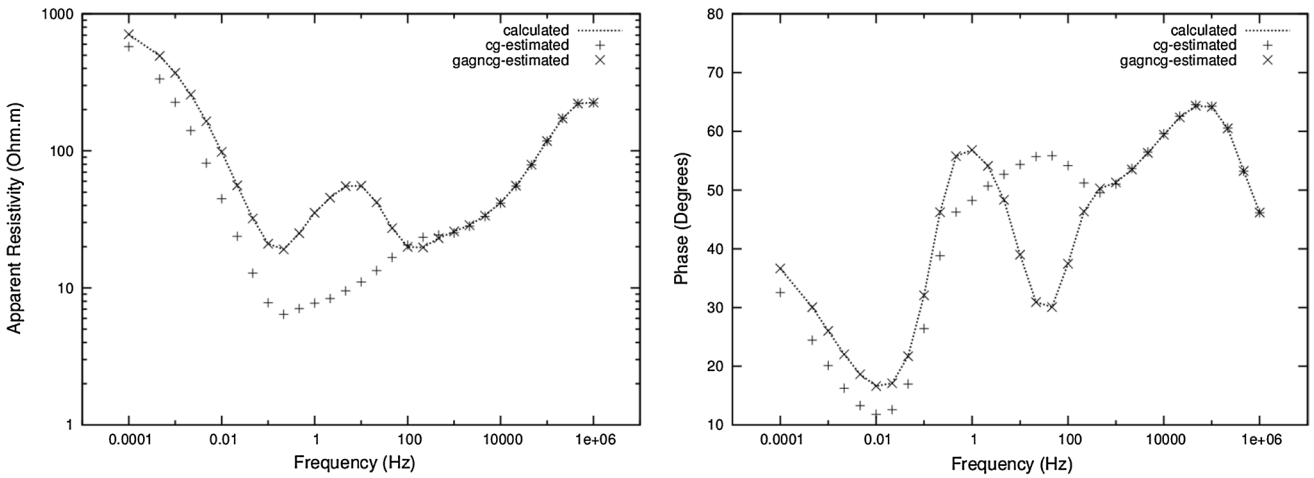


Fig. 2 Result of the joint inverse modeling of synthetic data of modulus and phase of the complex impedance. Left—Curves of the model studied and the result of the joint inversion using the GNCG and the GAGNCG algorithms. Right—Evolution of deviations of the inverted resistivity and phase by the GNCG and the GAGNCG algorithms and the maximum, medium and minimum data fitness showing the convergence of the parameters for each iteration

Table 1 Numerical comparison between the result of the GACG inversion and the respective 1D model

Physical properties	Exact model	Initial model	Final model	Percentage difference (%)
ρ_1 (Ωm)	200	100	200.1	− 0.001
ρ_2 (Ωm)	20	10	20.1	− 0.001
ρ_3 (Ωm)	200	100	185.3	− 0.15
ρ_4 (Ωm)	15	7.5	9.6	0.05
ρ_5 (Ωm)	1000	500	999.2	0.008
ϵ_1 (m)	10	5	9.9	0.001
ϵ_2 (m)	200	100	208.3	− 0.08
ϵ_3 (m)	1000	500	1160.6	− 1.60
ϵ_4 (m)	4000	2000	2518.2	14.82

Table 2 Numerical comparison between the result of the GAGNCG inversion and the respective 1D model

Physical properties	Exact model	Minimum parameter	Maximum parameter	Final model	Percentage difference (%)
ρ_1 (Ωm)	200	100	400	199.9	0.001
ρ_2 (Ωm)	20	10	30	20	0.0
ρ_3 (Ωm)	200	100	400	200.2	− 0.002
ρ_4 (Ωm)	15	7.5	30	15	0.0
ρ_5 (Ωm)	1000	2000	15	1000	0.0
ϵ_1 (m)	10	5	10	9.9	0.001
ϵ_2 (m)	200	100	400	200	0.0
ϵ_3 (m)	1000	500	2000	999.7	0.003
ϵ_4 (m)	4000	2000	8000	4001.4	− 0.01

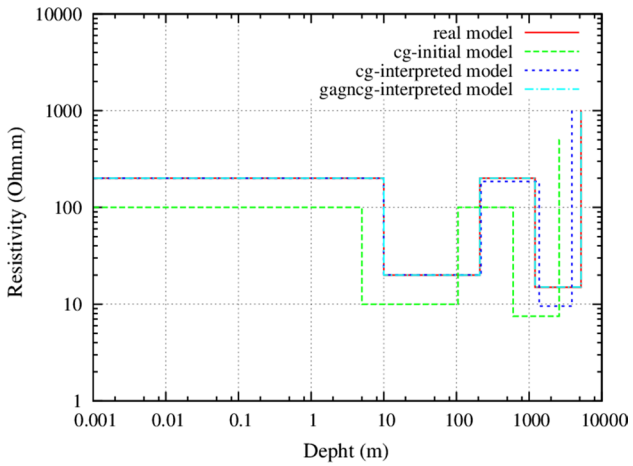


Fig. 3 Comparison between the resistivity-depth models for the synthetic data and those obtained with the GNCG and the GAGNCG inversion algorithms shown in Fig. 2

GNCG method, which tends to local minimum values, since due to its more global nature, GA has helped the GAGNCG to better exploit global minimum values.

Numerical results of the 2D inverse modeling

Synthetic data

We consider the model of a semi-infinite vertical fault and the outcropping vertical dike in Figs. 4 and 5. For the inverse modeling of the mentioned synthetic models, we used a rectangular mesh and uniform spacing, with dimensions $-5000\,\text{m} \le x \le 5000\,\text{m}$ and $0\,\text{m} \le z \le 4000\,\text{m}$, with 45 frequency values ranging from 10–2 Hz to 103 Hz in 24 measurement stations and 45 depth values. Therefore, the initial mesh used in GA2D has dimension of 24×45, being necessary operations to estimate 1080 parameters or values of resistivity in the nodes of that mesh. The GA2D developed in this research uses characteristics of genetic algorithms cited and some features of the SA method. These new implementations have been developed to replace the local methodology used in the 1D inversion of GAGNCG. For, even though it shows better outcomes for investments under the 1D condition, it presents a high computational cost due to the calculation of the sensitivity matrix in the 2D MT case. There are four main characteristics of the GA2D:

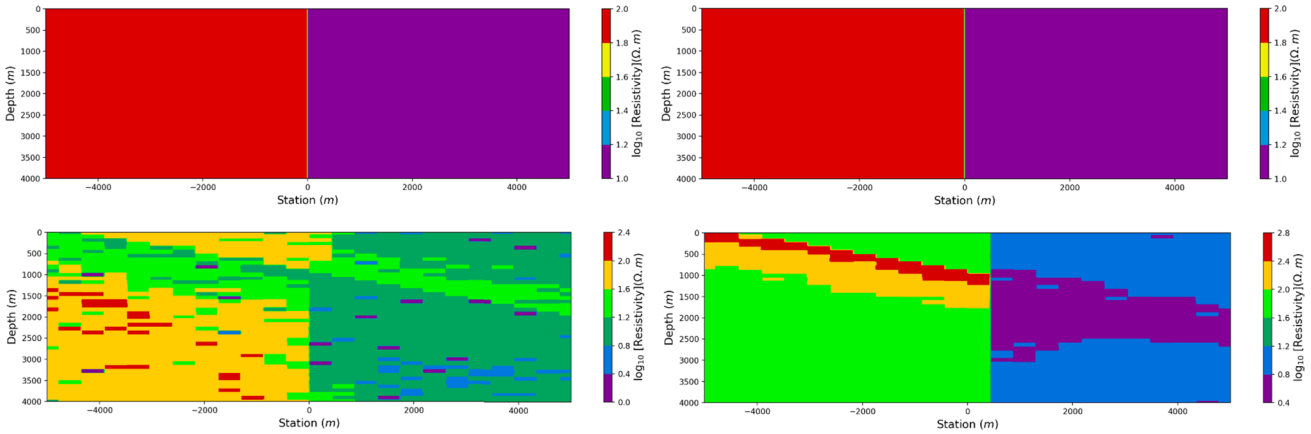


Fig. 4 Sections of the 2D inversion for the vertical fault model. Left—TE mode: the vertical fault model (top) and section corresponding to the interpreted model (bottom). Right—TM mode: the vertical fault model (top) and section corresponding to the interpreted model (bottom)

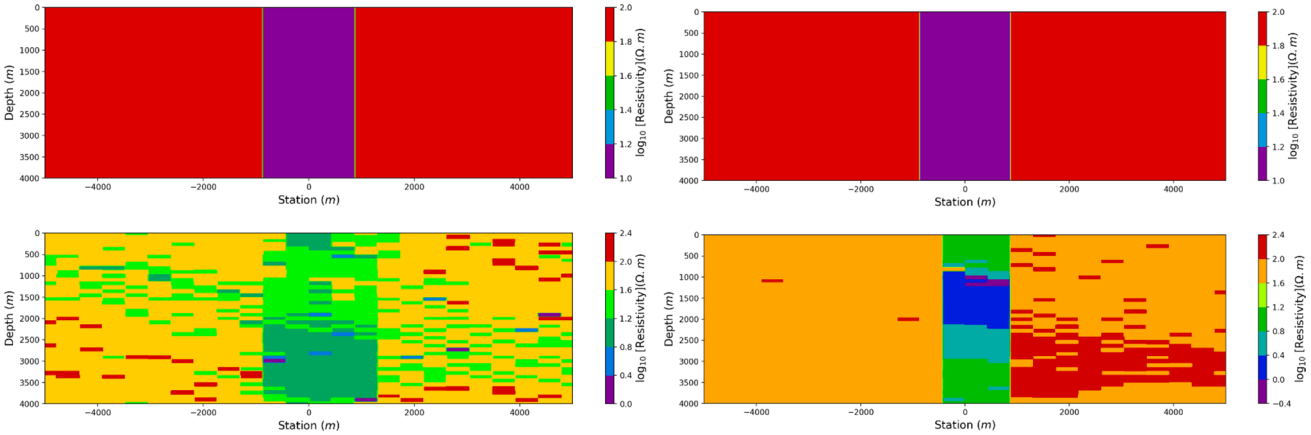


Fig. 5 Sections of the 2D inversion for the vertical fault model. Left—TE mode: the outcropping vertical dike model (top) and section corresponding to the interpreted model (bottom). Right—TM mode: the outcropping vertical dike model (top) and section corresponding to the interpreted model (bottom)

- Insertion of the probability function of the SA in the elitism modulus of the GA.
- Substitution of the guide function of the GAGNCG by the routine that interprets an initial set of 100 models based on the inflection points of the observed data curves to yield the initial model.
- Concomitant use of real values of amplitude and phase of the complex impedance to estimate the parameters of the model.
- Insertion of the interpolation function between the mutation and crossover probabilities of the GA modulus with the same consequences of the GAGNCG.

The GA2D does not start with the random generation of populations within a polling interval. Here, models or initial populations are interpreted automatically by the algorithm using the curves of modulus and phase of the calculated or observed complex impedance. The process begins with a subroutine identifying the inflection point in the phase curves, which delimits the passages with environment changes in the resistive mesh. Then, based on the minimum and maximum values of the apparent resistivity curves of these ranges, the subroutine estimates the resistivity values of the nodes belonging to each environment, as follows: $\rho = \rho_a^{\min} + r * \Delta, 0 \leq r \leq 1$, and $\Delta = \rho_a^{\max} - \rho_a^{\min}$.

After making this estimate, we continue with the calculation of fitness for each population created. The population or initial model is the group of true resistivity of each grid point that idealizes the modeled medium. GA2D also uses the linear relationship to define the probabilities of crossover operators and variable mutation, followed by genetic operations, which is applied to initially interpreted models. Thus, we computed the new fitness of the resulting population for this process, which constituted the elitism both of the

initially interpreted population and of the population interpreted after the genetic operations. The elitism process uses elements of the simulated annealing, because the probability for the acceptance of a new population depends on the probability $P(\Delta E)$. This first cycle ends with the selection of the fittest models with their skills.

The second cycle starts with the generation of the overpopulation. This concept applies to the cloning operation of the best models resulting from the previous cycle. After that, a second set of genetic operations acts upon this overpopulation. This transaction increases the probability of the inclusion of good features to the resulting models of this process, because there will be a selection of the fittest models of that overpopulation after these operations. A second elitism acts, then, among the population derived from overpopulation and the best people selected from the previous cycle. At the end of this second cycle, there will be a population of parameters which will bring together the feature of the best *fitness*.

In a third cycle, we compute both the current and the previous maximum populations, because the algorithm works with these two populations in parallel. Current population refers to a population that undergoes changes after the first two cycles. Previous population refers to the population interpreted initially during the first iteration and, subsequently, to the preservation of the best from the older populations. If the best model corresponds to the more apt or if the process reaches the maximum number of iterations, the algorithm ends.

However, to prevent premature convergence of the proposed algorithm, the algorithm compares the skills of the best models and updates an index. If the best previous population prevails for five generations (iterations), the algorithm creates a new interpreted population to substitute the best current population—followed by a new calculus of the skills and elitism of the current model population. Finally, it updates the best population parameters based on their fitness and heads to the resumption of the three cycles.

Figures 4, 5, 6, 7, 8, 9 and 10 display the results of the inverse modeling for TE and TM modes. On average, each model lasted two days to complete 200 iterations. Of course, more iterations could further improve the results. However, we did not pursue it further, because the results indicate a relatively satisfactory solution. This results show the sections with the best obtained inverse models, the section corresponding to the percentage deviation between observed and estimated models, and the pseudo-sections of the modulus and phase of the complex impedance. Figures 4, 6, 7 and 8 show the results for vertical fault. The model of Fig. 4 shows the largest discrepancies on the fault plane (see also Fig. 6). It is worth to note that the less discrepant values occur in the depth range between 0 and 1.5 km. Figures 5, 6, 9 and 10 show the results for the dike. The model of Fig. 5 shows the most significant errors in the dike area (see also Fig. 6). As in the fault model, the observed errors are concentrated on the interfaces between the media, what demonstrates the difficulty of the proposed method to recover the characteristics of the models in the presence of strong gradients.

Real data

This section illustrates the application of GA2D to real magnetotelluric data from 16 stations acquired in the Tucano basin, Bahia, Brazil. These stations constitute a band set of 48 frequency values in the broadband band, spaced 10 km. The mesh used by the GA2D was dimensioned with 16*x*48, a total of 768 values of resistivity to be estimated at the nodes of this mesh. We also assumed that this mesh has a regular spacing, where their minimum and maximum horizontal limits were defined by the

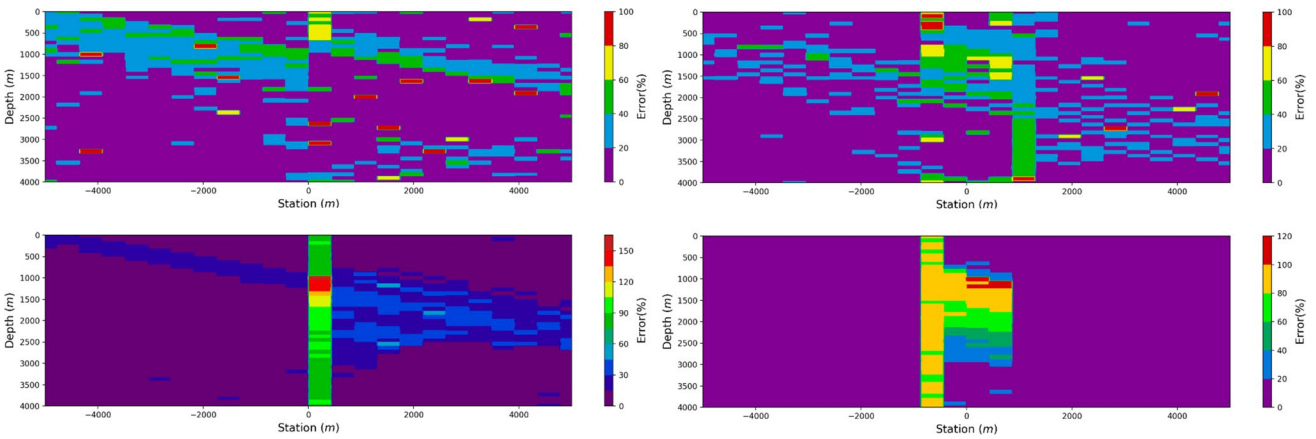


Fig. 6 Sections corresponding to the deviation between the observed and the estimated values for the synthetic models used. Left—Vertical fault model: TE mode (top) and TM mode (bottom). Right—Outcropping vertical dike model: TE mode (top) and TM mode (bottom)

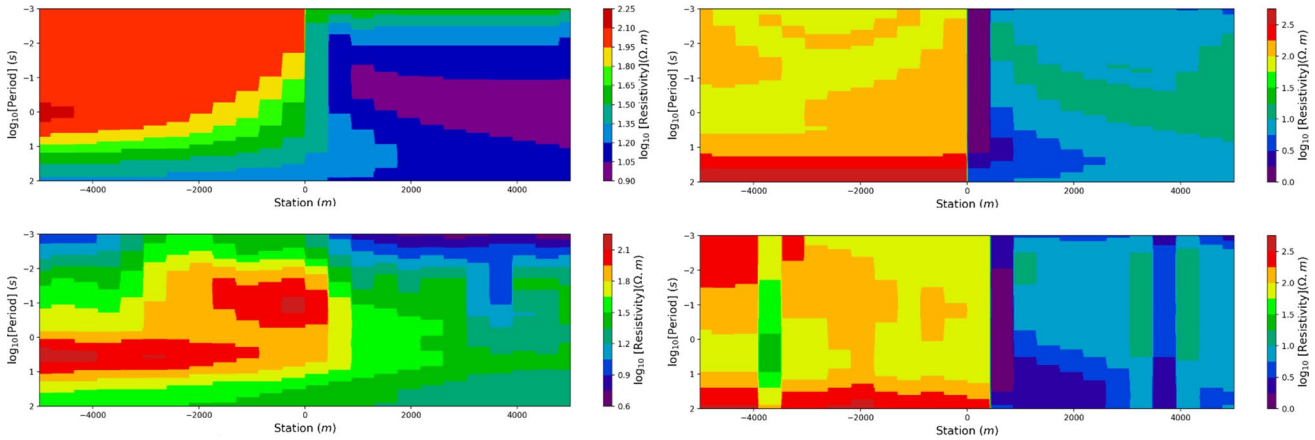


Fig. 7 Pseudo-sections of the modulus of the apparent resistivity as a function of period and horizontal distance for the vertical fault model. Left—TE mode: observed (top) and the estimated values (bottom). Right—TM mode: observed (top) and the estimated values (bottom)

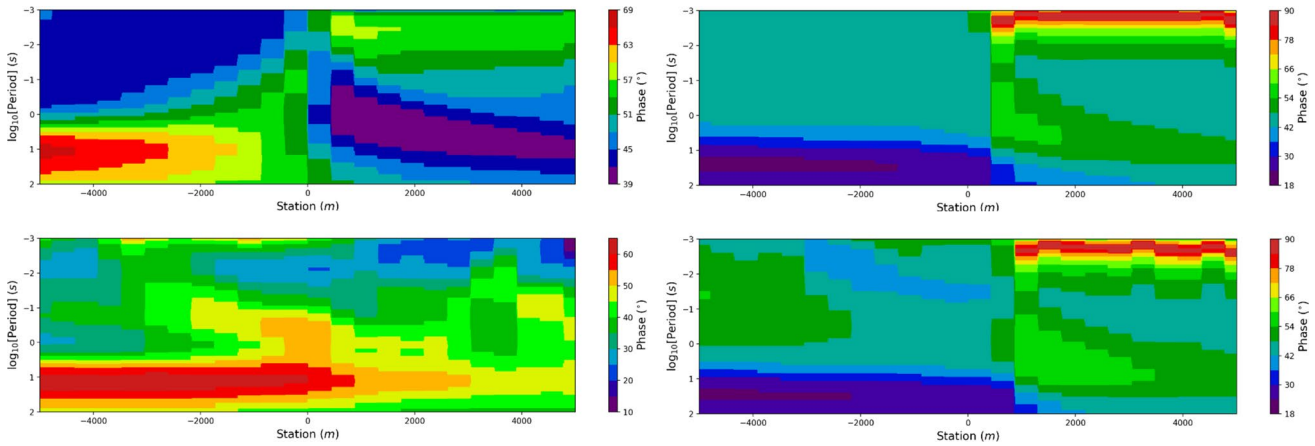


Fig. 8 Pseudo-sections of the phase of the apparent resistivity as a function of period and horizontal distance for the vertical fault model. Left—TE mode: observed (top) and the estimated values (bottom). Right—TM mode: observed (top) and the estimated values (bottom)

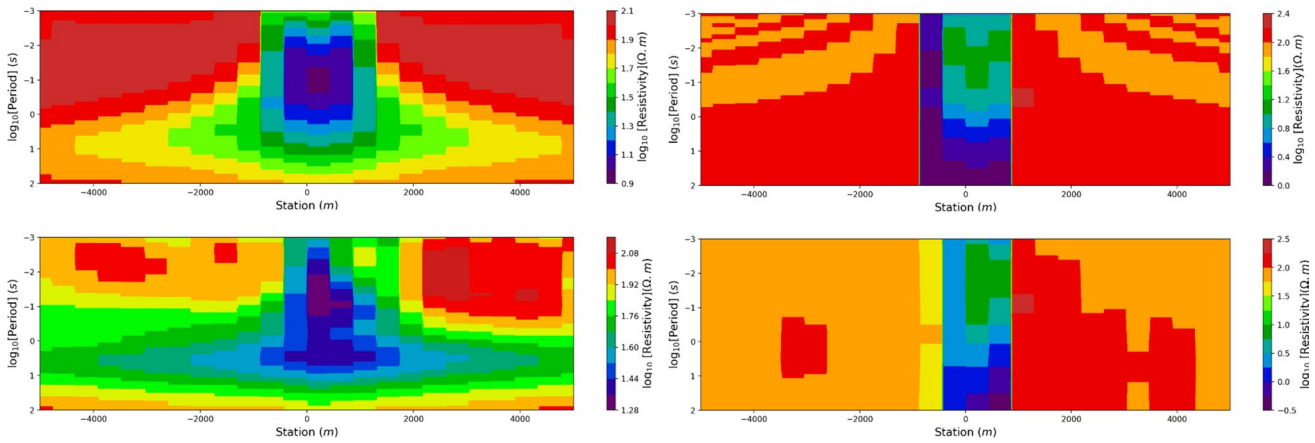


Fig. 9 Pseudo-sections of the modulus of the apparent resistivity as a function of period and horizontal distance for the outcropping vertical dike model. Left—TE mode: observed (top) and the estimated values (bottom). Right—TM mode: observed (top) and the estimated values (bottom)

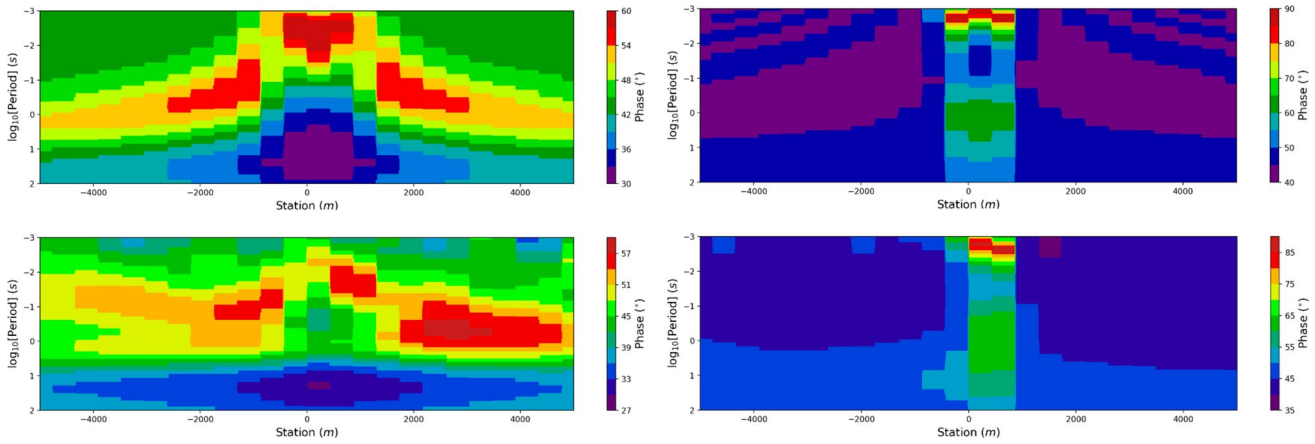


Fig. 10 Pseudo-sections of the phase of the apparent resistivity as a function of period and horizontal distance for the outcropping vertical dike model. Left—TE mode: observed (top) and the estimated values (bottom). Right—TM mode: observed (top) and the estimated values (bottom)

coordinates of the first and last measurement stations, while its maximum vertical limit was defined by the skin depth. The skin depth value (Eq. 2) was defined by the lowest values of frequency of the MT survey and apparent resistivity.

Figure 11 illustrates the best result of data inversions of a set of 200 generated models, and Fig. 12 shows the relative deviations between observed and predicted resistivity values of this interpreted model, confirming the coherence of the interpretative process.

A brief interpretation of the geoelectric section, shown in Fig. 11, illustrates the vertical and lateral boundaries of the Tucano Basin (black dashed line). The depth of this basin was interpreted to be about 5–7 km along two tracks: between stations 111 and 108, located about 40 km from the extreme *N* of the profile and between stations 201 and 205 at the south edge. On the other hand, the central part of the section—between stations 108 and 201—marks the asymmetry of this basin, with the presence of a 30 km wide structural high between two depocenters, indicating a depth varying between 10 and 15 km. Batista (2013) contains further details of the geology of the area and a more complete geophysical interpretation of these data.

Conclusions

This paper presents and analyzes implemented 1D and 2D algorithms for solving inverse modeling problems. The 1D inversion algorithms have been developed with the gradient technique (GNCG), genetic algorithm and a combination (hybridization) of these two approaches (GAGNCG). The 2D inversion has been developed based on the results of the 1D inversion applied to synthetic data and combines genetic algorithms and simulated annealing (GA2D).

The implementation and application of the inverse 1D methodology to synthetic data were important to better understand the advantages and disadvantages of these two techniques. The results showed that the hybrid technique (GAGNCG) was more efficient than the local minimization (GNCG) and, therefore, served as a basis for the development of the 2D inversion proposed algorithms.

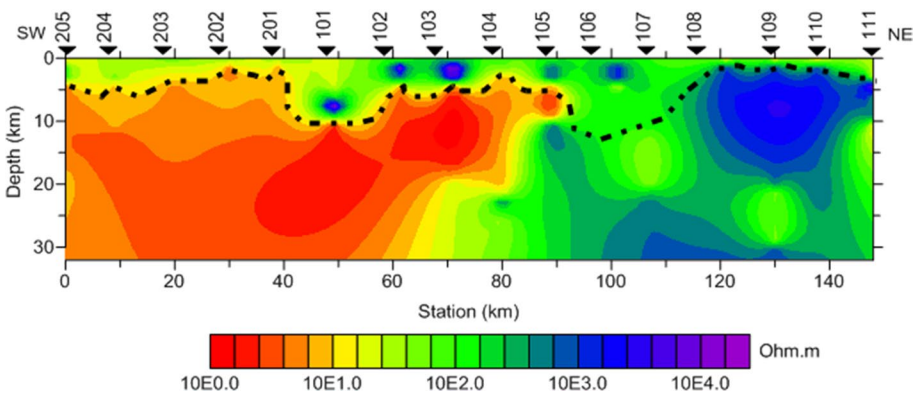


Fig. 11 Geoelectric profile model generated from the inverted models by using the GA2D, TE mode

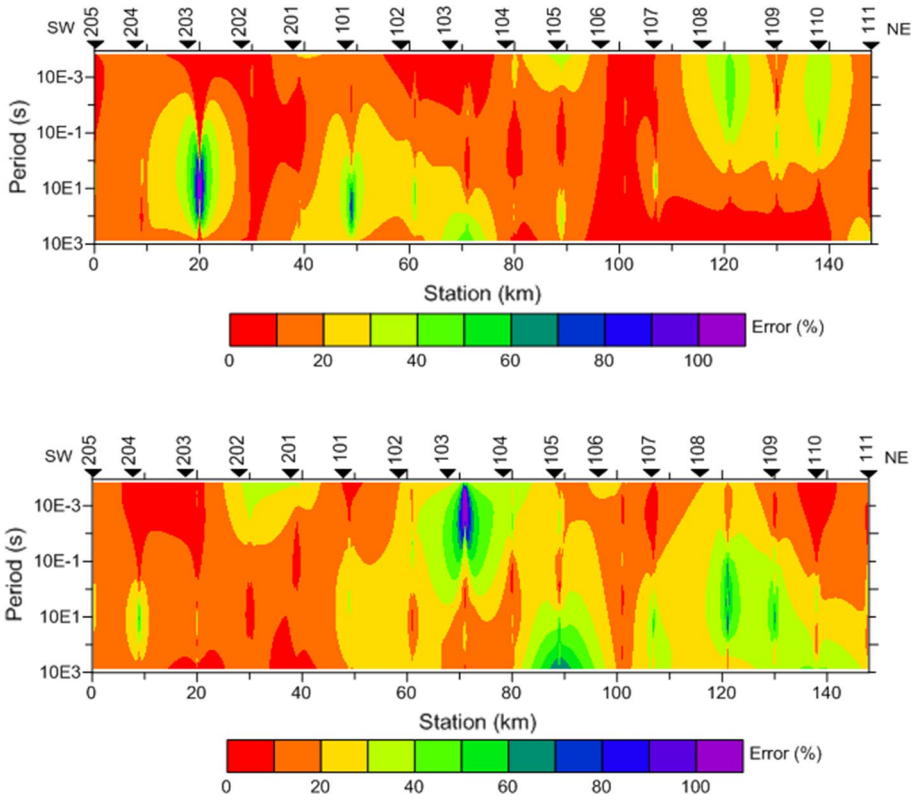


Fig. 12 Pseudo-section of percent relative error of the predicted apparent resistivity and phase values with respect to the observed apparent resistivity and phase values, TE mode

The interpretive scheme in the GA2D for automatic construct of initial models, associated with the probabilistic virtue of the simulated annealing in the elitism model, replaced the local methodology presented in GAGNCG and overcame the difficulty of circumventing the problem of defining a medium made up of many parameters. The selection of the estimated models prevented, therefore, the premature convergence of the algorithm, which is a peculiarity of the GA.

We initially applied the GA2D algorithm to two synthetic models, the fault and vertical dike. This technique showed reasonable results in the definition of true resistivity values and geometry of the modeled bodies, but it required a large processing time. The application of GA2D to real data acquired over the Tucano Basin, Brazil, allowed to generate a succinct interpretation of this basin, limiting it both vertically and horizontally, and demonstrating the applicability of the proposed algorithm. Therefore, the use of GA2D even with the amount of parameters estimated, for both synthetic and real data inversion, showed good results and can be improved with future implementations.

Acknowledgements We acknowledge our fellowships from CNPq (Brazilian National Council of Research). We also thank Dr. L. S. Batista for providing useful information about the FE method and the inversion codes.

References

Batista JC (2013) Modeling and Interpretation of Magnetotelluric Data in the Tucano Basin. Doctoral Thesis, Federal University of Bahia, Salvador (**in Portuguese**)

Batista LS, Porsani MJ (1991) Technical optimization of finite element modeling for electromagnetic geophysics. In: 2nd Brazilian Geophysical Congress, vol 1, Salvador, Bahia, Brazil

Becker BE, Garey GF, Oden JT (1981) Finite elements. Prentice-Hall Inc, Englewood, p 258

Cagniard L (1953) Basic theory of the magneto-telluric method of geophysical prospecting. Geophysics 18:605–635. https://doi.org/10.1190/1.1437915

Chunduru R, Sen MK, Stoffa PL, Nagendra R (1995) Non-linear inversion of resistivity profiling data for some geometrical bodies. Geophys Prospect 43:979–1003. https://doi.org/10.1111/j.1365-2478.1995.tb00292.x

Corana A, Marchesi M, Martini C, Ridella S (1987) Minimizing multimodal functions of continuous variables with the simulated annealing algorithm. ACM Trans Math Softw 13:262–280. https://doi.org/10.1145/29380.29864

de Groot-Hedlin C, Constable S (1990) Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data. Geophysics 55:1613–1624. https://doi.org/10.1190/1.1442813

Ferreira NR, Porsani MJ, Oliveira SP (2003) A hybrid genetic-linear algorithm for 2D inversion of sets of vertical electrical sounding. Braz J Geophys 21:235–248. https://doi.org/10.1590/S0102-261X2003000300003

Gersztenkorn A, Bednard JB, Lines LR (1986) Robust iterative inversion for the one-dimensional acoustic wave equation. Geophysics 51:357–368

Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, New York, p 403

Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor, p 14

Kelbert A, Meqbel N, Egbert GD, Tandon K (2014) ModEM: a modular system for inversion of electromagnetic geophysical data. Comput Geosci 66:40–53. https://doi.org/10.1016/j.cageo.2014.01.010

Mcgillivray P, Oldenburg DW (1990) Methods for calculating Fréchet derivatives and sensitivities for non-linear inverse problem: a comparative study. Geophys Prospect 38:499–524

Menke W (1989) Geophysical data analysis: discrete inverse theory. Academic Press, New York

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH (1953) Equations of state calculations by fast computing machines. J Chem Phys 21:1087–1092. https://doi.org/10.1063/1.1699114

Porsani MJ, Stoffa PL, Sen MK, Chunduru R, Wood WT (1993) A combined genetic and linear inversion algorithm for seismic waveform inversion. In: SEG technical program expanded abstracts, Society of Exploration Geophysicists, pp 692–695, https://doi.org/10.1016/S0921-9366(06)80009-5

Porsani MJ, Stoffa PL, Sen MK, Chunduru RK (2000) Fitness functions, genetic algorithms and hybrid optimization in seismic waveform inversion. J Seism Explor 9:143–164

Ramos RGN, Sampaio EES (1993) A comparative study of the asymptotic techniques mostly employed in the interpretation of magnetotelluric soundings. Braz J Geophys 11:65–79

Rodi W, Mackie RL (2001) Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion. Geophysics 66:174–187. https://doi.org/10.1190/1.1444893

Rothman DH (1985) Nonlinear inversion, statistical mechanics and residual static estimation. Geophysics 50:2784–2796. https://doi.org/10.1190/1.1441899

Santos AB, Sampaio EES, Porsani MJ (2005) A robust two-step inversion of complex magnetotelluric apparent resistivity data. Studia Geophysica et Geodaetica 49(1):109–125

Sen MK, Stoffa PL (1995) Global optimization methods in geophysical inversion. Elsevier, Amsterdam, p 281

Sen MK, Bhattacharya BB, Stoffa PL (1993) Nonlinear inversion of resistivity sounding data. Geophysics 58(4):496–507. https://doi.org/10.1190/1.1443432

Siripunvaraporn W, Egbert G (2000) An efficient data-subspace inversion method for 2-D magnetotelluric data. Geophysics 65:791–803. https://doi.org/10.1190/1.1444778

Siripunvaraporn W, Egbert G (2009) WSINV3DMT: vertical magnetic field transfer function inversion and parallel implementation. Phys Earth Planet Inter 173:317–329. https://doi.org/10.1016/j.pepi.2009.01.013

Stratton JA (1941) Electromagnetic theory. MaGraw-Hill, New York, p 615

Tarantola A, Valette B (1982) Inverse problems = quest for information. J Geophys 50:159–170

Tikhonov A (1950) Determination of the electrical characteristics of the deep strata of the Earth's crust. Doklady 73:295–297

Travis BJ, Chave AD (1989) A moving finite element method for magnetotelluric modeling. Phys Earth Planet Inter 53:432–443. https://doi.org/10.1016/0031-9201(89)90028-9

Ward SH, Hohmann GW (1988) Electromagnetic theory for geophysical applications. Electromagn Methods Appl Geophys 1:130–311

Whittall KP, Oldenburg DW (1992) Inversion of magnetotelluric data for a one-dimensional conductivity. In: Fitterman DV (ed) Geophysics monograph series no. 5. Society of Exploration Geophysicists, Tulsa, Oklahoma

Acta Geophysica (2019) 67:1379–1390
https://doi.org/10.1007/s11600-019-00326-x

RESEARCH ARTICLE - APPLIED GEOPHYSICS

Preliminary study on wave field and dispersion characteristics of channel waves in VTI coal seam media

Guangzhong Ji[1] · Hui Li[2] · Jiuchuan Wei[3] · Sitong Yang[3]

Received: 17 January 2019 / Accepted: 6 July 2019 / Published online: 15 July 2019

Abstract

Coal seam is a sedimentary rock with bedding, which can be regarded as VTI medium. VTI medium model is more suitable for real coal seam. However, existing channel wave theories generally assume that coal seams are isotropic for mathematical simplicity, and there is no study on the properties of channel waves in VTI media. In this paper, we deduce the theoretical dispersion equation of Love channel waves in the three-layer VTI model and analyze the parameter effects on the dispersion curve for the first time. The channel wave field in VTI media is simulated by three-dimensional staggered-grid finite difference method. The results indicate that polarization of both qP- and qSV-waves is not parallel or perpendicular to the orientation of the wave traveling in VTI media, while the polarization of SH wave is normal to wave propagation direction at horizontal plane. Therefore, it is wise to use Love channel waves to conduct field exploration because of the uniqueness of dispersion curves in the VTI media for the Love channel waves comparing with that in isotropic media. The velocities of the Love channel wave in VTI media are higher than that in isotropic media. The coal seam thickness primarily influences the Airy frequency phase, while the Airy phase velocity remains stable. Both the S-wave velocity and γ parameter of coal seam significantly affect Airy phase velocity. Severe errors may occur during in the coal thickness inversion when using the dispersion curve of Love channel wave in isotropic media, and dispersion curve in VTI media should be adopted. In terms of the amplitude distribution, Love channel waves appear to have similar patterns in both the VTI media and the isotropic media.

Keywords Channel wave · Coal seam · VTI media · Frequency dispersion

Introduction

Minor abnormal structures (e.g., faults and collapse columns) found in coal seams can severely endanger mining safety, such as roof collapse, water invasion and gas leaks. To predict these potential threats and minimize mining risks, channel wave (in-seam wave) exploration (Dresen and Rüter 1994) is generally applied to detect abnormal structures in coal seams.

Due to mathematical complexities and computational limits, coal seams are usually assumed to be 2D in homogenous elastic models to investigate the propagation features of channel waves (Krey et al. 1982). Since we cannot obtain the analytical solution of a channel wave equation in a complicated coal seam model, numerical methods are usually applied to qualitatively analyze the propagation features of channel waves (Korn and Emmerich 1988). Improvements in computational capability enable us to simulate the channel wave field and thoroughly understand the propagation features in a complicated 3D model by using the finite difference algorithm (Essen et al. 2007; Yang and Cheng 2012; He 2017; Wang et al. 2017). Essen et al. (2007) numerically simulated the channel wave propagation in models with faults and variable coal bed thicknesses and the results suggested that a weak reflected channel wave is observed when there is an abnormal structure. However, a coal seam intersection is not reflected. Based on the high-order staggered-grid finite difference algorithm, a mirror method (Ji et al. 2012, 2018) was proposed to process the roadway model.

✉ Hui Li
huili16@xjtu.edu.cn

[1] School of Earth and Environment, Anhui University of Science and Technology, Huainan, Anhui, China

[2] The School of Electronic and Information Engineering, Xi'an Jiaotong University, Shaanxi, China

[3] College of Earth Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong, China

Yang et al. (2016) studied the seismic wave field ahead of roadways. The results suggest that Rayleigh channel waves, which have high energy concentrations, are more appropriate for the exploration of waves ahead of roadways, while the weak Love channel waves may not be an appropriate option. Li et al. (2015) applied finite difference method to numerically study the propagation properties of channel waves in a viscoelastic model.

Although coal seams usually appear to be anisotropic due to their mineralogical, joint surface, and depositional environment characteristics, there is still a lack of quantitative investigations that have considered such physical properties in coal layer models (Dong 2008). Buchanan et al. (1983) found that the velocity anisotropy can reach 14% by calculating azimuthal velocity dispersion curves. Therefore, they argued that an anisotropic coal layer could significantly affect the propagation features of channel waves, which will mislead us to predict abnormal structures. Liu et al. (1991) studied the effects of cracks on the Love channel wave propagation in a 2D EDA media and found that there is reasonable agreement between the theoretical estimations and actual records of the travel time, amplitude and dispersion properties of these waves. At present, the study of channel waves propagating in anisotropic coal seams is rare, especially in the three-dimensional channel wave field of anisotropic coal seams.

Dispersion is also vital for channel waves. Krey (1963) calculated theoretical dispersion curves in an elastic, isotropic and horizontal three-layer coal model. Using theoretical derivations and numerical calculations, Rader et al. (1985) obtained a numerical solution to the channel wave dispersion curves in an elastic, isotropic and multilayer coal model. This study also analyzed the parameter effects on dispersion curves. Yang et al. (2014) investigated both the dispersive curves and an elasticity property of Rayleigh channel waves in a multilayer model and claimed that both the coal bed thickness and the quality factor of the coal seam can primarily influence the quality factor of a Rayleigh channel wave. He et al. (2017) concluded that the normal mode of the Rayleigh channel wave is theoretically observed, while both fundamental mode and first-order channel waves are usually recorded in field records. Feng and Zhang (2017) extracted the dispersion curve of Love channel wave by high-precision S-transform method. Some scholars (Wang et al. 2012; Hu et al. 2018) use the inverse method of dispersion curve to detect coal thickness.

The coal seam is a sedimentary rock with a bedding in the direction of the coal seam, which can be regarded as a Vertical Transverse Isotropy (VTI) medium. The undeformed coal can generally be regarded as a VTI medium. Obviously, the VTI medium model is more consistent with the real situation of the coal seam than isotropic. In addition, coal seam in different areas has weak anisotropy, moderate or strong anisotropy (Wang et al. 2012; Morcote et al. 2010), which is more complex and diverse.

Based on Thomsen theory, this paper specially studies the properties of channel waves in weak anisotropic and VTI coal seam media. The similarities and differences in channel waves between the VTI media and isotropic media are analyzed emphatically. First, the constitutive dispersion equation for the Love channel wave propagating in a three-layer VTI model is theoretically deduced. Second, the parameter effects on dispersion curves are systematically analyzed. Third, the high-order staggered-grid finite difference algorithm is considered to numerically model propagating features of channel waves.

VTI medium theory

The transversely isotropic (TI) media have axial symmetry. If the media are symmetrical along the vertical axis, then these are referred to as the VTI media. The VTI media consist of periodic thin layers (Fig. 1), which are similar to sedimentary strata.

According to the general stress–strain relation, the VTI media have five independent elastic parameters:

$$\mathbf{C} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{pmatrix} \tag{1}$$

$$C_{66} = \frac{C_{11} - C_{12}}{2}$$

Despite the elastic matrix C determining the relation of stress and strain, its physical meaning is not apparent. A set of

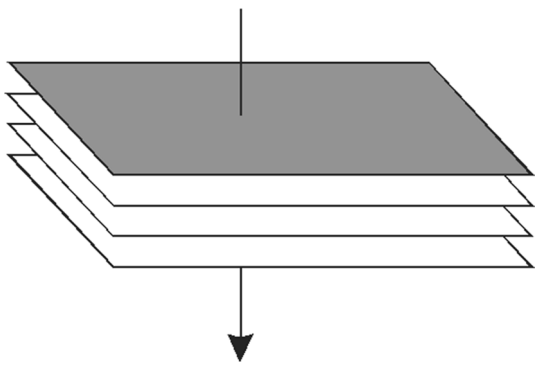


Fig. 1 VTI medium which has a vertical symmetry axis

anisotropic parameters are proposed to characterize the features of TI media to conduct theoretical investigations and field tests (Thomsen, 1986), and the following equations are used to characterize relations between the matrix C and Thomsen parameters of the VTI media:

$$\begin{aligned} V_P &= \sqrt{C_{33}/\rho} \\ V_S &= \sqrt{C_{44}/\rho} \\ \varepsilon &= \frac{C_{11}-C_{33}}{2C_{33}} \\ \gamma &= \frac{C_{66}-C_{44}}{2C_{44}} \\ \delta &= \frac{(C_{13}+C_{44})^2-(C_{33}-C_{44})^2}{2C_{33}(C_{33}-C_{44})} \end{aligned} \tag{2}$$

where ρ is density, V_p is the velocity of qP-wave propagating in the perpendicular orientation, V_s is either qSV-wave or pure SH wave velocity passing through the perpendicular direction. ε, γ and δ are Thomsen parameters. In the isotropic media, ε, γ and δ are 0. ε reflects the P-wave anisotropy; the higher ε is, the greater the anisotropy of the qP-wave. δ is the qP-wave anisotropy transfer coefficient. γ indicates the S-wave anisotropy. The characteristics of SH waves are described by V_s and γ.

Dispersion equation of Love channel waves in the three-layer VTI model

Since the Love channel wave is generally applied to detect the coal seam structures during the practical coal exploration, so we focus on the Love channel wave dispersion equation of VTI media, and take the commonly used three-layer horizontal medium as the model. VTI medium has vertical symmetrical axis, so the horizontal plane is isotropic, the vertical plane is anisotropic, and the vertical planes in all directions have the same anisotropic properties. We take the xoz plane which is formed by horizontal x-axis and vertical z-axis (Fig. 2) as an example to solve the dispersion curve of Love channel wave in VTI media.

Theoretical solution to the dispersion equation of Love channel waves

The SH wave equations in the VTI media can be written as:

$$\begin{aligned} \rho\frac{\partial^2 v}{\partial t^2} &= \frac{\partial \sigma_{xy}}{\partial x}+\frac{\partial \sigma_{yz}}{\partial z} = C_{66}\frac{\partial^2 v}{\partial x^2}+C_{44}\frac{\partial^2 v}{\partial z^2} \\ \sigma_{yz} &= C_{44}\frac{\partial v}{\partial z} \\ \sigma_{xy} &= C_{66}\frac{\partial v}{\partial x} \end{aligned} \tag{3}$$

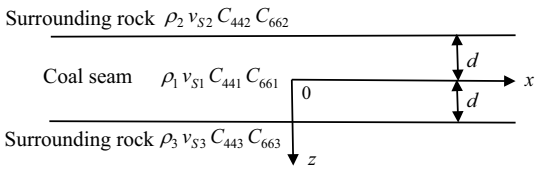


Fig. 2 Asymmetrical three-layer coal seam VTI model

The plane-wave solution to the dispersion equation is $v = v_0 e^{\beta z} e^{i\omega(t-x/c_L)}$ (Dresen and Rüter 1994). v is the displacement. v_0 is the displacement at zero time. β is the coefficient of the amplitude decaying exponentially with depth. c_L denotes the phase velocity of channel wave. ω is the circle frequency.

Figure 2 displays schematic of three-layer VTI model. Both the top and bottom layer media are the rocks. The density, vertical shear wave velocity and elastic parameters of the top rock are represented by ρ_1, v_{s1} and C_{441} and C_{661}, respectively. The density, vertical shear wave velocity and elastic parameters of the bottom rock are represented by ρ_3, v_{s3} and C_{443} and C_{663}, respectively. Meanwhile, the middle layer of three-layer VTI model shows the coal seam, and its respective parameters are ρ_2, v_{s2}, and C_{442} and C_{662}. The velocity and density of the coal seam are less than those of the surrounding rocks. The coal seam thickness is represented by $2d$. Furthermore, as shown in Fig. 2, the origin of coordinate assumes at the geometric center of the three-layer VTI model. The x-axis is along the boundary of three-layer VTI model, while the z-axis is perpendicular to that and oriented vertically downward.

The displacement of SH waves in the horizontal three-layer VTI model can be written as:

$$\begin{cases} v_1 = a_3 e^{\beta_1(z+d)} e^{i\omega(t-x/c_L)} & z < -d \\ v_2 = (a_1\cos\beta_2 z + a_2\sin\beta_2 z) e^{i\omega(t-x/c_L)} & -d \le z \le d \\ v_3 = a_4 e^{-\beta_3(z-d)} e^{i\omega(t-x/c_L)} & z > d \end{cases} \tag{4}$$

In Eq. (4), $e^{i\omega(t-x/c_L)}$ is the harmonic factor, and the left part of the harmonic factor is the displacement amplitude. a_1, a_2, a_3 and a_4 are the amplitude coefficients. β_1, β_2 and β_3 are the coefficients of the amplitude decaying exponentially with depth. If we substitute these equations into Eq. (3), then β can be obtained:

$$\begin{aligned} \beta_1 &= \frac{\omega}{\sqrt{C_{441}}}\sqrt{-\rho_1+\frac{C_{661}}{c_L^2}} \\ \beta_2 &= \frac{\omega}{\sqrt{C_{442}}}\sqrt{\rho_2-\frac{C_{662}}{c_L^2}} \\ \beta_3 &= \frac{\omega}{\sqrt{C_{443}}}\sqrt{-\rho_3+\frac{C_{663}}{c_L^2}} \end{aligned} \tag{5}$$

where c_L represents the phase velocity of the channel wave and β represents a positive real number; therefore, $\sqrt{\frac{C_{662}}{\rho_2}} \le c_L \le \min\left(\sqrt{\frac{C_{661}}{\rho_1}}, \sqrt{\frac{C_{663}}{\rho_3}}\right)$. By combining these equations with Eq. (2), we can obtain $v_2\sqrt{1+2\gamma_2} \le c_L \le \min(v_1\sqrt{1+2\gamma_1}, v_3\sqrt{1+2\gamma_3})$ which shows the velocity range of c_L.

Both the particle motion displacement and stress at the interfaces ($z=-d$ and $z=d$) should satisfy following boundary conditions:

$$\begin{cases} v_1\big|_{z=-d} = v_2\big|_{z=-d} \\ v_2\big|_{z=d} = v_3\big|_{z=d} \\ (\sigma_{yz})_1\big|_{z=-d} = (\sigma_{yz})_2\big|_{z=-d} \\ (\sigma_{yz})_2\big|_{z=d} = (\sigma_{yz})_3\big|_{z=d} \end{cases} \tag{6}$$

According to Hooke's law, $\sigma_{yz} = C_{44}\varepsilon_{yz} = C_{44}\frac{\partial v}{\partial z}$; if this equation is substituted into Eq. (6), then the equation of amplitude coefficient is obtained:

$$\begin{cases} -\cos(\beta_2 d)a_1 + \sin(\beta_2 d)a_2 + a_3 = 0 \\ -C_{442}\beta_2\sin(\beta_2 d)a_1 + C_{442}\beta_2\cos(\beta_2 d)a_2 + C_{443}\beta_3 a_4 = 0 \\ -C_{442}\beta_2\sin(\beta_2 d)a_1 - C_{442}\beta_2\cos(\beta_2 d)a_2 + C_{441}\beta_1 a_3 = 0 \\ \cos(\beta_2 d)a_1 + \sin(\beta_2 d)a_2 - a_4 = 0 \end{cases} \tag{7}$$

if a_1, a_2, a_3 and a_4 are not equal to zero, then the coefficient determinant needs to be zero:

$$\begin{bmatrix} -\cos(\beta_2 d) & \sin(\beta_2 d) & 1 & 0 \\ -C_{442}\beta_2\sin(\beta_2 d) & C_{442}\beta_2\cos(\beta_2 d) & 0 & C_{443}\beta_3 \\ -C_{442}\beta_2\sin(\beta_2 d) & -C_{442}\beta_2\cos(\beta_2 d) & C_{441}\beta_1 & 0 \\ \cos(\beta_2 d) & \sin(\beta_2 d) & 0 & -1 \end{bmatrix} = 0 \tag{8}$$

Two solutions can be obtained:

$$\tan(\beta_2 d) = \frac{C_{443}\beta_3 C_{441}\beta_1 - C_{442}^2\beta_2^2 + \sqrt{(C_{442}^2\beta_2^2 + C_{443}^2\beta_3^2)(C_{442}^2\beta_2^2 + C_{441}^2\beta_1^2)}}{C_{442}\beta_2(C_{441}\beta_1 + C_{443}\beta_3)} \tag{9}$$

$$\tan(\beta_2 d) = \frac{C_{443}\beta_3 C_{441}\beta_1 - C_{442}^2\beta_2^2 - \sqrt{(C_{442}^2\beta_2^2 + C_{443}^2\beta_3^2)(C_{442}^2\beta_2^2 + C_{441}^2\beta_1^2)}}{C_{442}\beta_2(C_{441}\beta_1 + C_{443}\beta_3)} \tag{10}$$

If the top and bottom surrounding rock are identical, then the solutions of the Love channel wave are:

$$\tan(\beta_2 d) = \frac{C_{441}\beta_1}{C_{442}\beta_2} \tag{11}$$

$$\tan(\beta_2 d) = -\frac{C_{442}\beta_1}{C_{441}\beta_2} \tag{12}$$

Equation (12) can be changed to:

$$\tan(\beta_2 d - \frac{\pi}{2}) = \frac{C_{441}\beta_1}{C_{442}\beta_2} \tag{13}$$

Obviously, Eqs. (11) and (13) can be unified into one equation by taking the arctangent function:

$$\beta_2 d = \arctan\frac{C_{441}\beta_1}{C_{442}\beta_2} + \frac{n\pi}{2}, \quad n = 0, 1, 2\ldots \tag{14}$$

When n is 0, it is called the fundamental mode dispersion curve, and when n is 1, it is called the first-order dispersion curve, and so on.

Furthermore, we can solve the amplitude distribution in depth of Love channel wave. Substituting the values on the dispersion curve (Eq. (9)) into Eq. (7) and letting the coefficient $a_1 = 1$, the other three coefficients a_2, a_3, a_4 can be solved. Taking the coefficients into Eq. (4), then the amplitude distribution equation of Love channel wave in the VTI media can be obtained.

Dispersion analysis of the Love channel wave

The group velocity is easily deduced from the phase velocity (Dresen and Rüter 1994). If we take the parameters in Table 1 as an example to calculate the zero- to second-order theoretical dispersion curves of the Love channel wave in both the VTI media and isotropic media (Fig. 3), the largest difference is in the velocity. The minimum velocity of the isotropic media is 1100 m/s in the coal seam, whereas the minimum velocity of the VTI media is 1250 m/s, which is approximately 14% higher than that via the isotropic media. The dominant frequencies of the fundamental mode of the Airy phase in the two media are almost identical (150 Hz), but there is a large

Table 1 Elastic parameters for the symmetrical three-layer VTI model

Parameters	V_s (m/s)	ρ (kg/m^3)	γ
Coal seam	1100	1300	0.15
Surrounding rock	2000	2400	0

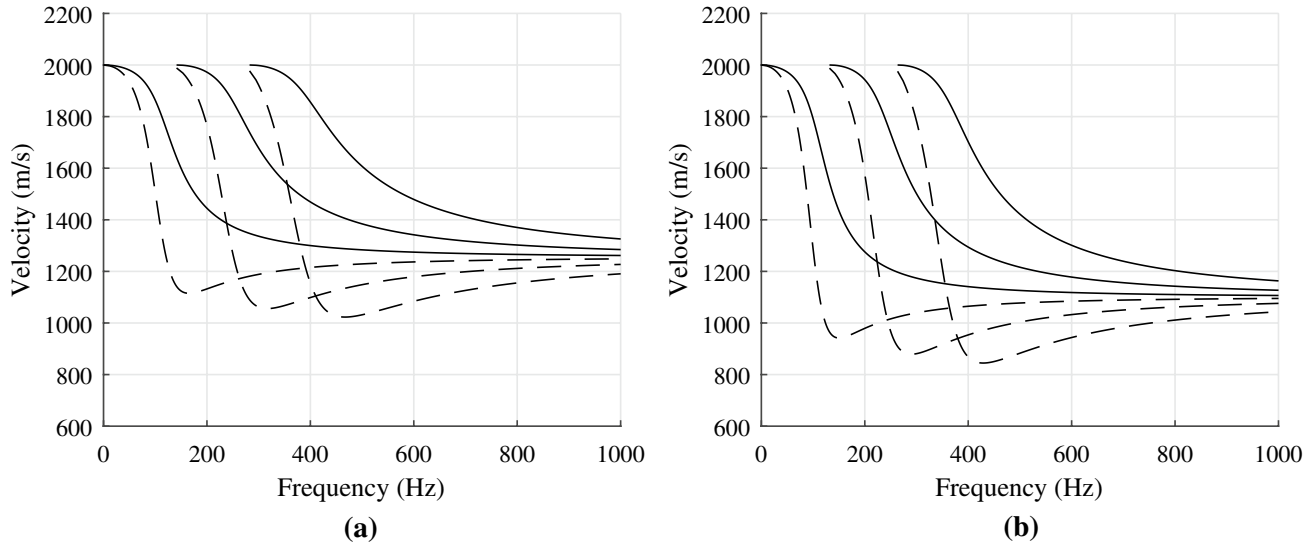


Fig. 3 Theoretical dispersion curves of the zero- to second-order Love channel wave in both the VTI media and isotropic media. **a** VTI medium dispersion curves. **b** Isotropic medium dispersion curves. The solid line indicates the phase velocity. The dashed line indicates the group velocity

difference between them in terms of the high-order Airy phase. The frequency of the Airy phase in the second-order VTI is 440–500 Hz, while that of the isotropic media is 400–450 Hz.

When detecting coal thickness in the working face, we invert for the velocity distribution of a certain frequency of Love channel wave, and then find out the corresponding coal thickness according to Love channel wave dispersion curve of various coal thickness (Wang et al. 2012; Hu et al. 2018). The dispersion curve of isotropic medium is used at present, but the dispersion of VTI medium and isotropic medium is quite different, so there may be a big error in the current method of detecting coal thickness, and the dispersion curve of VTI medium should be adopted.

The dispersion curves of Love channel waves in the isotropic media are dramatically influenced by both the coal bed thickness and its S-wave velocity. The dispersion equation in the VTI media (Eq. (14)) is similar to the corresponding equation in the isotropic media. Therefore, the thickness of the coal seam, S-wave velocity of the coal seam, and γ of the coal seam significantly influence the dispersion curve, as these parameters also influence the isotropic media. Because γ in the surrounding rock is relatively small, its effect is negligible. If we take the parameters in Table 1 as a basis, stabilize the other parameters and alter the thicknesses of the coal seams to be 2 m, 3 m, 5 m and 8 m (Table 2), then four groups of dispersion curves can be calculated (Fig. 4a). Because the Airy phase of the fundamental mode group velocity is most important, we focus on analyzing it.

The variable thickness of the coal seam seems to have a significant influence on the frequency range of the Airy phase, while the Airy phase velocity remains constant

Table 2 Single parameter changes of Love channel waves in the VTI model

Parameters	1	2	3	4
Coal thickness	2 m	3 m	5 m	8 m
Coal v_s	800 m/s	950 m/s	1100 m/s	1250 m/s
Coal γ	0.05	0.1	0.15	0.2

(Fig. 4a). Specifically, the thicker the coal seam is, the lower the frequency of the Airy phase. The frequency of the Airy phase is approximately 100 Hz in the 8-m coal seam model. However, it is 400 Hz, with a broad frequency band, in the 2-m coal seam model. The frequency of the Airy phase varies nonlinearly with the coal seam thickness. In addition, the S-wave velocity of the coal seam also significantly influences the velocity of the Airy phase. Specifically, the velocity of the Airy phase varies from 700 to 1300 m/s (Fig. 4b). The S-wave velocity has a minor effect on the frequency of the Airy phase; the frequency only varies from 100 to 200 Hz. The γ of the coal seam primarily influences the velocity of the Airy phase, which varies from 1000 to 1170 m/s, and it has a slight effect when compared with the S-wave velocity of the coal seam (Fig. 4c). Specifically, it has little effect on the frequency of the Airy phase.

Polarization characteristics and amplitude distribution analysis of the channel waves in VTI media

In the VTI media, the polarizations of the qP- and qSV-waves are not parallel or perpendicular to the orientation of

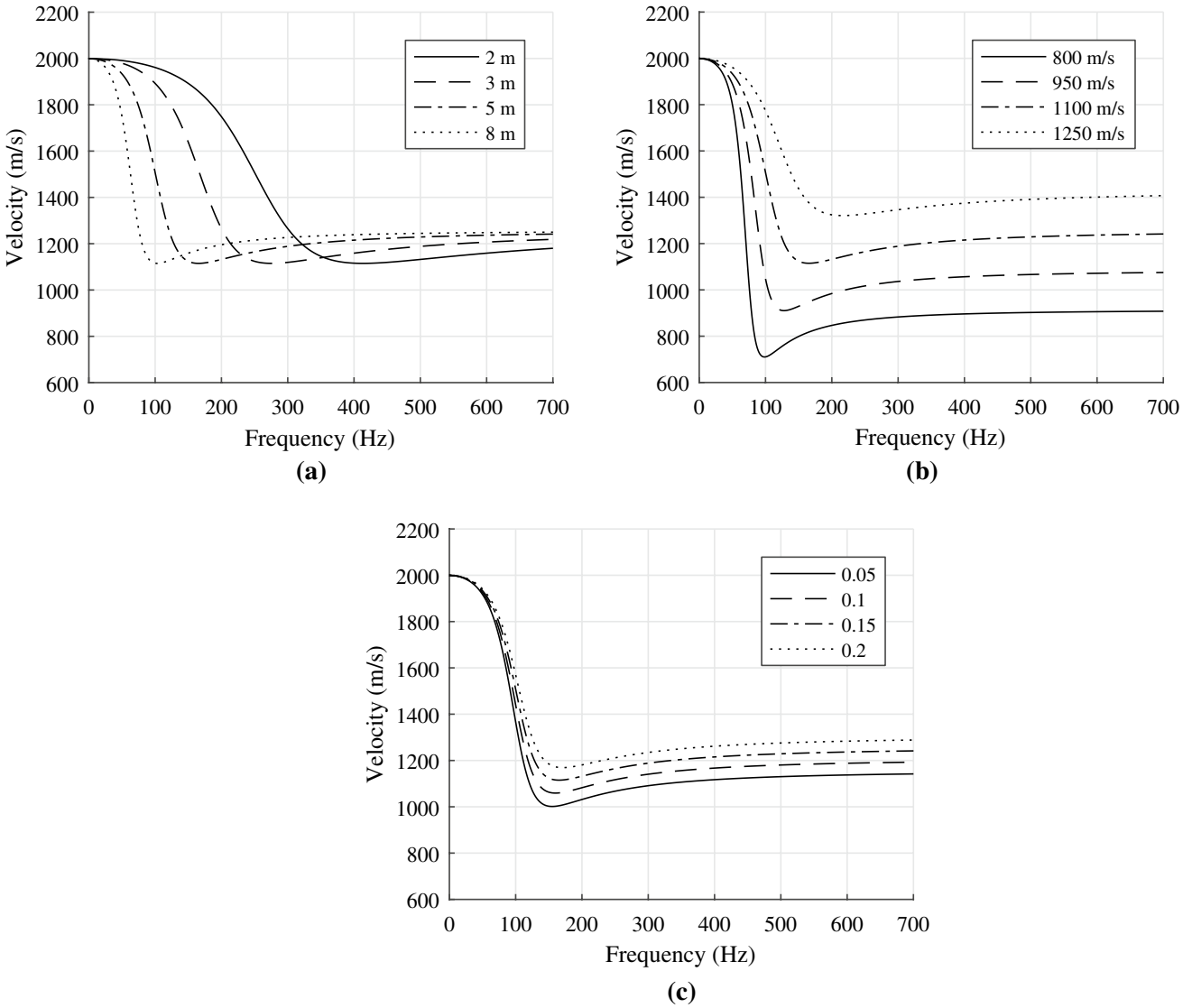


Fig. 4 Comparison of the fundamental mode dispersion curves for group velocity with different parameters. **a** Change in coal thickness. **b** Change in coal v_s. **c** Change in coal γ

wave propagation, while those of the Love channel wave in the horizontal plane is still norm to the orientation of wave propagation. The polarization of Rayleigh channel wave in the VTI media is very different from that in the isotropic media; therefore, it is difficult to extract the Rayleigh channel wave. However, since the polarization of the Love channel wave in the VTI media is similar to that in the isotropic media along the horizontal plane, we can extract the Love channel wave from a complicated wave field. Moreover, the SH wave shows no anisotropic velocity along the horizontal plane, which allows us to efficiently process the channel wave. Consequently, the Love channel wave, instead of the Rayleigh channel wave, is considered when conducting field exploration.

Taking Table 1 as the medium parameter, when the phase velocity is 1700 m/s, the fundamental frequency corresponding to the phase velocity based on the dispersion curve is 112 Hz, the first-order frequency is 293.34 Hz, and the second-order frequency is 456.29 Hz. According to the solution method of Love channel wave amplitude in the previous section, we calculate the amplitude-depth distribution of Love channel wave under this condition (Fig. 5).

In Fig. 5, the amplitude distribution of Love channel waves in the VTI media is similar to that in the isotropic media. That is, the amplitude of the fundamental mode and second-order channel waves is maximum in the middle of coal seam and has an even symmetry. However, the amplitude of the first-order channel wave in the middle of the

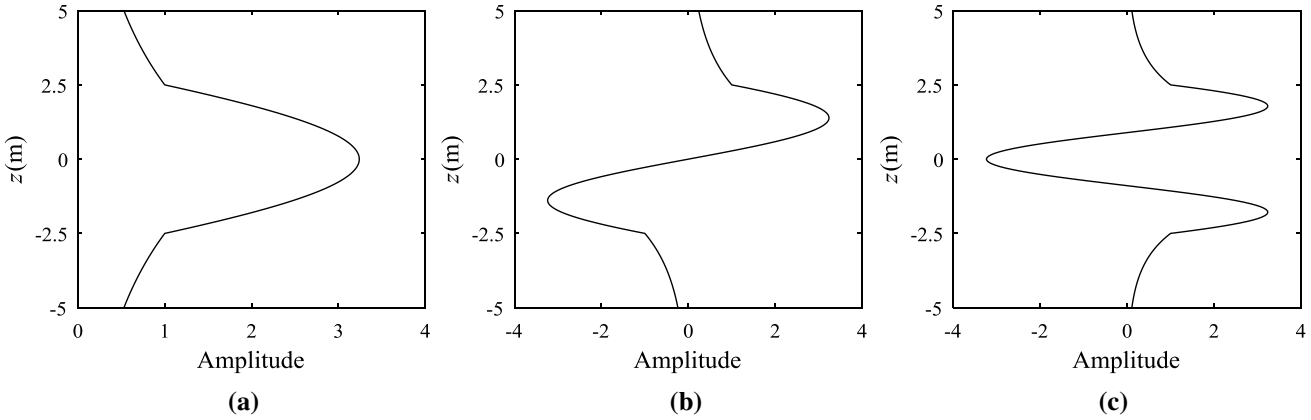


Fig. 5 Depth distribution of the Love channel wave amplitude in VTI medium. **a** Amplitude distribution of fundamental mode. **b** Amplitude distribution of first-order. **c** Amplitude distribution of second-order

coal seam equals zero, while shows the maximum values at a distance equal to one-fourth of the coal seam thickness, which indicates odd symmetry.

Numerical simulation of channel waves in the 3D VTI model

First-order velocity–stress elastic equation in the 3D VTI model

To eliminate computing the second derivative for displacement, the variables V_x, V_y, and V_z in the first derivative for displacement are introduced. Without considering the external force, the first-order velocity–stress elastic equations can be written as:

$$\begin{aligned}
\frac{\partial \sigma_{xx}}{\partial t} &= C_{11}\frac{\partial V_x}{\partial x} + (C_{11} - 2C_{66})\frac{\partial V_y}{\partial y} + C_{13}\frac{\partial V_z}{\partial z} \\
\frac{\partial \sigma_{yy}}{\partial t} &= C_{11}\frac{\partial V_y}{\partial y} + (C_{11} - 2C_{66})\frac{\partial V_x}{\partial x} + C_{13}\frac{\partial V_z}{\partial z} \\
\frac{\partial \sigma_{zz}}{\partial t} &= C_{33}\frac{\partial V_z}{\partial z} + C_{13}\frac{\partial V_x}{\partial x} + C_{13}\frac{\partial V_y}{\partial y} \\
\frac{\partial \tau_{yz}}{\partial t} &= C_{44}\left(\frac{\partial V_y}{\partial z} + \frac{\partial V_z}{\partial y}\right) \\
\frac{\partial \tau_{xz}}{\partial t} &= C_{44}\left(\frac{\partial V_x}{\partial z} + \frac{\partial V_z}{\partial x}\right) \\
\frac{\partial \tau_{xy}}{\partial t} &= C_{66}\left(\frac{\partial V_x}{\partial y} + \frac{\partial V_y}{\partial x}\right)
\end{aligned} \tag{15}$$

We perform numerical simulations of the channel wave in the 3D geological model by using the staggered-grid high-order finite difference algorithm. Mirror image method is used to treat free surface on the wall of the roadway (Ji et al. 2012; Li et al. 2015).

Numerical modeling

Figure 6 displays the 3D model with the dimensions of 200 × 200 × 25 m ($X \times Y \times Z$). Specifically, the middle layer of the model is the coal seam with the thickness of 5 m. The top and bottom media indicate surrounding rocks, which have similar physical properties. The grid unit of the 3D model is 1 × 1 × 0.25 m ($X \times Y \times Z$), and the temporal sampling interval is 0.05 ms. For the coal seam, the vertical P- and S-wave velocities are 1900 m/s and 1100 m/s, respectively. Its elastic parameters ε, γ and δ are equal to 0.1, 0.15 and − 0.1, respectively. Bulk density of the coal is 1300 kg/m^3. For the surrounding rocks, the vertical P- and S-wave velocities are 3500 m/s and 2000 m/s, respectively. Their elastic parameters ε, γ and δ are equal to 0, indicating an isotropic case. Bulk density of the rock is 2400 kg/m^3. There are two vacuum roadways: one is at x = 11–15 m, y = 10–190 m and z = 11–14 m; another is at x = 186–190 m, y = 10–190 m and z = 11–14 m. The roadway section is 4 m × 3 m. Survey line 1 (x = 185 m, z = 12.5 m) is located along the wall of the roadway to the right, and survey line 2 (y = 100 m, z = 12.5 m) passes through the source point (black lines in Fig. 6b). The seismic source, which is the dominant frequency of the 150-Hz Ricker wavelet, is located at x = 16 m and y = 100 m (circled in Fig. 6b).

Figure 7 displays the 60 ms snapshot. The fastest wave front is a refracted P-wave followed by a high-order Rayleigh channel wave. The fundamental channel wave propagates the slowest, with a high energy concentration. Meanwhile, the roadway channel wave is also observed, and it is generally slower than the traditional channel wave.

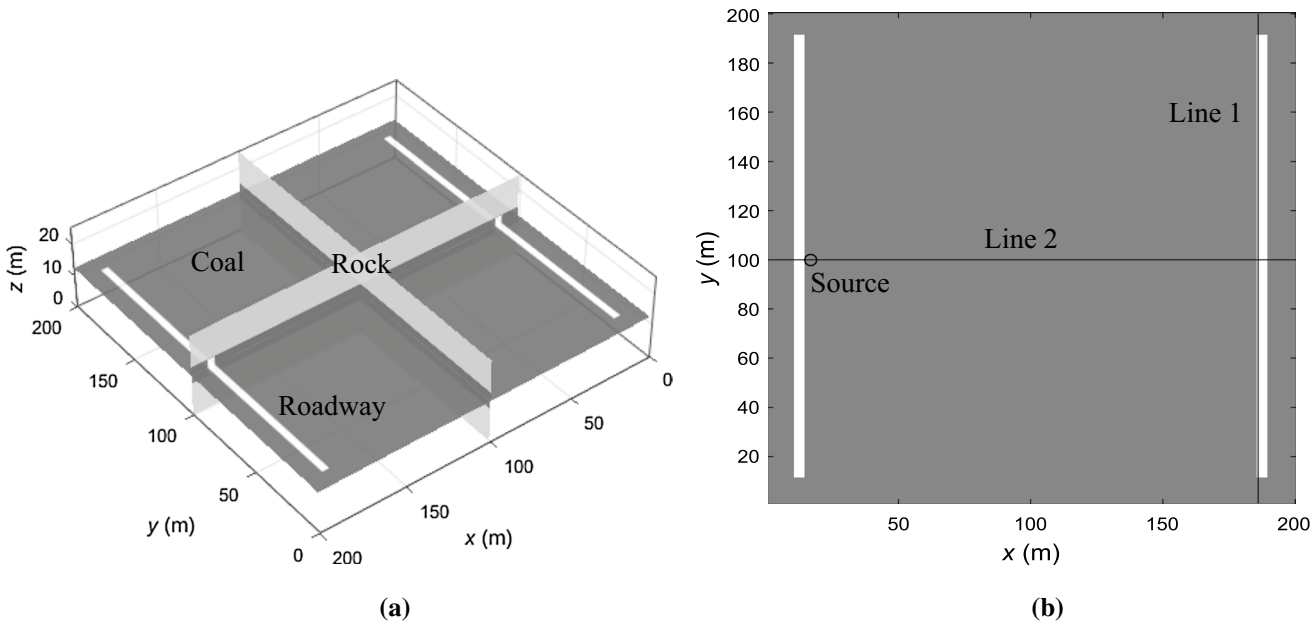


Fig. 6 Working face model of a coal seam with roadways. **a** 3D model slices. Dark gray represents the surrounding rock, light gray represents the coal seam and white represents the roadway. **b** The horizontal slice in the center of the coal seam ($z = 12.5$ m). The black circle represents the source, and the black line represents the survey lines

Additionally, the z-component of the channel wave is also slower than both its x- and y-components.

Figure 8 shows the transmitted channel wave record of survey line 1. The figure shows that the fundamental mode of the channel wave is intensive, but it is also the slowest. Therefore, it is a high-order Rayleigh channel wave. The velocities of the x- and y-components are faster than that of the z-component, but their energies are more concentrated; the group velocity is approximately 1100 m/s and includes both the Love and Rayleigh channel wave. The fundamental mode of the z-component, which mainly consists of a Rayleigh channel wave, is not very concentrated. The energies of both the x- and y-components are primarily concentrated along the side of the survey line, which is due to the roadway effect.

We analyze the wave record and frequency dispersions for survey line 2 (Fig. 9). Because line 2 passes through the source, it is obvious that it only generates Love channel waves in the y-component. The dispersion power spectra in the V-f domain are extracted by using the 2D fast Fourier transform (FFT). The dotted line (Fig. 9b) represents the theoretical dispersion curve calculated by the dispersion formula (Eq. 14). The theoretical curve and extracted dispersion image from the simulation is consistent which shows that the theoretical dispersion formula derived is correct. Figure 9b shows that the fundamental mode of the Love channel wave is mainly observed. The energy is concentrated within 150–250 Hz, and the velocity ranges from 1250 to 2000 m/s, with a minimum velocity of 1250 m/s. If the model is isotropic, then the minimum velocity of the coal seam is 1100 m/s. Obviously, the two models are quite different.

If we use the isotropic model rather than the VTI model for the coal seam, the wave fields are similar to those in survey line 1 (Fig. 10); however, there is a large velocity difference. Compared with that in the isotropic media, the velocity is higher in the VTI media. The group velocity of the fundamental mode in both the x- and y-components is 950 m/s, whereas that in the VTI media is 1100 m/s (Fig. 3). In addition, the high-order Rayleigh channel wave in the isotropic media is more intense than that in the VTI media. The dispersion curves of survey line 2 can distinctly show the difference in velocity between the isotropic media and VTI media (Fig. 11). The minimum phase velocity of the VTI media is 1250 m/s, while that of the isotropic media is 1100 m/s. This obvious velocity difference significantly influences the processed data for the channel wave. From the perspective of distributed energy in the frequency dispersion (Fig. 11b), we can see that these two cases are almost similar, which suggests that the VTI media may have a minor effect on the energy distribution of a channel wave.

Discussion

The anisotropy of coal seam is complicated. When there are near vertical cracks in coal seams (Chen et al. 2010; Ji et al. 2019), Horizontal Transverse Isotropy (HTI) medium

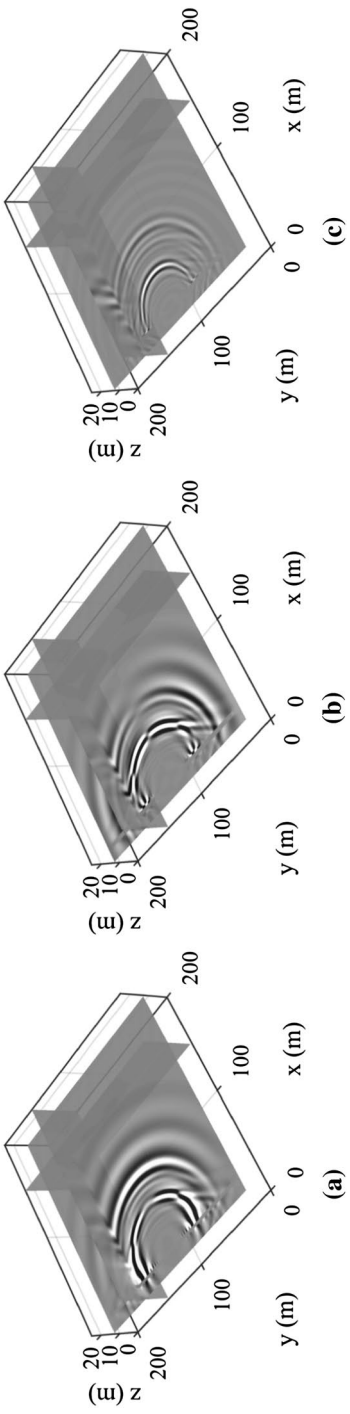


Fig. 7 Wave field snapshot at 60 ms. **a** *X*-component. **b** *Y*-component. **c** *Z*-component

can be used to be equivalent. When coal seams have both bedding and cracks, orthotropic medium can be used to be equivalent. Therefore, it is difficult to summarize the coal seam properties with one theoretical model.

At present, little is known about the properties of channel waves in anisotropic coal seams. This paper begins with the study of weak anisotropic and VTI model coal seams, so as to avoid the analysis difficulty caused by complex models which should be further studied. In addition, other properties in anisotropic media such as multilayer Love channel wave dispersion, Rayleigh channel wave dispersion and channel wave field in coal seam with structures also need to be studied in the future.

In this paper, the channel waves in VTI medium are only studied in theory, and it is necessary to compare with the real data in the future.

Conclusions

The propagation patterns of the channel waves and the dispersion features of Love channel waves are investigated in VTI media which is weak anisotropy based on Thomsen theory. The conclusions are as follows.

In VTI media, the polarization of qP- and qSV-waves is not parallel or norm to the orientation of wave propagation. However, the horizontal plane in the VTI media is still isotropic, and the polarization direction of the SH wave is still perpendicular to the direction of wave propagation, so Love channel waves can be accurately separated out. Therefore, it is wise to use Love channel waves to conduct field exploration.

The dispersion curves of Love channel waves in VTI media are quite different from those in isotropic media. The velocity of the Love channel wave in VTI media is higher than that in isotropic media. However, the Airy phase frequency of the fundamental mode seems to be similar in both models. The thickness of the coal seam plays an important role in influencing the frequency of the Airy phase. Both the S-wave velocity and γ parameter of the coal seam significantly affect the velocity of the Airy phase. In addition, the γ of the coal seam has little effect on the frequency of the Airy phase. In terms of the amplitude distribution, Love channel waves appear to have similar patterns in both the VTI media and the isotropic media.

According to our findings, big errors may occur in the inversion of the coal seam thickness when using the Love channel wave dispersion curve in isotropic media. Therefore, we propose using the more accurate dispersion curve of VTI media instead of that of the isotropic media.

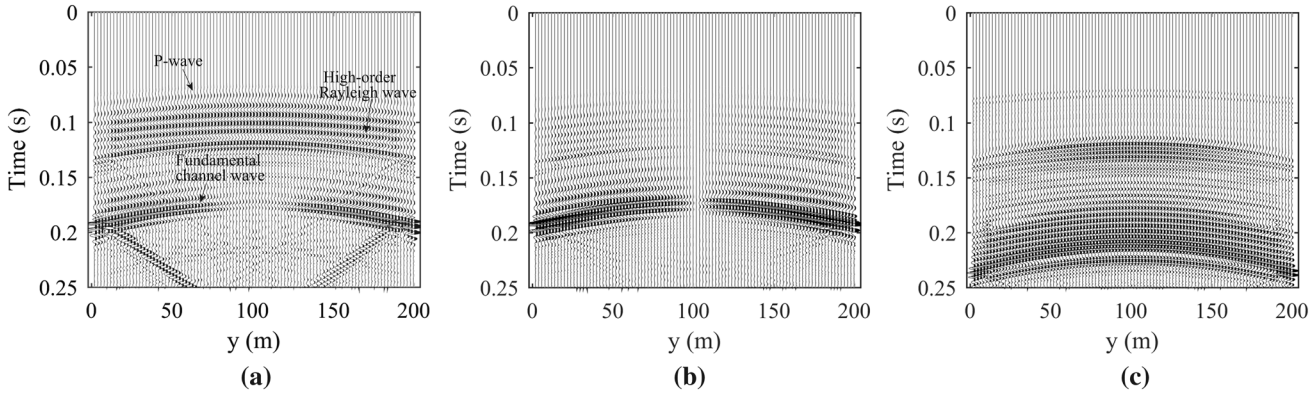


Fig. 8 Channel wave record by survey line 1 in VTI medium. **a** *X*-component. The refracted P-wave is the fastest, the high-order Rayleigh channel wave is the second-fastest, and the fundamental mode channel wave is the slowest. **b** *Y*-component. **c** *Z*-component

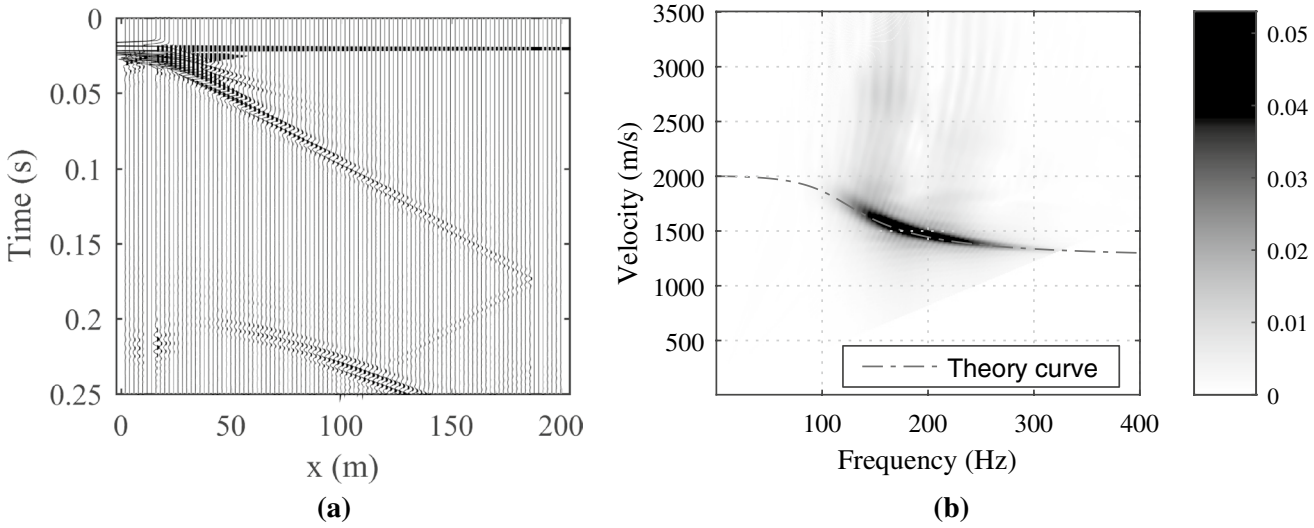


Fig. 9 *Y*-component channel wave record by survey line 2 in VTI medium and its power spectrum in the V–f domain. **a** *Y*-component channel wave record. **b** Power spectrum in the V–f domain. The dotted line represents the theoretical dispersion curve

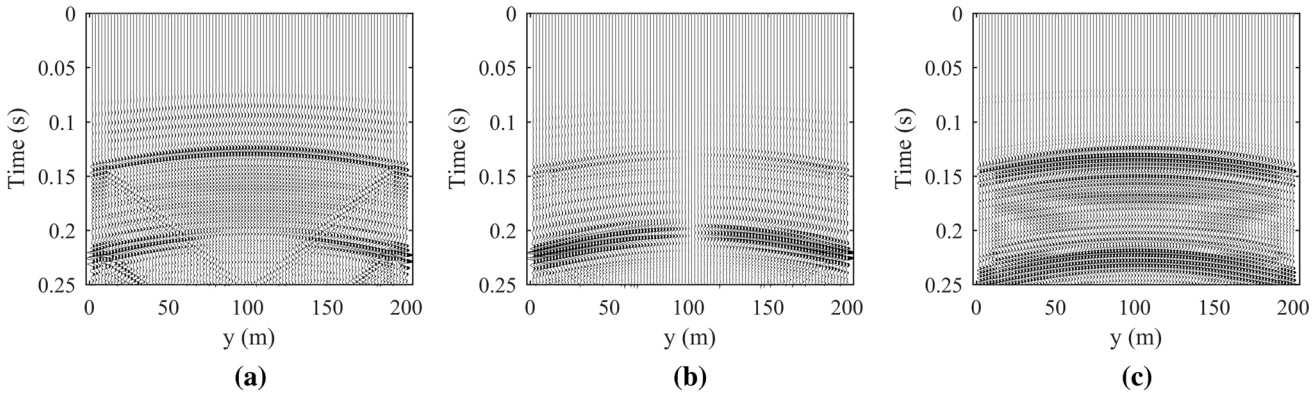


Fig. 10 Channel wave record by survey line 1 in isotropy medium. **a** *X*-component. **b** *Y*-component. **c** *Z*-component

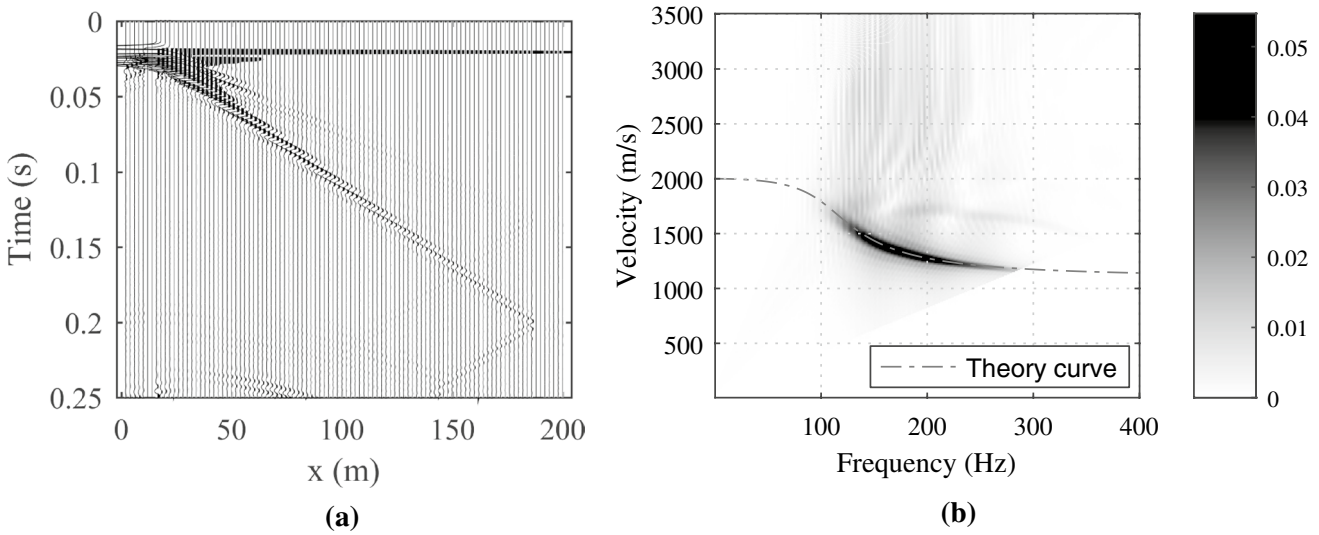


Fig. 11 *Y*-component channel wave record by survey line 2 in isotropy medium and its power spectrum in the V–f domain. **a** *Y*-component channel wave record. **b** Power spectrum in the V–f domain. The dotted line represents the theoretical dispersion curve

Acknowledgements This research is financially supported by National Key R&D Program of China (2017YFC0804105), China Postdoctoral Science Foundation (2019M652434) and Qingdao Postdoctoral Applied Research Project.

References

Buchanan DJ, Jackson P, Davis D (1983) Attenuation and anisotropy of channel waves in coal seams. Geophysics 48:133–147. https://doi.org/10.1190/1.1441453

Chen TJ, Wang X, Cui RF (2010) The detectability analysis on HTI tectonic coal cracks by azimuthal AVO's forward modeling. J China Coal Soc 35(4):640–644. https://doi.org/10.13225/j.cnki.jccs.2015.04.015 **(in Chinese)**

Dong SH (2008) Test on elastic anisotropic coefficients of gas coal. Chin J Geophys 51(3):947–952. https://doi.org/10.1002/cjg2.1257 **(in Chinese)**

Dresen L, Rüter H (1994) Seismic coal exploration, part B: in-seam seismics. Pergamon, Oxford

Essen K, Bohlen T, Friederich W, Meier T (2007) Modelling of Rayleigh-type seam waves in disturbed coal seams and around a coal mine roadway. Geophys J Int 170:511–526. https://doi.org/10.1111/j.1365-246X.2007.03436.x

Feng L, Zhang Y (2017) Dispersion calculation method based on S-transform and coordinate rotation for Love channel waves with two components. Acta Geophys 65(4):757–764. https://doi.org/10.1007/s11600-017-0069-y

He WX (2017) Three dimensions in-seam wave field simulation and detection method research of working face fault. China University of Mining and Technology, Xuzhou

He WX, Ji GZ, Dong SH et al (2017) Theoretical basis and application of vertical Z-component in-seam wave exploration. J Appl Geophys 138(3):91–101. https://doi.org/10.1016/j.jappgeo.2017.01.008

Hu ZA, Zhang PS, Xu GQ (2018) Dispersion features of transmitted channel waves and inversion of coal seam thickness. Acta Geophys 66(5):1001–1009. https://doi.org/10.1007/s11600-018-0192-4

Ji GZ, Cheng JY, Zhu PM et al (2012) 3-D numerical simulation and dispersion analysis of in-seam wave in underground coal mine. Chin J Geophys 55(2):645–654. https://doi.org/10.6038/j.issn.0001-5733.2012.02.028 **(in Chinese)**

Ji GZ, Wei JC, Yang ST et al (2018) Three-component polarization migration of channel waves for prediction ahead of coal roadway. J Appl Geophys 159(12):475–483. https://doi.org/10.1016/j.jappgeo.2018.09.028

Ji GZ, Wei JC, Yang ST et al (2019) Preliminary study on wave field and dispersion characteristics of channel waves in HTI coal seam medium. Chin J Geophys 62(2):789–801. https://doi.org/10.6038/cjg2019M0230 **(in Chinese)**

Korn M, Emmerich H (1988) Waves in discontinuous coal seams with absorption: finite difference simulations. Explor Geophys 19(2):106–108. https://doi.org/10.1071/EG988106

Krey TC (1963) Channel waves as a tool of applied geophysics in coal mining. Geophysics 28:701–714. https://doi.org/10.1190/1.1439258

Krey T, Arnetzb H, Knecht M (1982) Theoretical and practical aspects of absorption the application of in-seam seismic coal exploration. Geophysics 47(12):1645–1656. https://doi.org/10.1190/1.1441314

Li H, Zhu PM, Ji GZ, Zhang Q (2015) Modified image algorithm to simulate seismic channel waves in 3D tunnel mode. Geophys Prospect 64(5):1259–1274. https://doi.org/10.1111/1365-2478.12351

Liu E, Crampins S, Roth B (1991) Modelling channel waves with synthetic seismograms in an anisotropic in-seam seismic survey. Geophys Prospect 40(5):13–540. https://doi.org/10.1111/j.1365-2478.1992.tb00539.x

Morcote A, Mavko G, Prasad M (2010) Dynamic elastic properties of coal. Geophysics 75(6):E227–E234. https://doi.org/10.1190/1.3508874

Rader D, Schott W, Dresen L, Ruter H (1985) Calculation of dispersion curves and amplitude-depth distributions of Love channel waves in horizontally-layered media. Geophys Prospect 33:800–816. https://doi.org/10.1111/j.1365-2478.1985.tb00779.x

Thomsen L (1986) Weak elastic anisotropy. Geophysics 51(10):1954–1966. https://doi.org/10.1190/1.1442051

Wang W, Gao X, Li SY et al (2012) Channel wave tomography method and its application in coal mine exploration: an example from Henan Yima Mining Area. Chin J Geophys 55(3):1054–1062. https://doi.org/10.6038/j.issn.0001-5733.2012.03.036 **(in Chinese)**

Wang B, Liu S, Zhou F et al (2017) Diffraction characteristics of small fault ahead of tunnel face in coal roadways. Earth Sci Res J 21(2):95–99. https://doi.org/10.15446/esrj.v21n2.64938

Yang ST, Cheng JL (2012) The method of small structure prediction ahead with Rayleigh channel wave in coal roadway and seismic wave field numerical simulation. Chin J Geophys 55(2):655–662. https://doi.org/10.6038/j.issn.0001-5733.2012.02.029 **(in Chinese)**

Yang XH, Cao SY, Li DC et al (2014) Analysis of quality factors for Rayleigh channel waves. Appl Geophys 11(1):107–114. https://doi.org/10.1007/s11770-014-0409-5

Yang ST, Wei JC, Cheng JL et al (2016) Numerical simulations of full-wave fields and analysis of channel wave characteristics in 3-D coal mine roadway models. Appl Geophys 13(04):621–630. https://doi.org/10.1007/s11770-016-0582-9

Acta Geophysica (2019) 67:1391–1402
https://doi.org/10.1007/s11600-019-00332-z

RESEARCH ARTICLE - APPLIED GEOPHYSICS

Optimized compact finite difference scheme for frequency-domain acoustic wave equation

Aman Li[1,2,3] · **Hong Liu**[1,2,3]

Received: 17 April 2019 / Accepted: 14 July 2019 / Published online: 18 July 2019

Abstract

Frequency-domain numerical simulation is the most important foundation of frequency-domain full-waveform inversion and reverse time migration. The accuracy of numerical simulation seriously affects the results of the seismic inversion and image. In this article, we develop an optimized compact finite difference scheme for acoustic wave equation in frequency domain to improve numerical simulation accuracy. For the sake of avoiding the extra memory and computational costs caused by solving the inverse of a pentadiagonal band matrix, we calculate the optimized compact finite difference discrete operator for the Laplace operator in the numerical simulation. Although the optimized compact finite difference scheme has only second-order formal accuracy, it has a spectral-like resolution feature. This method can significantly reduce the numerical dispersion and the numerical anisotropy. We find that the results of the optimized compact finite difference scheme agree well with the analytic solution according to accuracy analysis. Two numerical simulations are done to verify the theoretical analysis of the optimized compact finite difference scheme.

Keywords Optimization · Compact finite difference · Frequency domain · Modeling

Introduction

In recent years, due to the improvement of computing ability, reverse time migration (RTM) (McMechan 1983; Chang and McMechan 1987; Baysal et al. 1983; Whitmore 1983; Zhang et al. 2010) and full-waveform inversion (FWI) (Tarantola 1984; Mora 1987; Pratt et al. 1998; Virieux and Operto 2009; Shin and Cha 2008; Yuan et al. 2019) have developed rapidly and played a significant role in the field of exploration geophysics. As the basis of the seismic image and inversion, seismic wave numerical simulation can seriously affect the quality of the results. Seismic wave numerical simulation can be done in frequency domain and time domain. Time-domain forward modeling has been greatly developed because of its wide applicability. However, for some problems, frequency-domain modeling is more convenient and efficient than time-domain forward modeling. For example, there are no stability problems for frequency-domain modeling (Marfurt 1984) due to the implicit relationship. And the attenuation effects of the wave propagation in viscous media are easier to implement in frequency domain (Pratt 1990). Moreover, for multi-shot parallel computation, the numerical simulation in frequency domain by the way of direct solver is more efficient than that in time domain.

Lysmer and Drake (1972) firstly did numerical simulations in frequency domain for seismology. They calculated the propagation of seismic wave in the subsurface medium by the finite element method. Then, Marfurt (1984) and Marfurt and Shin (1989) examined the effect of this method and further developed it into wave equation for exploration geophysics. At the same time, the finite difference method became more and more popular for solving geophysical problems due to its ease of implementation, characteristic of parallelism and high efficiency (Alford et al. 1974; Virieux 1984; Alterman and Karal 1968; Virieux 1986; Shi et al. 2018). The frequency-domain forward modeling with finite difference method was firstly applied to the crosshole seismic imaging and inversion by Pratt (1990) and Pratt and Worthington (1990). Compared

✉ Aman Li
liaman16@mails.ucas.ac.cn

[1] Key Laboratory of Petroleum Resources Research, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China

[2] Institutions of Earth Science, Chinese Academy of Science, Beijing 100029, China

[3] University of Chinese Academy of Sciences, Beijing 100049, China

with the finite element method forward modeling in frequency domain, finite difference method can reduce computational costs. The spatial derivative is approximated by the second-order central finite difference operator, so it requires too many sampling grid points per wavelength to achieve higher accuracy. To deal with the problem, Jo et al. (1996) presented an optimized 9-point scheme for the acoustic wave equation by using the weighted average mass acceleration term and the technique of rotated coordinate. They used the steepest descent method to optimize the numerical phase velocity curves to get the optimal coefficients. This method could reduce the value of required sampling points for per wavelength to 4 and successfully reduce computational costs. Based on the same method, by combining the finite difference operators of four rotated angles, Shin and Sohn (1998) presented an optimized 25-point scheme for the acoustic wave numerical simulation. This new scheme reduced the number of required sampling points for per wavelength to 2.5. Hustedt et al. (2004) generalized this method to the numerical simulation of the variable density acoustic equation, and Operto et al. (2007) extended it to 3D case. It can be argued that those optimal schemes based on the coordinate rotation technique have achieved considerable success. However, the method has a drawback that it demands the equal space sampling intervals, which seriously influences its practical applications. To overcome the limitation of the rotated coordinate method, Chen (2012) proposed an optimized 9-point scheme for scalar wave equation by the way of average-derivate method (ADM). In this method, the finite difference of spatial derivative is approximated by the average derivative of three grid points in the orthogonal direction, which can be applied to unequal space sampling intervals, increasing the flexibility and widening its application range. Compared with the rotated coordinate method, average-derivate method has higher accuracy, so it has been widely used. Tang et al. (2015) presented an optimized ADM 17-point scheme to further improve the accuracy. Chen (2014) and Chen and Cao (2016) generalized this method to 3D case and elastic wave numerical simulation in frequency domain, respectively. All of the above methods are using explicit finite difference operators to discretize spatial partial derivatives. In fact, compact (implicit) finite difference operators can also be used to discrete spatial partial derivatives. Lele (1992) introduced a series of schemes for calculating spatial partial derivatives based on the pentadiagonal compact relationship. Those schemes have a spectral-like resolution, although they are only in the form of fourth-order accuracy. Subsequently, Kim and Lee (1996) optimized the high-order compact finite difference scheme for first derivatives to achieve maximum-resolution characteristics. Liu and Sen (2009) developed compact finite difference method for seismic numerical simulation in time domain. Chu and Stoffa (2012) used compact finite difference operator for numerical simulation in frequency domain, and they improved the accuracy and got better simulation results.

In this article, we propose an optimized compact finite difference scheme for acoustic wave modeling in frequency domain. The optimized compact finite difference scheme has a spectral-like resolution, although it just has the form of second-order accuracy. In order to avoid the extra memory and computational costs caused by solving the inverse of a pentadiagonal band matrix, we calculate the optimized compact finite difference operators for the Laplace operator. In the following part, we firstly introduce the optimal compact finite difference scheme for acoustic wave. Next, we obtain the optimal coefficients by optimizing the numerical phase velocity curves and evaluate the errors of the different schemes by the dispersion analysis. Then, we analyze the precision of numerical solutions of different schemes by comparing with the analytical solution. Finally, we use two forward modeling examples to further verify the effectiveness of our method.

Theory

The optimized compact scheme

For a homogeneous medium, the frequency-domain acoustic wave equation in two-dimensional Cartesian coordinates is:

$$\nabla^2 P + \frac{\omega^2}{v^2} P = \frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial z^2} + \frac{\omega^2}{v^2} P = s, \tag{1}$$

where v is the velocity of the wave, P is the pressure wavefield, ω is the circular frequency and s is the source in frequency domain. By referring to the idea of Lele (1992), we introduce a compact finite difference that only needs five points to approximate the second-order space derivative $\partial^2 P / \partial x^2$

$$\begin{aligned} &\beta P''_{i-2} + \alpha P''_{i-1} + P''_i + \alpha P''_{i+1} + \beta P''_{i+2} \\ &\quad = b\frac{P_{i+2} - 2P_i + P_{i-2}}{4h^2} + a\frac{P_{i+1} - 2P_i + P_{i-1}}{h^2}, \end{aligned} \tag{2}$$

where P_i is the pressure wavefield at the grid x_i, h is the space sampling interval, P''_i is the second-order space derivative of the pressure wavefield at the grid x_i and α, β, a, b are the coefficients. Then, do Taylor expansion on both sides of the compact scheme (2). By comparing the same terms on both sides of the equation, we can obtain the relationships of α, β, a, b. The first mismatched coefficient of the relationships determines the formal truncation error of the compact finite difference approximation (2). These constraint relationships are:

$$b + a = 1 + 2\alpha + 2\beta \quad \text{(second-order)} \tag{3}$$

$$(2)^2 b + (1)^2 a = \frac{4!}{2!}\left(\alpha + \beta * 2^2\right) \quad \text{(fourth-order)} \tag{4}$$

$$(2)^4 b + (1)^4 a = \frac{6!}{4!}\left(\alpha + \beta * 2^4\right) \quad \text{(sixth-order)} \tag{5}$$

$$(2)^6 b + (1)^6 a = \frac{8!}{6!}\left(\alpha + \beta * 2^6\right) \quad \text{(eighth-order)} \tag{6}$$

Because there are four coefficients that we need to determine, only the eighth-order scheme has unique coefficients. The coefficients of other lower-order schemes are not determined because there are not enough constraints, which gives us the opportunity to optimize the coefficients to improve the resolution characteristics.

Fourier analysis is a common method to measure the accuracy of finite difference approximation. This method can provide the quantitative analysis of the optimized compact scheme. The compact Eq. (2) can be rewritten as:

$$\begin{aligned} &\beta P''(x-2h) + \alpha P''(x-h) + P''(x) \\ &\quad + \alpha P''(x+h) + \beta P''(x+2h) \\ &= b\frac{P(x+2h) - 2P(x)_i + P(x-2h)}{4h^2} \\ &\quad + a\frac{P(x+h) - 2P(x)_i + P(x-h)}{h^2}. \end{aligned} \tag{7}$$

By applying Fourier transform on both sides of the above equation, we can get:

$$\begin{aligned} &(i\zeta)^2\left[\beta\exp(-2i\zeta h) + \alpha\exp(-i\zeta h) + 1\right. \\ &\quad \left. + \alpha\exp(i\zeta h) + \beta\exp(2i\zeta h)\right]\tilde{P} \\ &= \frac{b}{4h^2}\left[\exp(2i\zeta h) - 2 + \exp(-2i\zeta h)\right]\tilde{P} \\ &\quad + \frac{a}{h^2}\left[\exp(i\zeta h) - 2 + \exp(-i\zeta h)\right]\tilde{P}. \end{aligned} \tag{8}$$

Simplifying Eq. (8), we obtain the scaled numerical wavenumber:

$$\tilde{k}^2 = \frac{2a(1-\cos(k)) + b/2(1-\cos(2k))}{1+2\alpha\cos(k) + 2\beta\cos(2k)}, \tag{9}$$

where $k^2 = \zeta^2 h^2$ is the true wavenumber. The effect of compact finite difference approximation (2) to the partial derivate depends on the degree of agreement between the true wavenumber k^2 and the scaled numerical wavenumber $\tilde{k}^2$. So we can optimize the coefficients that work best for numerical simulations.

Optimization and dispersion analysis

By minimizing the errors between the true wavenumber k^2 and the scaled numerical wavenumber $\tilde{k}^2$ over a wavenumber range band, we can obtain the following integral error:

$$E(\alpha, \beta, a, b) = \int_{k_l}^{k_u} \left(\tilde{k}^2 - k^2\right)^2 \eta(k)\mathrm{d}k, \tag{10}$$

where $\eta(k)$ is a weighting function, k_u and k_l and are the up and low limits of the integral. Kim and Lee (1996) found that a good weighting function could make Eq. (10) analytically integrable and reduce the high wavenumber influence. Here, we choose the most widely used weighting function as the following form:

$$\eta(k) = [1+2\alpha\cos(k) + 2\beta\cos(2k)]^2. \tag{11}$$

There are many methods to solve the problem of (10). Here, we make use of the conjugate gradient method to determine the optimization coefficients. To obtain the best optimal values of α, β, a, b, in our study, we keep second-order formal accuracy by using Eq. (3) as a constraint condition to the problem of (11). Then, we obtain the coefficients of the optimized pentadiagonal compact finite difference scheme:

$$\begin{aligned} &\alpha = 0.4912, \quad \beta = 0.0390, \\ &a = 0.2241, \quad b = 1.8363. \end{aligned} \tag{12}$$

The optimized compact finite difference (FD) scheme has a spectral-like resolution, although it just has the form of second-order accuracy. Figure 1 shows that the scaled numerical wavenumber varies with the true wavenumber for different schemes. As illustrated in Fig. 1, the optimized second-order compact FD scheme stays close to the accurate differentiation compared with the conventional eighth-order compact FD scheme and the conventional fourth-order explicit FD scheme.

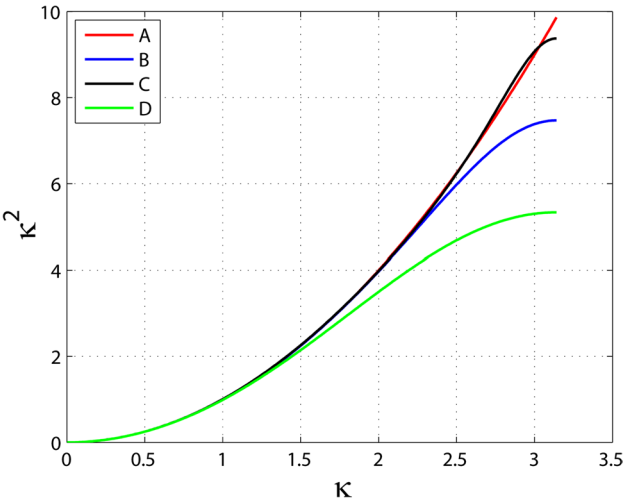


Fig. 1 The differencing error for second derivative: curve A stands for the exact differentiation, curve B stands for the conventional eighth-order compact FD scheme, curve C stands for the optimized second-order compact FD scheme and curve D stands for the conventional fourth-order explicit FD scheme

In order to verify the effect of the optimization compact scheme on the partial differential Eq. (1), we derive its normalized numerical phase velocity:

$$\frac{v_{\mathrm{ph}}}{v} = \frac{G}{2\pi}\sqrt{A+B}$$
$$A = \frac{2a\left(1-\cos\left(\frac{2\pi\sin\theta}{G}\right)\right)+b/2\left(1-\cos\left(\frac{4\pi\sin\theta}{G}\right)\right)}{1+2\alpha\cos\left(\frac{2\pi\sin\theta}{G}\right)+2\beta\cos\left(\frac{4\pi\sin\theta}{G}\right)}$$
$$B = \frac{2a\left(1-\cos\left(\frac{2\pi\cos\theta}{G}\right)\right)+b/2\left(1-\cos\left(\frac{4\pi\cos\theta}{G}\right)\right)}{1+2\alpha\cos\left(\frac{2\pi\cos\theta}{G}\right)+2\beta\cos\left(\frac{4\pi\cos\theta}{G}\right)}, \tag{13}$$

where θ is the wave propagation direction angle and $G = 2\pi/\zeta h$ is the value of sampling points per wavelength. Figure 2 shows the normalized numerical phase velocity curves of the optimized second-order compact FD scheme, the conventional eighth-order compact FD scheme and the conventional fourth-order explicit FD scheme for different propagation angles θ. As shown in Fig. 2, the numerical dispersion errors of the optimized compact FD scheme are smaller than those of the other schemes. To keep in the same normalized numerical phase velocity errors, the optimized compact FD scheme needs fewer G, which means that the optimized compact FD scheme can do numerical simulations on a coarse grid to reduce the computation cost.

The numerical phase velocity errors of FD schemes exhibit in the characteristic of the numerical anisotropy for multi-dimensional problems. Ideally, the normalized phase velocity should be constantly one for all waves (any direction and any wavenumber) in the isotropic medium. However, numerical simulation method leads to the numerical anisotropy. Figure 3 shows the numerical anisotropic propagation of the optimized second-order compact FD scheme, the conventional eighth-order compact FD scheme and the fourth-order explicit FD scheme. The curves in Fig. 3 are polar plots for all propagation angles θ. For each FD scheme, the cases of $G = 1, 2, 3, 4$ are plotted. As illustrated in Fig. 3, we can see that the effect of FD numerical simulation is better for the larger G. The

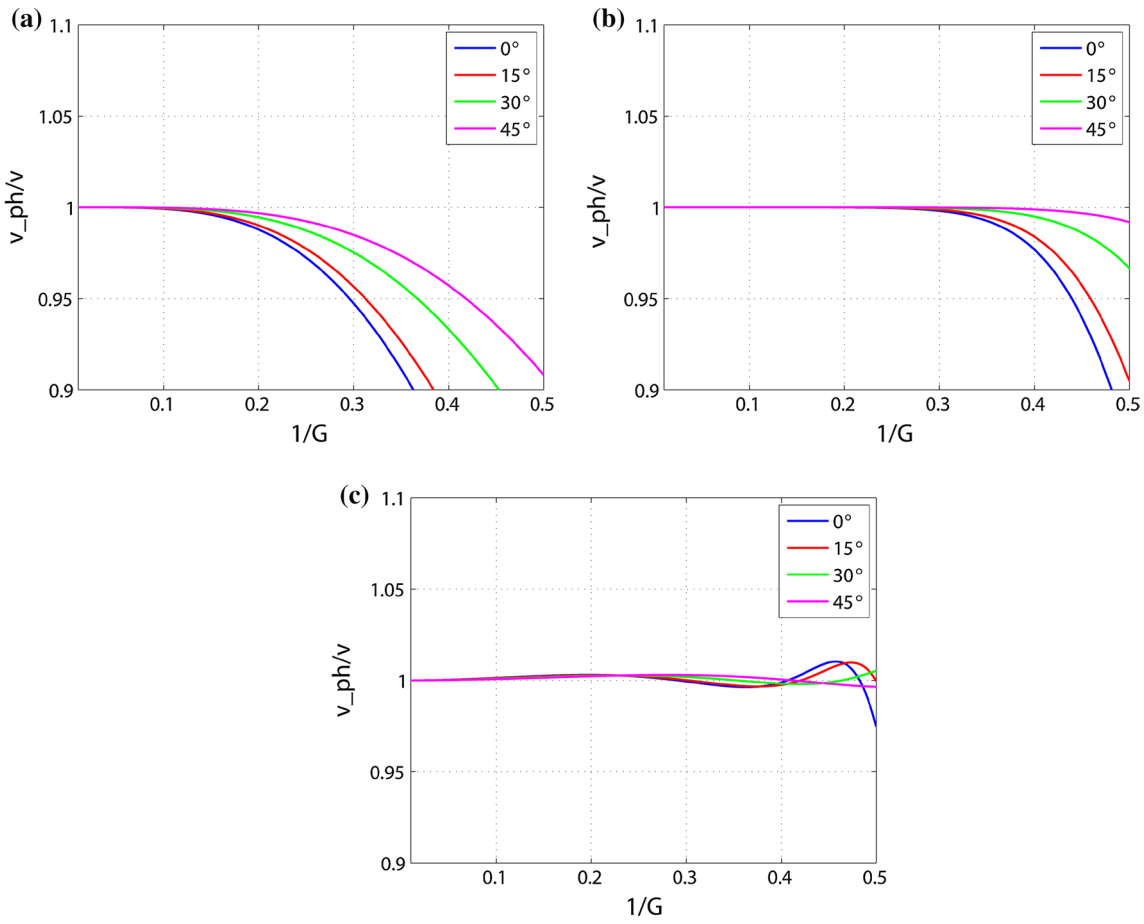


Fig. 2 The normalized numerical phase velocity curves of **a** the conventional fourth-order explicit FD scheme, **b** the conventional eighth-order compact FD scheme and **c** the optimized second-order compact FD scheme for different propagation angles θ

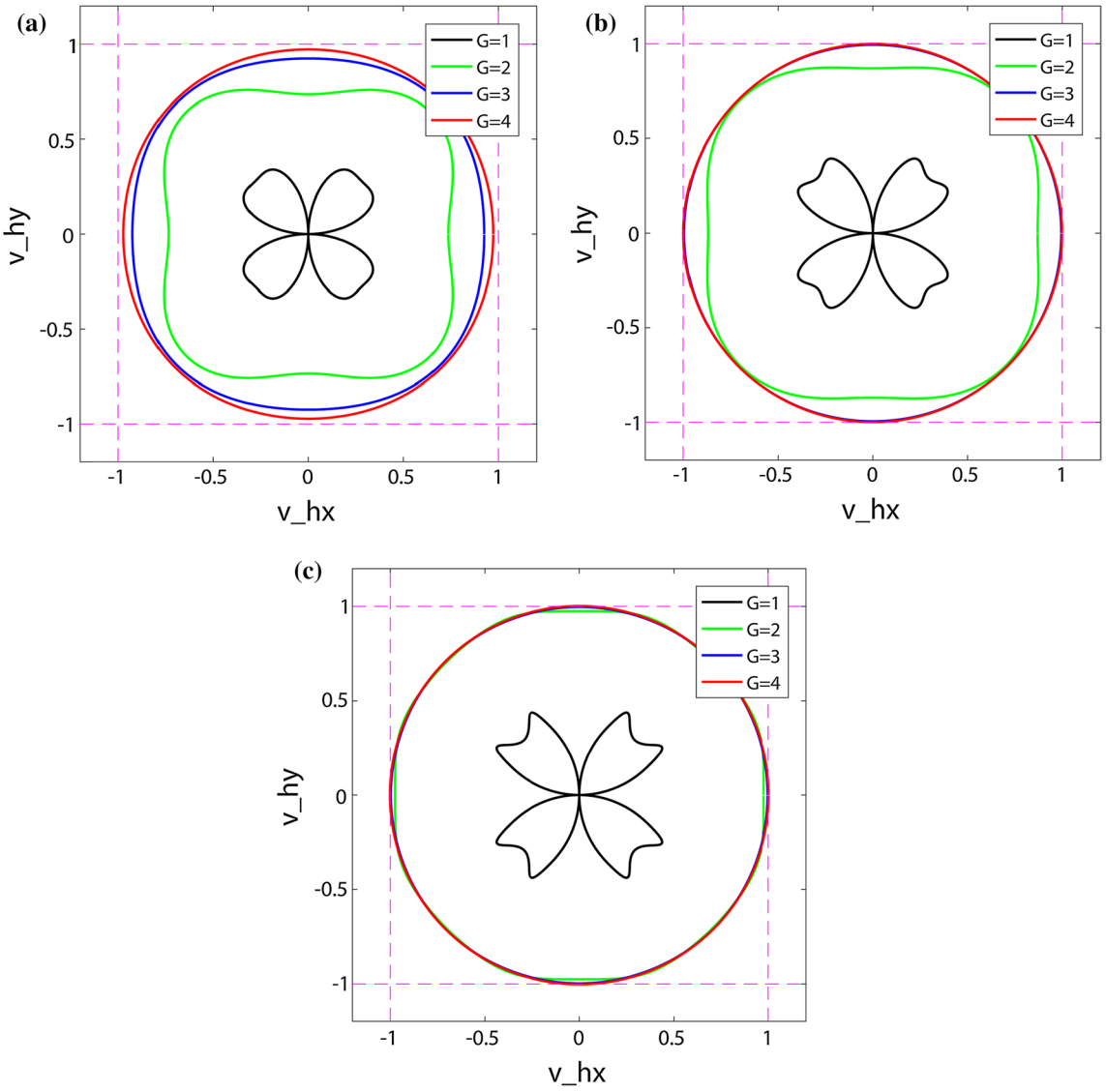


Fig. 3 Polar plot of the normalized phase velocity numerical anisotropy of second derivative for different schemes. The case of $G=1$, 2, 3, 4 is plotted: **a** the conventional fourth-order explicit FD scheme, **b** the conventional eighth-order compact FD scheme and **c** the optimized second-order compact FD scheme

wave propagation is close to numerical isotropy, and the normalized numerical phase velocity is closer to 1. As G decreases, the numerical anisotropy of wave propagation increases. There is a phenomenon that, when the propagation angles is near 45 degrees, the numerical anisotropic effect and the dispersion error are the least, which is consistent with the dispersion curves in Fig. 2. As shown in Fig. 3, in contrast to the conventional fourth-order explicit FD scheme and the conventional eighth-order compact FD scheme, the optimized second-order compact FD scheme can not only reduce the numerical dispersion, but also reduce the numerical anisotropy.

Solve the compact scheme in frequency domain

The optimized second-order compact FD scheme has obvious advantages for frequency-domain forward modeling, but we need to solve the compact FD system. Here, we derive the compact FD operator for the Laplacian operator based on the idea of Chu and Stoffa (2012). The corresponding compact spatial FD scheme can be expressed as:

$$\sum_{m=-2}^{2} \beta_m^x \frac{\partial}{\partial x^2} P_{m,0} = \sum_{i=-2}^{2} \alpha_i^x P_{i,0}, \tag{14}$$

$$\sum_{n=-2}^{2} \beta_n^z \frac{\partial}{\partial z^2} P_{0,n} = \sum_{j=-2}^{2} \alpha_j^z P_{0,j} \,. \tag{15}$$

We construct the compact FD scheme for 2D Laplacian operator, which has the following form:

$$\sum_{m=-2,n=-2}^{2,2} \tilde{\beta}_{m,n} \nabla^2 P_{m,n} = \sum_{i=-2,j=-2}^{2,2} \tilde{\alpha}_{i,j} P_{i,j} \,. \tag{16}$$

By applying Fourier transform on both sides of Eqs. (14), (15) and (16), we can obtain the following equations:

$$-k_x^2 = \frac{\sum_{i=-2}^{2} \alpha_i^x e^{i\Delta x k_x \sqrt{-1}}}{\sum_{m=-2}^{2} \beta_m^x e^{m\Delta x k_x \sqrt{-1}}}, \tag{17}$$

$$-k_z^2 = \frac{\sum_{j=-2}^{2} \alpha_j^z e^{j\Delta z k_z \sqrt{-1}}}{\sum_{n=-2}^{2} \beta_n^z e^{n\Delta z k_z \sqrt{-1}}}, \tag{18}$$

$$-k_x^2 - k_z^2 = \frac{\sum_{i=-2,j=-2}^{2,2} \tilde{\alpha}_{i,j} e^{(i\Delta x k_x + j\Delta z k_z)\sqrt{-1}}}{\sum_{m=-2,n=-2}^{2,2} \tilde{\beta}_{m,n} e^{(m\Delta x k_x + n\Delta z k_z)\sqrt{-1}}} \,. \tag{19}$$

Substituting Eqs. (17) and (18) into Eq. (19), we can get the following relationship:

$$\tilde{\alpha}_{i,j} = \alpha_i^x \beta_j^z + \alpha_j^z \beta_i^x, \tag{20}$$

$$\tilde{\beta}_{m,n} = \beta_m^x \beta_n^z \,. \tag{21}$$

Applying Eq. (16) to Eq. (1), we can get the discretized compact FD linear equation:

$$\left(A_\beta^{-1} A_\alpha + V\right) P = s, \tag{22}$$

where A_β and A_α are the discrete diagonal FD matrices, V is a diagonal matrix formed by the mass acceleration term ω^2 / v^2. Multiplying both sides of Eq. (22) by A_β, we have the following linear systems:

$$\left(A_\alpha + A_\beta V\right) P = A_\beta s \,. \tag{23}$$

Then, we can get the result of numerical simulation by solving linear Eq. (23).

Accuracy analysis

In order to confirm the accuracy of our optimized second-order compact FD scheme, we make the accuracy analysis in contrast to the analytical solution. For the acoustic wave equation in a homogeneous medium, Alford et al. (1974) gave the analytical solution as the following expression:

$$P(x, z, t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} -i\pi H_0^{(2)}\left(\frac{\omega}{v} r\right) F(\omega) \mathrm{d}\omega. \tag{24}$$

where $H_0^{(2)}$ is the second kind of Hankel function of zero order, r is the distance from the receiver to source, i is the imaginary unit and $F(\omega)$ is the frequency-domain source.

Figure 4a shows the geometric relationship between the receiver point and the source point in the homogeneous medium. We take the velocity of the homogeneous model to be 2000 m/s. The numbers of vertical and horizontal spatial grid points of the homogeneous model are both 100. The space sampling interval is taken as $h = 18$ m. We put a Ricker wavelet whose dominant frequency is 20 Hz in the center of the homogeneous model as the source. The receiver is placed at $nx = 20$ and $nz = 50$. Figure 4b shows the results computed by the optimized second-order compact FD scheme, the conventional eighth-order compact FD scheme, the conventional fourth-order explicit FD scheme and the analytical solution. Figure 4c, d shows the corresponding frequency-domain normalized seismograms for $f = 10$ HZ and $f = 30$ HZ. It is obvious that the numerical solution of the optimized second-order compact FD scheme is more consistent with the analytic solution than the other FD schemes.

Numerical examples

We examine the validity of the optimized second-order compact FD scheme on two models through numerical simulations. By contrast, we also use the conventional fourth-order explicit FD scheme and the conventional eighth-order compact FD scheme to do numerical simulation in frequency domain on two models. As a reference, we also do FD numerical simulation in time domain with the second order in time and the 12th order in space on two models.

Two-layer model

Firstly, we consider a two-layer model to test the optimized second-order compact FD scheme as shown in Fig. 5a. The up-layer velocity of the two-layer model is 2000 m/s, and the low-layer velocity is 3500 m/s.

The numbers of vertical and horizontal spatial grid points of the two-layer model are both 101. The space sampling interval is taken to be 18 m. We take the time sampling interval to be 0.002 s and the length of recorded seismogram time to be 1.0 s. The PML boundary conditions are designed around the two-layer model to eliminate the artificial boundary reflections (Berenger 1994; Yuan et al. 2014) during forward modeling. The number of the PML thickness is 25. The Ricker wavelet is used as the source, and its dominant

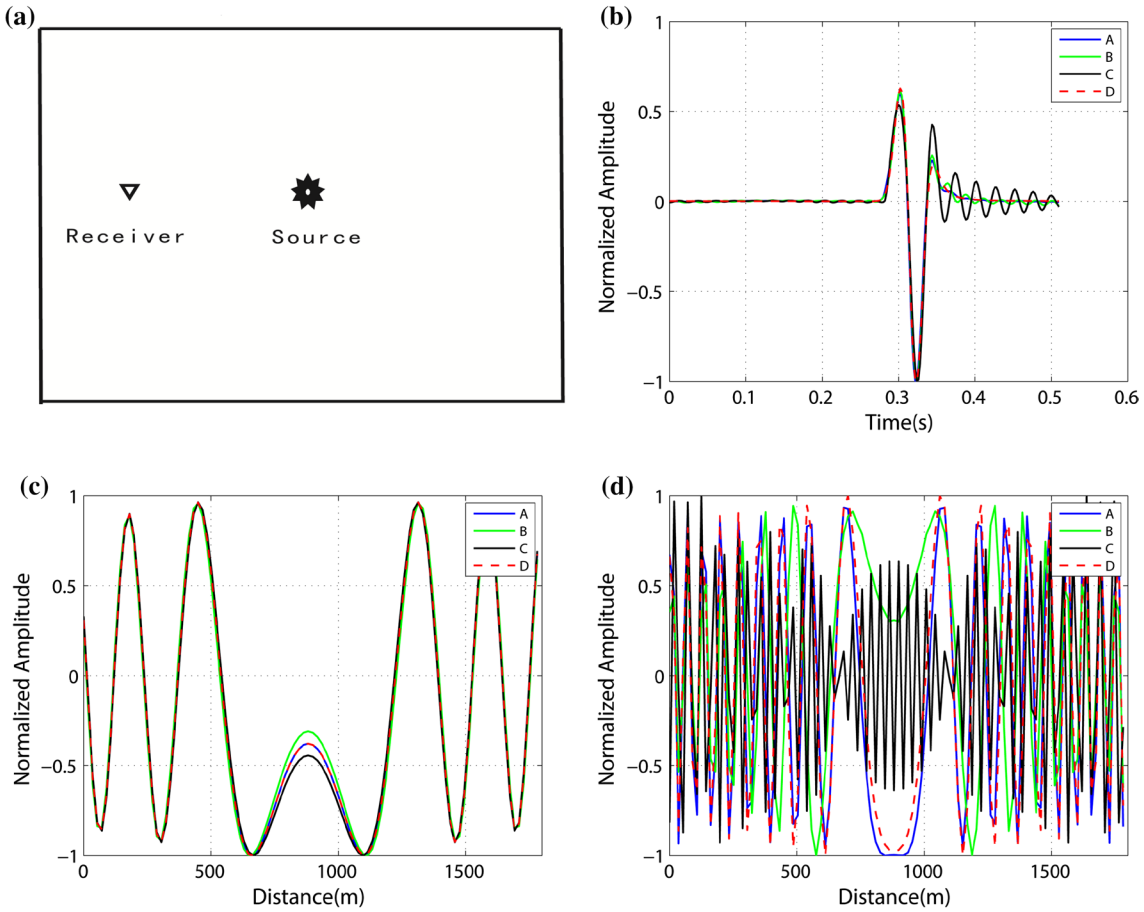


Fig. 4 **a** The geometric relationship of the receiver and the source in homogeneous model. **b** The time-domain normalized seismograms. **c** The frequency-domain normalized seismograms for $f = 10$ HZ. **d** The frequency-domain normalized seismograms for $f = 30$ HZ. Curves A, B, C and D stand for the results of the optimized second-order compact FD scheme, the conventional eighth-order compact FD scheme, the conventional fourth-order explicit FD scheme and the analytical solution

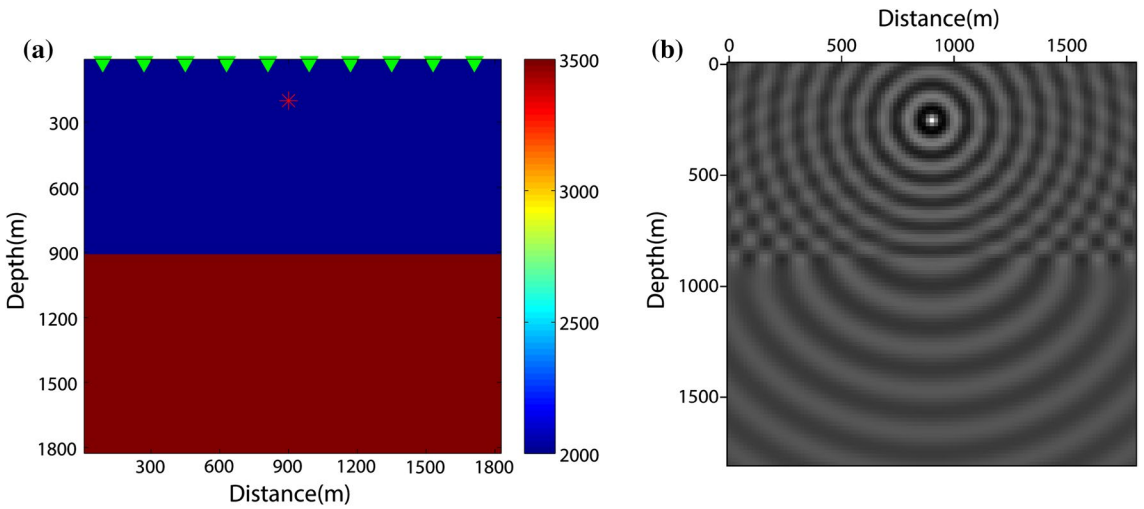


Fig. 5 **a** The two-layer model and **b** 22-Hz monochromatic wavefield computed by the optimized second-order compact FD scheme in frequency domain

frequency is 20 Hz. The source is located at $nx = 51$ and $nz = 10$. The receiver array is placed at the second level of the two-layer model.

Figure 5b shows the 22-Hz monochromatic wavefield computed by optimized second-order compact FD scheme. The position of the reflection interface is clearly visible in Fig. 5b. Figure 6 shows the time-domain seismograms computed by the optimized second-order compact FD scheme, the conventional eighth-order compact FD scheme, the conventional fourth-order explicit FD scheme in frequency-domain and FD numerical simulation in time domain with the second order in time and the 12th order in space for the two-layer model. From Fig. 6a, one can find that the conventional fourth-order explicit FD scheme has a serious numerical dispersion. As shown in Fig. 6b, the conventional eighth-order compact FD scheme can reduce the numerical dispersion in contrast to the conventional fourth-order explicit FD scheme. Figure 6c shows that the optimized second-order compact FD scheme can more significantly reduce the numerical dispersion relative to the conventional eighth-order compact FD scheme. The optimized compact FD scheme can even achieve the effect of high-precision numerical simulation in time domain. Figure 7 shows the time-domain snapshots corresponding to the above methods at 0.54 s, respectively. The incident wave, transmitted wave and reflected wave are clearly visible at the interface of medium. As shown in Fig. 7, the time-domain snapshots computed by the optimized compact FD scheme have relatively less numerical dispersion. So, we obtain the conclusion that our optimized compact FD scheme can significantly decrease the numerical dispersion and improve the precision of numerical simulation.

Salt model

As illustrated in Fig. 8a, we use the two-dimensional truncated section of SEG/EAGE salt model to further prove the effectiveness of our optimized compact FD scheme.

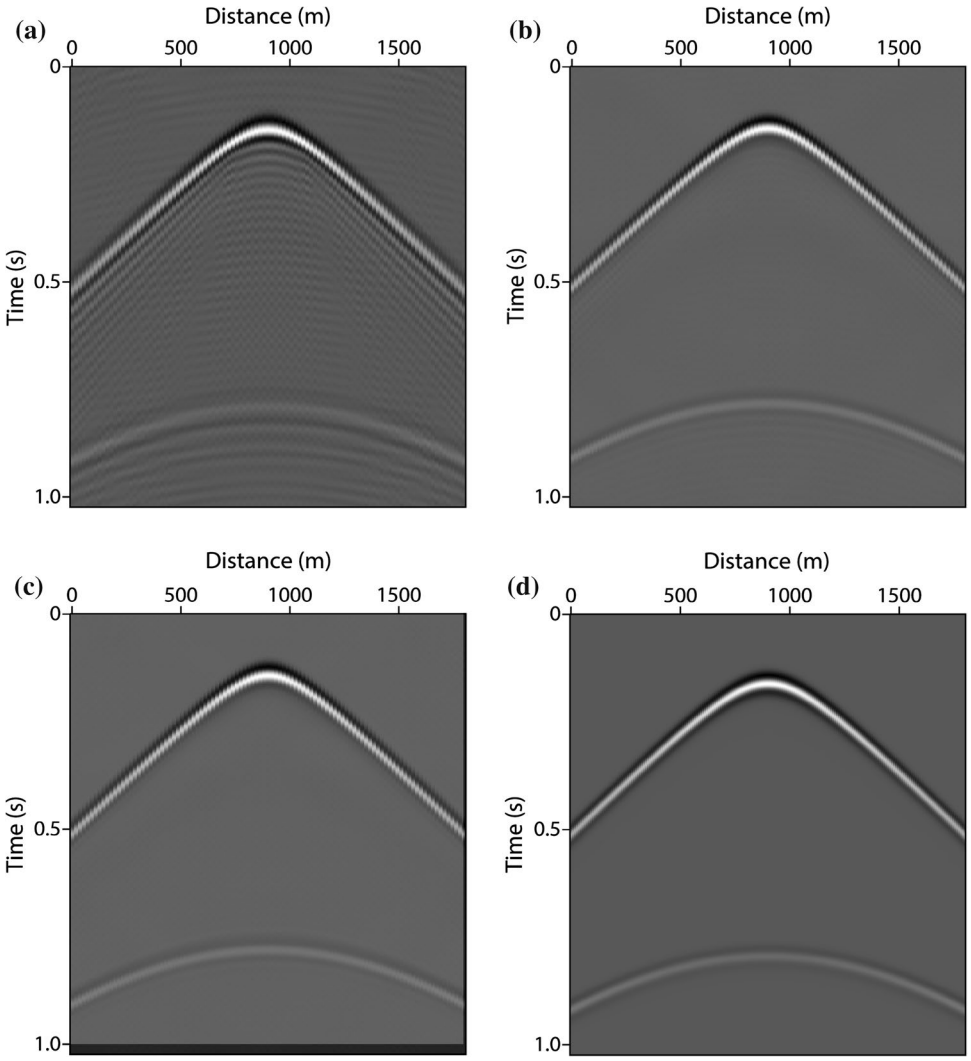


Fig. 6 Time-domain seismograms of the two-layer model. Time-domain seismograms computed by **a** the conventional fourth-order explicit FD scheme, **b** the conventional eighth-order compact FD scheme, **c** the optimized second-order compact FD scheme in frequency domain and **d** the FD numerical simulation in time domain with the second order in time and the 12th order in space

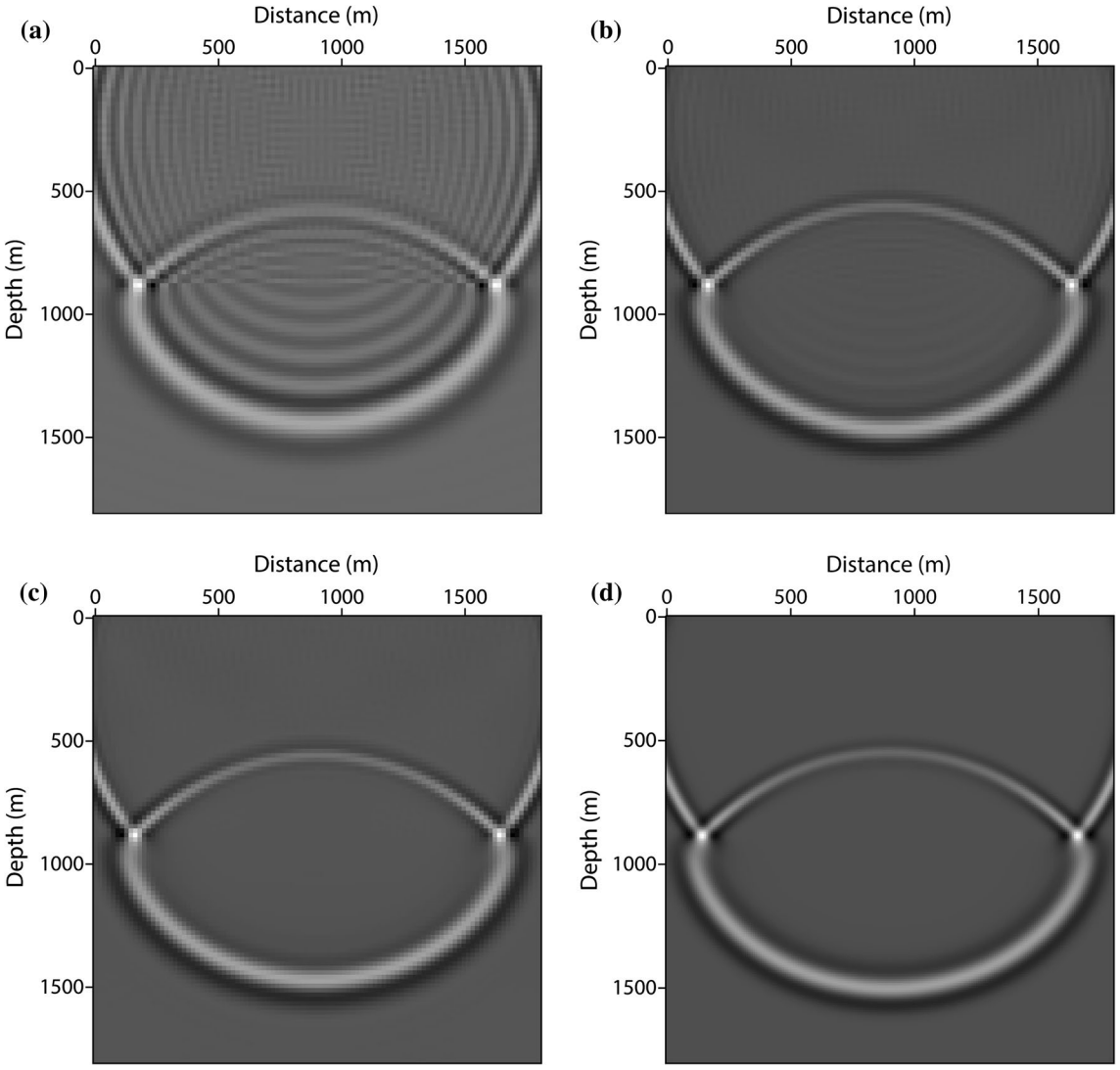


Fig. 7 Snapshots at 0.54 s. Snapshots computed by **a** the conventional fourth-order explicit FD scheme, **b** the conventional eighth-order compact FD scheme, **c** the optimized second-order compact FD scheme in frequency domain and **d** the FD numerical simulation in time domain with the second order in time and the 12th order in space

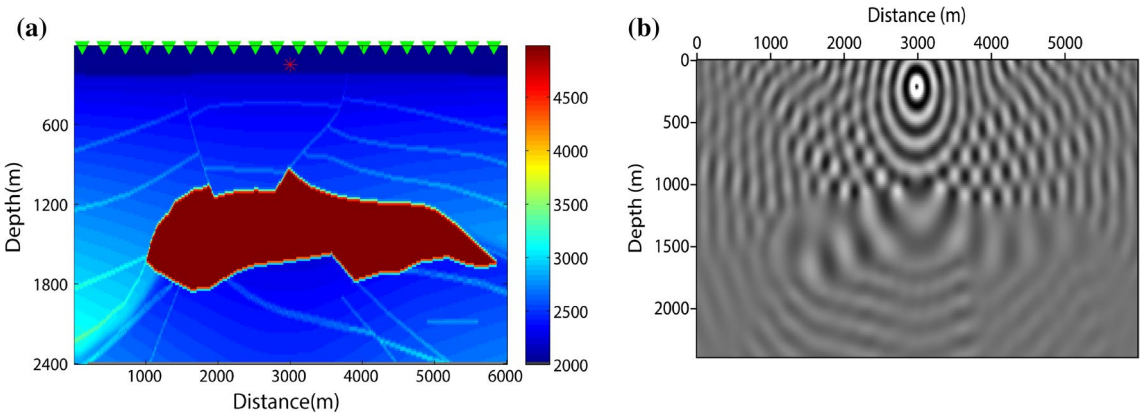


Fig. 8 **a** The section of the SEG/EAGE salt model and **b** 25-Hz monochromatic wavefield computed by the optimized second-order compact FD scheme in frequency domain

The velocity range of the salt model is from 1999.6 to 4980.8 m/s. The spatial sampling interval is set to be 15 m. The vertical and horizontal spatial grid points of the model are $nz = 160$ and $nx = 400$, respectively. The time sampling interval is set to be 0.002 s, and the recorded seismogram time is 2.5 s. We still set the PML boundary condition around the salt model to attenuate the undesired boundary reflection, and the number of PML thickness is 25. The Ricker wavelet is used as the source, and its dominant frequency is 20 Hz. The source is located at $nx = 200$ and $nz = 10$. The receiver array is placed at the first level of the salt model.

Figure 8b shows the 25-Hz monochromatic wavefield computed by optimized second-order compact FD scheme in frequency domain. The irregular variable velocity interface leads to a complicated frequency-domain wavefields in Fig. 8b. Figure 9 shows the time-domain seismograms computed by the optimized second-order compact FD scheme, the conventional eighth-order compact FD scheme, the conventional fourth-order explicit FD scheme

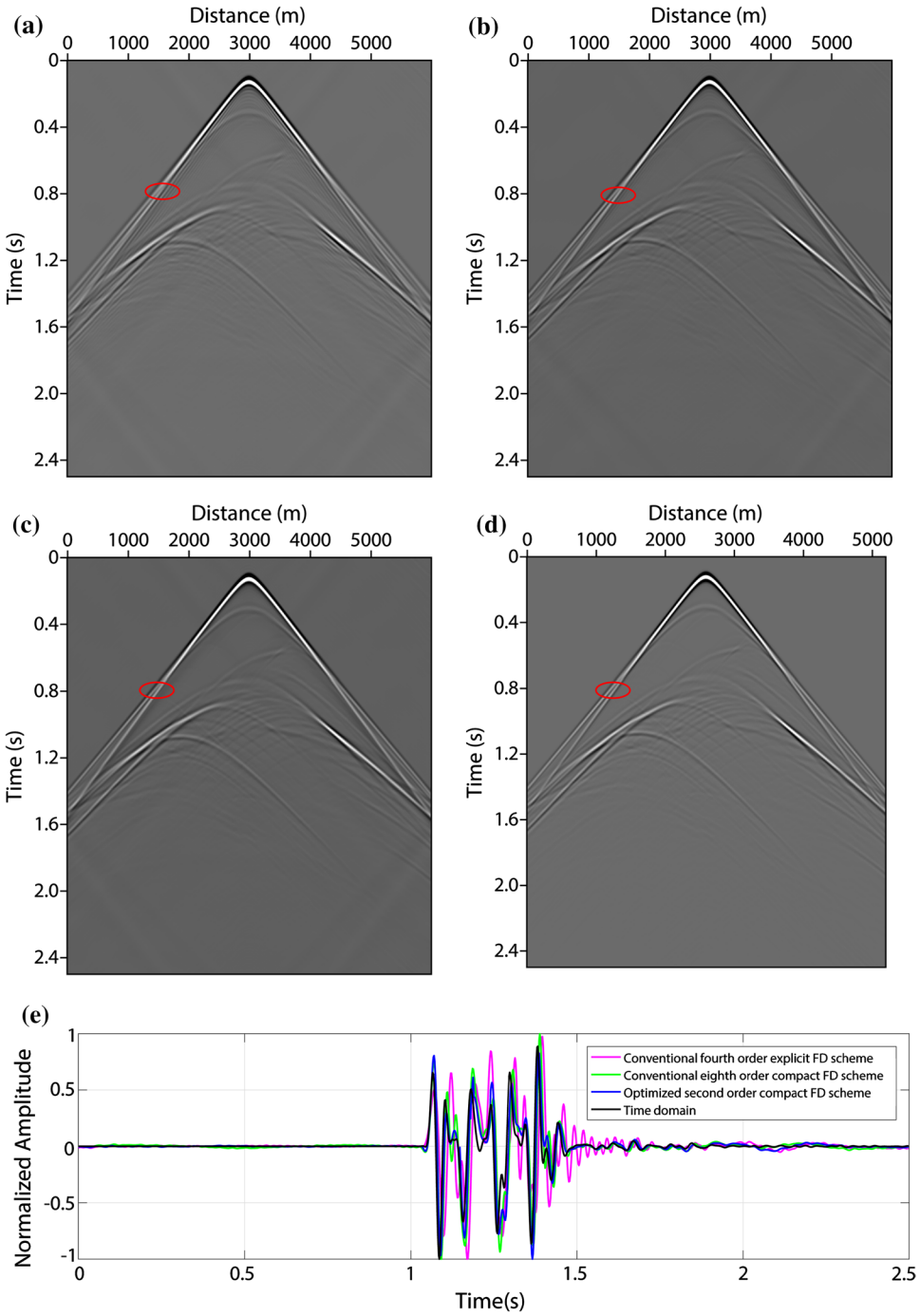


Fig. 9 Time-domain seismograms of the salt model. Time-domain seismograms computed by **a** the conventional fourth-order explicit FD scheme, **b** the conventional eighth-order compact FD scheme, **c** the optimized second-order compact FD scheme in frequency domain and **d** the FD numerical simulation in time domain with the second order in time and the 12th order in space. **f** Comparison of single trace extracted from the above results at 1250 m

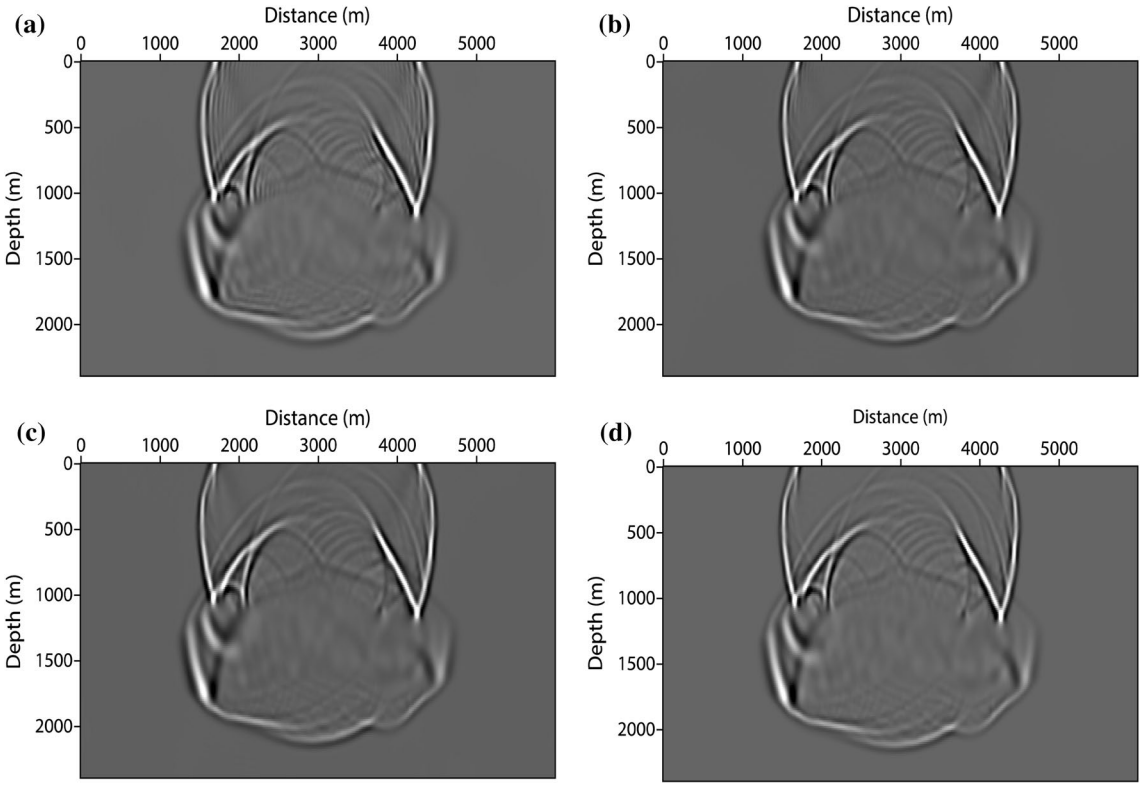


Fig. 10 Snapshots at 0.72 s. Snapshots computed by **a** the conventional fourth-order explicit FD scheme, **b** the conventional eighth-order compact FD scheme, **c** the optimized second-order compact FD scheme in frequency domain and **d** the FD numerical simulation in time domain with the second order in time and the 12th order in space

in frequency-domain and FD numerical simulation in time domain with the second order in time and the 12th order in space for the salt model. Figure 9f shows single trace extracted from the time-domain seismograms computed by different methods at 1250 m. Figure 10 shows the time-domain snapshots corresponding to the above methods at 0.72 s, respectively. The irregular variable velocity interface also makes the time-domain seismograms and snapshots become very complex. By comparing and analyzing the results of the salt model, we still obtain the conclusion that the optimized compact FD scheme can significantly decrease the numerical dispersion and improve the precision of numerical simulation for the salt model.

Conclusions

We have introduced an optimized compact FD scheme for acoustic wave numerical simulation in frequency domain. We perform optimization to obtain the coefficients of the optimized compact FD scheme and make dispersion analysis. The dispersion analysis shows that the dispersion errors of the optimized compact FD scheme are smaller than those of the conventional compact FD scheme, and it can also reduce the numerical anisotropy. To avoid the extra memory and computational costs in the numerical simulation, we use the method that calculates the compact FD operator for the Laplace operator. This is followed by accuracy analysis. Accuracy analysis shows the numerical solution of the optimized second-order compact FD scheme is more consistent with the analytical solution than the results of conventional eighth-order compact FD scheme. Finally, the numerical simulations of the two-layer model and salt model are used to verify the theoretical analysis of the optimized compact FD scheme. We use the results of the optimized second-order compact FD scheme, the conventional fourth-order explicit FD scheme and the conventional eighth-order compact FD scheme to make a comparison. As a reference, we also do time-domain forward modeling with the second order in time and the 12th order in space on two models.

Acknowledgements This research is supported by National Key Scientific Instrument and Equipment Development Project (Grant No. 2018YFF01013503), National Natural Science Foundation of China (Grant No. 41630319) and Major State Research Development Program of China (Grant No. 2016YFC0601101).

References

Alford RM, Kelly KR, Boore DM (1974) Accuracy of finite-difference modeling of the acoustic wave equation. Geophysics 39:834–842. https://doi.org/10.1190/1.1440470

Alterman Z, Karal FC (1968) Propagation of elastic waves in layered media by finite difference methods. Bull Seismol Soc Am 58:367–398

Baysal E, Dan DK, Sherwood JWC (1983) Reverse time migration. Geophysics 48:1514–1524

Berenger JP (1994) A perfectly matched layer for the absorption of electromagnetic waves. J Comput Phys 114:185–200

Chang WF, McMechan GA (1987) Elastic reverse-time migration. Geophysics 52:1365–1375. https://doi.org/10.1190/1.1442249

Chen JB (2012) An average-derivative optimal scheme for frequency-domain scalar wave equation. Geophysics 77:T201–T210. https://doi.org/10.1190/geo2011-0389.1

Chen JB (2014) A 27-point scheme for a 3D frequency-domain scalar wave equation based on an average-derivative method. Geophys Prospect 62:258–277

Chen JB, Cao J (2016) Modeling of frequency-domain elastic-wave equation with an average-derivative optimal method. Geophysics 81:T339–T356. https://doi.org/10.1190/geo2016-0041.1

Chu C, Stoffa PL (2012) An implicit finite-difference operator for the Helmholtz equation. Geophysics 77:97

Hustedt B, Operto S, Virieux J (2004) Mixed-grid and staggered-grid finite-difference methods for frequency-domain acoustic wave modelling. Geophys J R Astron Soc 157:1269–1296

Jo C, Shin C, Suh JH (1996) An optimal 9-point, finite-difference, frequency-space, 2-D scalar wave extrapolator. Geophysics 61:529–537

Kim JW, Lee DJ (1996) Optimized compact finite difference schemes with maximum resolution. J AIAA 34:887–893

Lele SK (1992) Compact finite difference schemes with spectral-like resolution. J Comput Phys 103:16–42

Liu Y, Sen MK (2009) A practical implicit finite-difference method: examples from seismic modelling. J Geophys Eng 6:231

Lysmer J, Drake LA (1972) A finite element method for seismology. Methods Comput Phys Adv Res Appl 11:181–216

Marfurt KJ (1984) Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations. Geophysics 49:533–549

Marfurt KJ, Shin CS (1989) The future of iterative modeling in geophysical exploration. Handb Geophys Explor Seism Explor 21:203–228

McMechan GA (1983) Migration by extrapolation of time-dependent boundary values. Geophys Prospect 31:413–420

Mora P (1987) Nonlinear two-dimensional elastic inversion of multi-offset seismic data. Geophysics 52:1211–1228

Operto S, Virieux J, Amestoy P, Giraud L, L'Excellent JY (2007) 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: a feasibility study. Geophysics 72:195

Pratt RG (1990) Frequency-domain elastic wave modeling by finite differences: a tool for crosshole seismic imaging. Geophysics 55:626–632

Pratt RG, Worthington MH (1990) Inverse theory applied to multi-source cross-hole tomography. Part I: Acoustic wave-equation method. Geophys Prospect 38:287–310

Pratt RG, Shin C, Hick 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

Shi P, Angus D, Nowacki A, Yuan S, Wang Y (2018) Microseismic full waveform modeling in anisotropicmedia with moment tensor implementation. Surv Geophys 39:567–611. https://doi.org/10.1007/s10712-018-9466-2

Shin C, Cha YH (2008) Waveform inversion in the Laplace domain. Geophys J Int 173:922–931

Shin C, Sohn H (1998) A frequency-space 2-D scalar wave extrapolator using extended 25-point finite-difference operator. Geophysics 63:289–296. https://doi.org/10.1190/1.1444323

Tang X, Liu H, Zhang H, Liu L, Wang Z (2015) An adaptable 17-point scheme for high-accuracy frequency-domain acoustic wave modeling in 2D constant density media. Geophysics 80:T211–T221. https://doi.org/10.1190/geo2014-0124.1

Tarantola A (1984) Inversion of seismic reflection data in the acoustic approximation. Geophysics 49:1259–1266

Virieux J (1984) SH-wave propagation in heterogeneous media: velocity-stress finite-difference method. Geophysics 49:1933–1942. https://doi.org/10.1190/1.1441605

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

Virieux J, Operto S (2009) An overview of full-waveform inversion in exploration geophysics. Geophysics 74:WCC1–WCC26

Whitmore ND (1983) Iterative depth migration by backward time propagation. Seg Tech Program Expand Abstr 2:646

Yuan S, Wang S, Sun W, Miao L, Li Z (2014) Perfectly matched layer on curvilinear grid for the second-order seismic acoustic wave equation. Explor Geophys 45:94–104. https://doi.org/10.1071/EG13066

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 Y, Xu S, Tang B, Bai B, Huang Y, Huang T (2010) Angle gathers from reverse time migration. Lead Edge 29:1364–1371. https://doi.org/10.1190/1.3517308

Acta Geophysica (2019) 67:1403–1411
https://doi.org/10.1007/s11600-019-00333-y

RESEARCH ARTICLE - APPLIED GEOPHYSICS

Optical fiber marine seismic exploration system feasibility study

Xiangchun Wang[1] · Xueliang Li[2] · Haibing Zhang[3] · Xiaozhu Hao[4] · Kaiben Yu[5]

Received: 30 May 2019 / Accepted: 14 July 2019 / Published online: 17 July 2019

Abstract
The signal-to-noise ratio (SNR), vertical resolution and fidelity of marine seismic data are closely related to the characteristics of hydrophones, namely the sensitivity and the frequency band range. In this article, we analyzed and processed seismic data collected in 2017 from a towed marine streamer equipped with optical fiber cable and hydrophones. Results from a pre-stack time migration section revealed that the optical fiber seismic data yielded high SNR, vertical resolution and fidelity. For validation purpose, a comparison of these data with those acquired by classic piezoelectric hydrophones along a migration section located in a very close location was carried out. This comparison showed that the SNR, fidelity and vertical resolution from the two means of measurements were comparable.

Keywords Fidelity · Hydrophones · Migration section · Optical fiber · Signal-to-noise ratio (SNR) · Vertical resolution

Introduction

Hydrophones constitute the main means of measurement of marine seismic data, and their performance is evaluated based on the signal-to-noise ratio [SNR = 10log(Ps/Pn), here Ps and Pn are the effective power of signal and noise], vertical resolution and fidelity. Piezoelectric hydrophones have been typically used (Hart 2007) for marine seismic exploration but present various disadvantages such as low sensitivity, small dynamic range and poor anti-interference ability. These shortcomings lead to the loss of abundant low-frequency information contained in the seismic data. Over the past two decades, the optical fiber technology has been developing rapidly and has since become increasingly used in geophysical exploration (Shao et al. 2011; Luo et al. 2012; Wang et al. 2010; Ni et al. 2004), as the optical fiber allows for the fast and continuous transmission of massive amounts of data (Peng and Zhang 2018; Keith 2018; Yun et al. 2012; Zhang and Ni 2004). Optical fiber hydrophones have also a wider bandwidth, higher sensitivity and a better high frequency response. For example, the low frequency of optical fiber hydrophones can be 1 Hz while the low frequency of piezoelectric hydrophones now usually is 3 Hz. These characteristics have greatly improved the quality of the marine seismic data, although several aspects have yet to be improved.

In this study, an innovative seismic data hardware collection system using optical fiber cable and hydrophones is introduced in "Hardware collection system" section. The acquisition parameters of one marine seismic exploration experiment conducted in 2017 by Guangzhou Marine Geological Survey with this system is then presented in "Data acquisition" section, followed by the analysis of the different types of noise measured and a description of the various steps necessary to process the data, such as noise elimination, velocity analysis and suppression of multiples ("Data processing" section). The results of a pre-stack time migration section were then analyzed in "Migration results" section, revealing the high SNR, vertical resolution and fidelity of the system. A comparison with traditional piezoelectric hydrophones seismic data is also presented in this section, as a way to validate our results. Concluding remarks are provided in "Discussion and conclusion" section.

✉ Xiangchun Wang
wangxc@cugb.edu.cn

[1] School of Geophysics and Information Technology, China University of Geosciences (Beijing), Beijing 100083, China

[2] Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100081, China

[3] Appsoft Technology Co., Ltd, Beijing 100085, China

[4] Guangzhou Marine Geological Survey of China Geological Survey, Guangzhou 510760, China

[5] National Deep Sea Center, Ministry of Natural Resources, Qingdao 266237, China

Hardware collection system

The hardware collection system was developed by Beijing Appsoft Technology Co., Ltd. The system uses interferometric fiber-optic hydrophones and is composed of an optical fiber hydrophone array, a light source optical modulation amplification system, a photoelectric signal processing system, an hydro-acoustic signal processing unit, and an integrated information storage display control setup (Hao et al. 2018). The complete system basic principle can be summarized as follows: The laser light emitted by the laser device is divided into two paths by fiber coupling with one path constituting the signal arm of the fiber interferometer which receives the modulation of the sound wave. The other path constitutes the reference arm, which does not receive the modulation of the acoustic wave, contrary to that of the information arm. The optical signal receiving the modulated acoustic wave is reflected by the rear end reflection film and then returned to the fiber coupler causing an interference. The interfered optical signal is then converted into an electrical signal by a photodetector, following which the information of the acoustic wave can be obtained via signal processing.

Data acquisition

In 2017, the optical fiber seismic data collection system was used to collect marine seismic data along the survey line shown in Fig. 1. The line is located in the northern continental margin of the South China Sea, and in this area the gas hydrate has been found by many means. The acquisition parameters were as follows: The low-cut frequency was 5 Hz (the low-cut frequency can be selected, indeed the low frequency of optical fiber hydrophones can be 1 Hz), the cable length 800 m, the number of traces in one shot gather was 256, the first part (A) of the cable was located at a depth of 5 m, and the second part (B) at 10 m. The channel spacing was 3.125 m and there was only one hydrophone per channel. The shot spacing was 25 m and the time sampling interval 0.25 ms, whereas the time record length was 7 s and the minimum offset was 191 m. The total length of the survey line was 48 km.

To compare the optical fiber hydrophones' ability for marine seismic exploration, seismic data using piezoelectric hydrophones collected in 2015 by Guangzhou Marine Geological Survey are also processed. The acquisition system was SEAL428 with the following acquisition parameters: the low-cut frequency was 3 Hz, the cable length was 750 m, the number of traces in one shot gather was 120, and the cable depth was 5 m under the sea surface. The

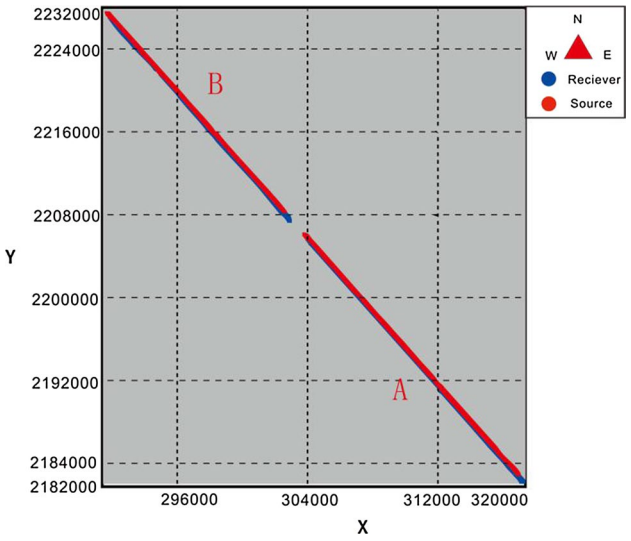


Fig. 1 Survey Geometry. The red line is the source line and the blue line is the receiver line. As ocean waves occurred during the experiment, the source and receiver lines do not coincide perfectly

channel spacing was 6.25 m and the shot spacing 25 m. The time sampling interval was 0.25 ms with a recording time length of 5 s, and the minimum offset was 100 m. The total length of the survey line was also 48 km.

The shot lines of the two surveys are almost at the same position, and the receiver lines' interval of the two surveys was about 320 m; therefore, the reflection point lines' interval of the two surveys was about 160 m.

Data processing

Processing workflow

Standard marine seismic data processing techniques are adopted for processing both the optical fiber hydrophones' data and the piezoelectric hydrophones' data. Because the channel interval of the optical fiber acquisition system is 3.125 m and that of the piezoelectric acquisition system is 6.25 m, two channel data of the optical fiber system were selected for processing. In order to reduce the amount of data and further harmonize the parameters of the two survey's data, the time sample interval was resampled to 1 ms for the two sets of data. The processing steps and the processing parameters of the two datasets are similar except for the noise elimination of the optical fiber system, whose data include noises which present different characteristics. The whole processing workflow is shown in Fig. 2. The main processing techniques are a pre-stack noise attenuation for improving the SNR, amplitude compensation for improving the energy of the deep reflection events, predictive deconvolution for enhancing the data's vertical resolution, multiples

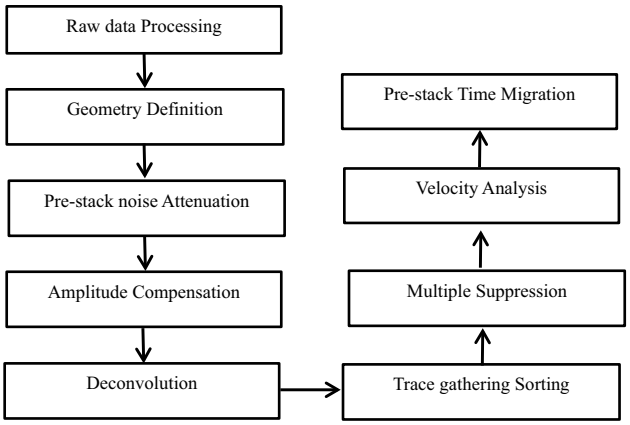


Fig. 2 Seismic data processing workflow

attenuation for eliminating multiple echoes, and velocity analysis for achieving the velocity model. Finally, the pre-stack time migration is conducted and the final pre-stack time migration sections were obtained.

Noise analysis and elimination

Noise suppression is a very important step in the processing of marine seismic data. Under normal circumstances, underwater cables are mainly affected by three types of noise: mechanical noise, turbulent noise and acoustic noise. Acoustic noise is caused by the external environment, while mechanical noise and turbulence noise are related to the hydrophones. Here, four shot gather optical fiber seismic records are shown in Fig. 3 with the water depth decreasing from the left to the right panels. The characteristics of the noise are different in each shot gather, which differs from the piezoelectric hydrophones' data whose noise is almost similar in each shot gather. Figure 3a–d shows that all the low-frequency linear noise is from the shot side to the receiver side, and its apparent velocity is about 1019 meters per second (m/s).

The spatial periodic noise in the seismic record within a distance of 16 channels (50 m) is shown in Fig. 4. The whole optical fiber cable consists of 16 short optical fiber cables whose length is 50 m, connected with each other by pairs of short cables. Possible flaws in the connection design may cause some spatial periodic noise in the data record. Another type of noise, possibly caused by surge waves, is apparent from the data and is shown in Fig. 5. It is relatively concentrated, has a strong energy, and its velocity is about 30 m/s for a main frequency of 3.3 Hz.

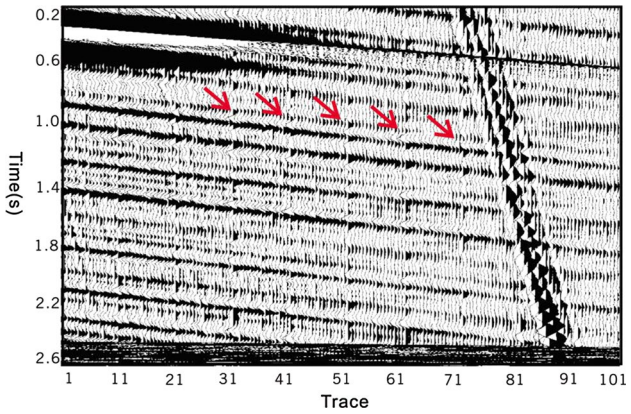


Fig. 4 Spatial periodic noise. The noise shown here may be caused by the connection of the cable

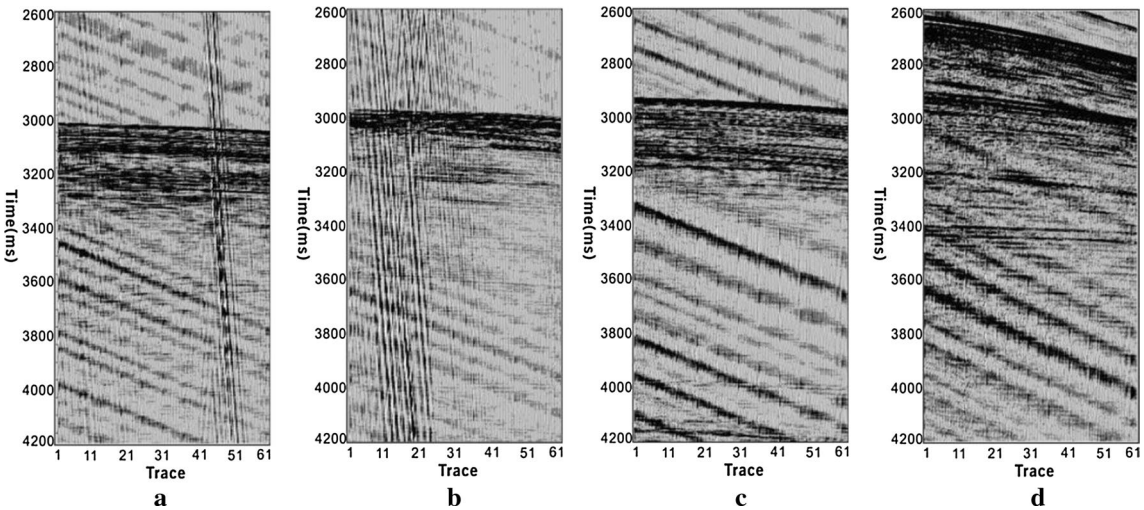


Fig. 3 Four shot gather seismic records. From left to the right, the water depth decreases. Note that the noise in the four shot gathers data is not consistent. This phenomenon differs from the piezoelectric hydrophones' noise, which is similar in each shot gather

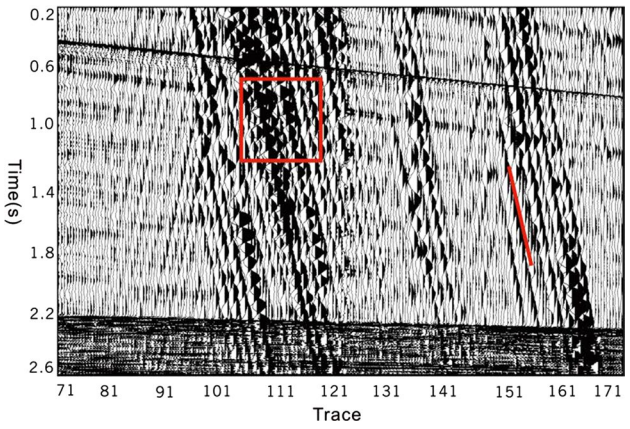


Fig. 5 Low-velocity and low-frequency noise. The energy of the noise is strong; its main frequency is about 3.3 Hz, and the apparent velocity is about 30 m/s. This may be a surge noise

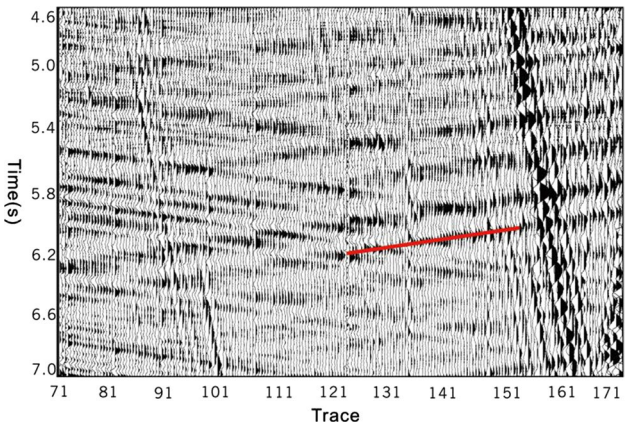


Fig. 6 Low-velocity linear noise. The linear noise is seen spreading from the receiver side to the shot side with an apparent velocity of about 1093 m/s

Lastly, a low-velocity and linear noise is seen spreading from the receiver side to the shot side in Fig. 6 with an apparent velocity of about 1093 m/s. It presents good linear and regular features.

Overall, the optical fiber system data are affected by four kinds of noise against only one type of noise (surge noise) for the piezoelectric acquisition system data. The noise from the optical fiber system is suppressed via the frequency division de-noising technique. The seismic data are first separated into several frequency bands, whose noise is then suppressed. The final processed signal is then obtained by combining the various processed bands. Figure 7a, b, which displays the shot gather prior and following the noise suppression processing, demonstrates the efficiency of the method, as most of the noise has been suppressed and only weak residual noise remains.

Multiples elimination

As a strong impedance difference interface exists between the air and the sea water, multiple echoes (hereafter referred to as multiples) are generated at the interface during the wave propagation. The suppression of multiples is not trivial as the maximum offset of the data is less than 1000 m. Here, we use successively several methods to suppress the multiples, i.e., the surface-related multiple elimination (SRME), predictive deconvolution and the radon transform. The SRME technology is a data-driven mode, as it convolutes the data with itself in both the time and space domains to predict the occurrence of multiple waves, and does not require any information about the bottom properties. The results of the multiples suppression are exemplified by the stack section shown in Fig. 8, before (Fig. 8a) and after (Fig. 8b) implementation of this processing. The number and intensity of

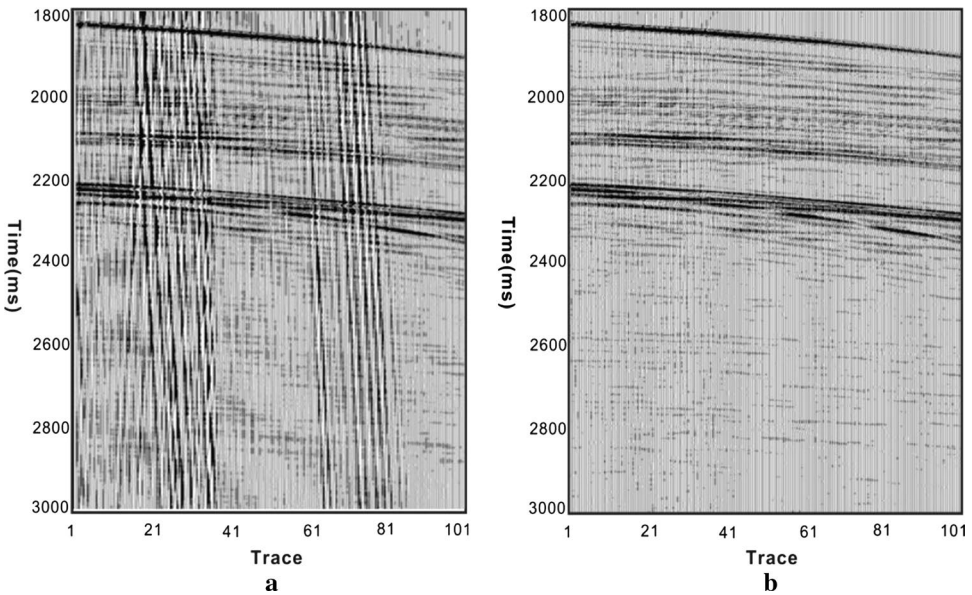


Fig. 7 Effects of noise suppression by the frequency division technique. **a** Shot gather prior and **b** after the noise suppression processing

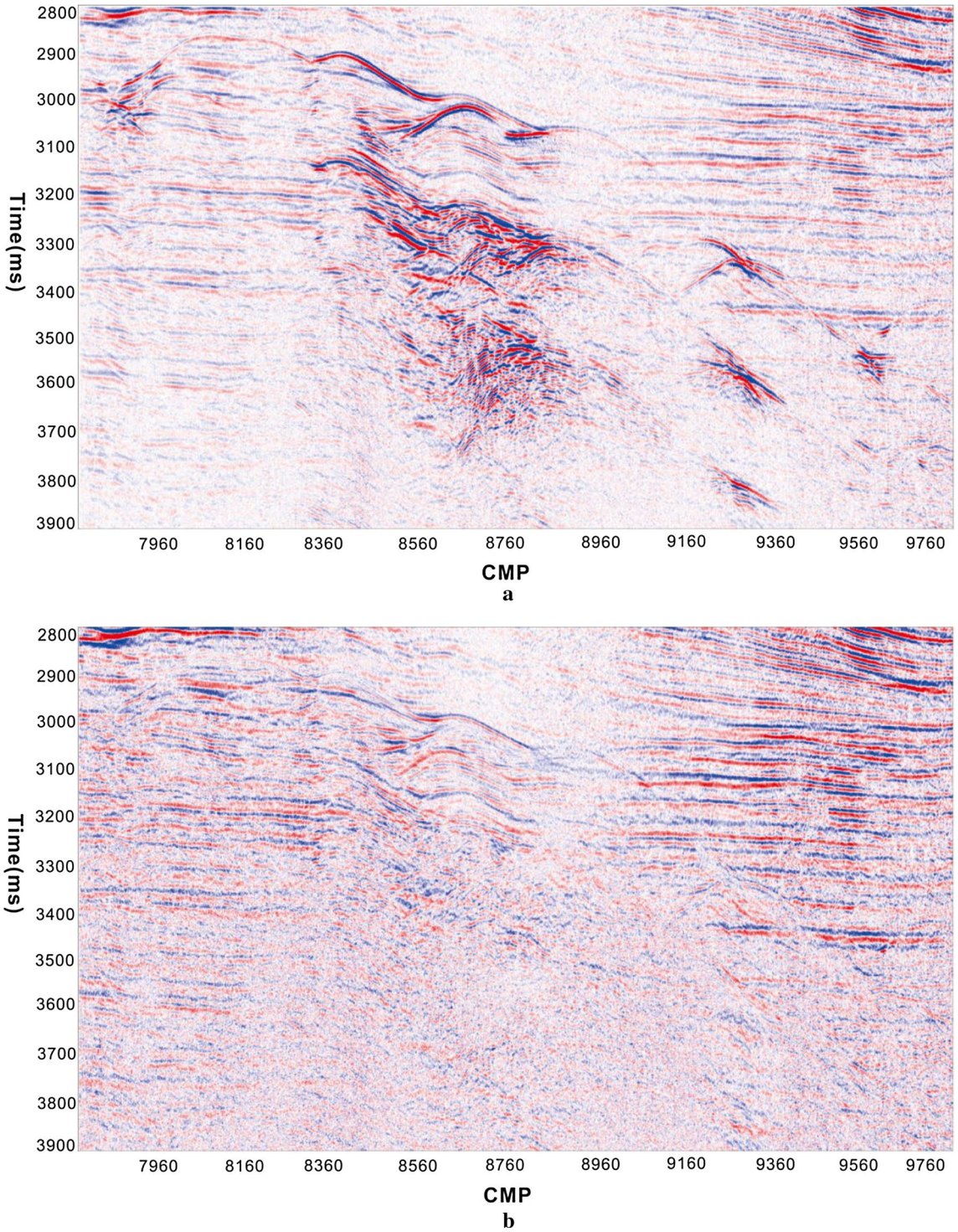


Fig. 8 Multiples suppression. Example of a portion of the stack section **a** before and **b** after suppression of the multiples. The data were processed by Radon transform, the SRME and predictive deconvolution

the multiples have been considerably reduced, and the few remaining residuals are possibly peg-leg multiples that were impossible to eliminate.

Velocity analysis and velocity model

Based on the approximately hyperbolic shape of the reflection events, the seismic data are extracted into common

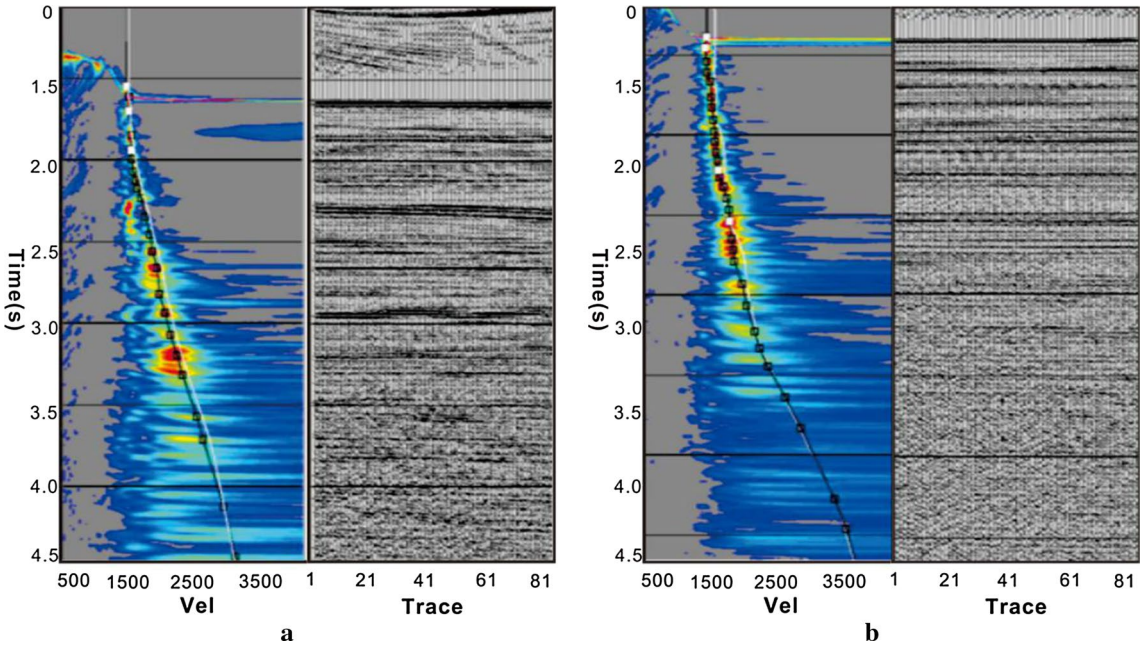


Fig. 9 Velocity spectra and the CMP gather. CMP gather (right) and its velocity spectra (left) before (**a**) and after (**b**) the applications of the various processing steps

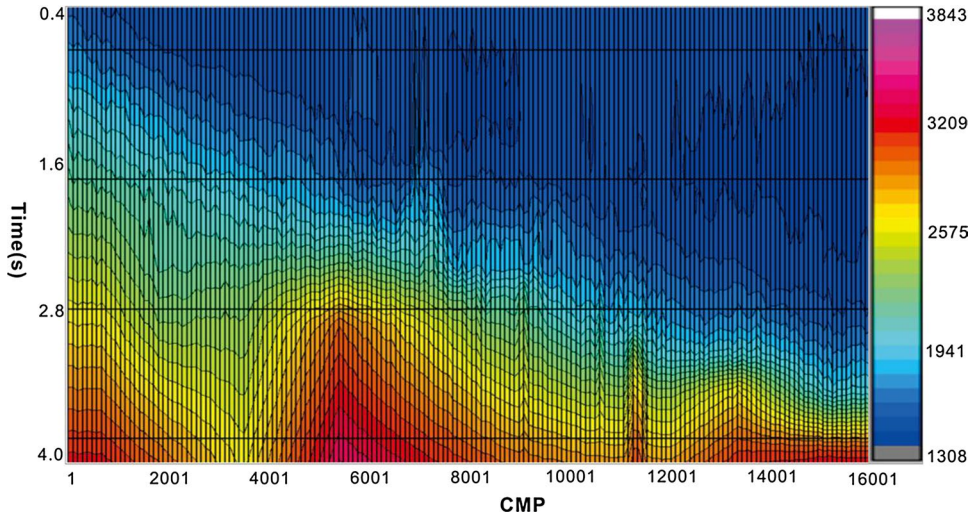


Fig. 10 Velocity Field. A local high velocity is visible in the middle of the survey section, and the velocity distribution is consistent with the structural characteristics

middle point (CMP) gathers in order to obtain the velocity field via a velocity analysis. Figure 9a, b presents the velocity spectra from a CMP trace before and after the various processing steps, respectively. The velocity spectrum after correction is seen to be better focused, which makes the velocity curve (black line in Fig. 9) easier to detect. Note that the trace of multiples velocity in the spectrum is hardly distinguishable following the suppression of the multiples in the data. Following the noise suppression processing, the CMP gather data present a higher SNR, while the reflection events are almost flattened after applying the normal move out (NMO) correction.

Using the velocity profile detected after the various processing steps, the velocity field is obtained, as shown in Fig. 10. Here, the velocity distribution is consistent with the structural characteristics of the bottom, and a local high velocity area can be seen in the middle of the survey line. This velocity field is further used for the pre-stack time migration processing.

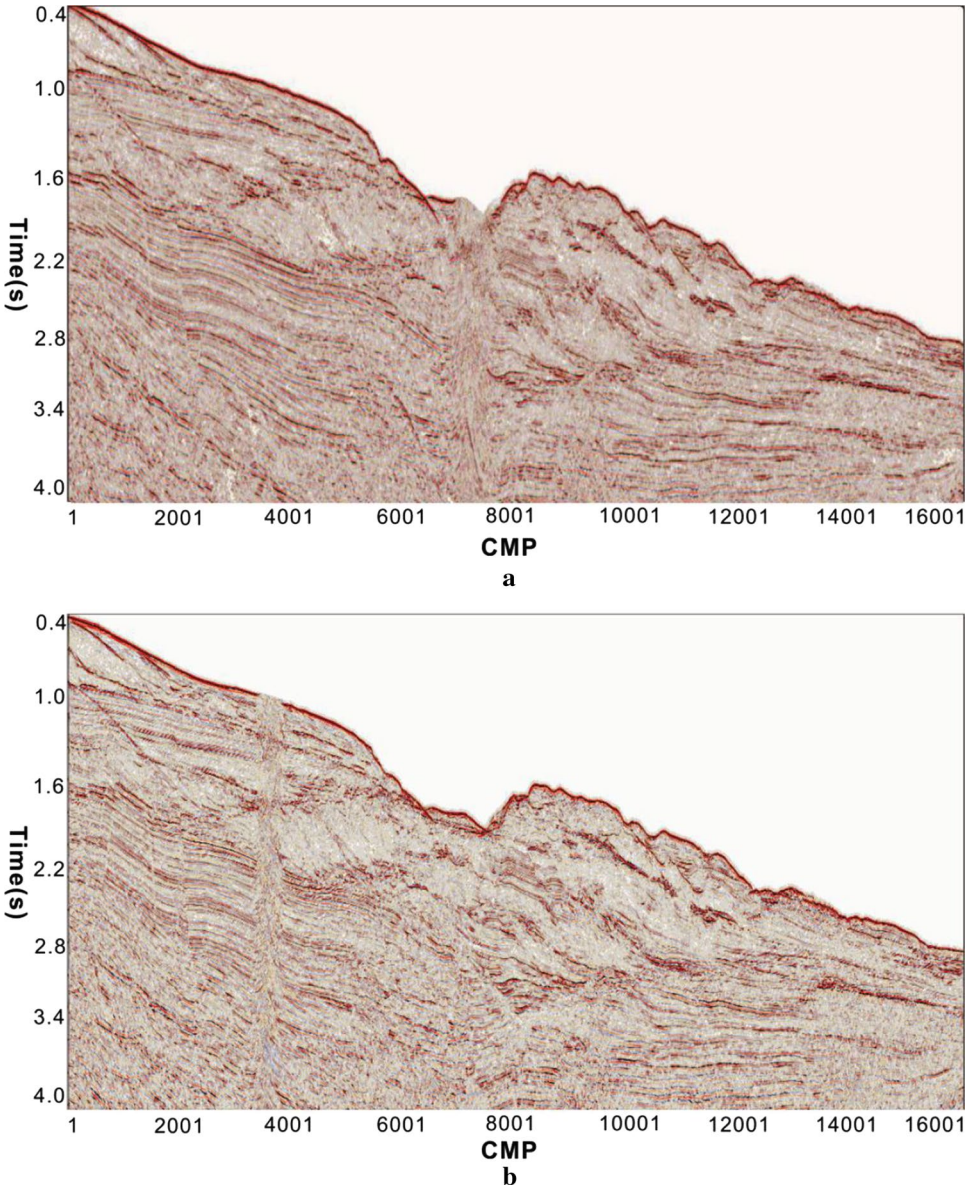


Fig. 11 Migration sections. Pre-stack time migration section from **a** the optical fiber hydrophones' data and **b** the piezoelectric hydrophones' data

Migration results

Figure 11a, b presents the pre-stack time migration section obtained from the optical fiber hydrophones' data and from the piezoelectric hydrophones' data, respectively, while Fig. 12a, b shows a close-up view of the pre-stack time migration section from each of the two systems. These figures reveal that the overall geological structure is clear, and the two sections resolution is equivalent and present a similar SNR. In Fig. 11a, there is a blur area around CMP 8001; this is because there are no data collected. Because the length of the migration sections is 48 km, it is hard to distinguish their differences in Fig. 11. Only limited differences exist in some geological details in Fig. 12.

Discussion and conclusion

The seismic data from the optical fiber system typically present four kinds of noise. The presence of the two types of noise whose velocity approximates 1000 m/s may be caused by our use of the piezoelectric system's auxiliary devices for this experiment, while the noise whose velocity is about 30 m/s may be a surge noise. The difference between the noises from each system may be explained by the presence of only one hydrophone per channel for the optical system versus several hydrophones per channel for the piezoelectric hydrophones system. The periodic noise issue has been proved to be caused by faulty cable connections and has since been solved. After suppression of these various types of noise, the comparison of the pre-stack

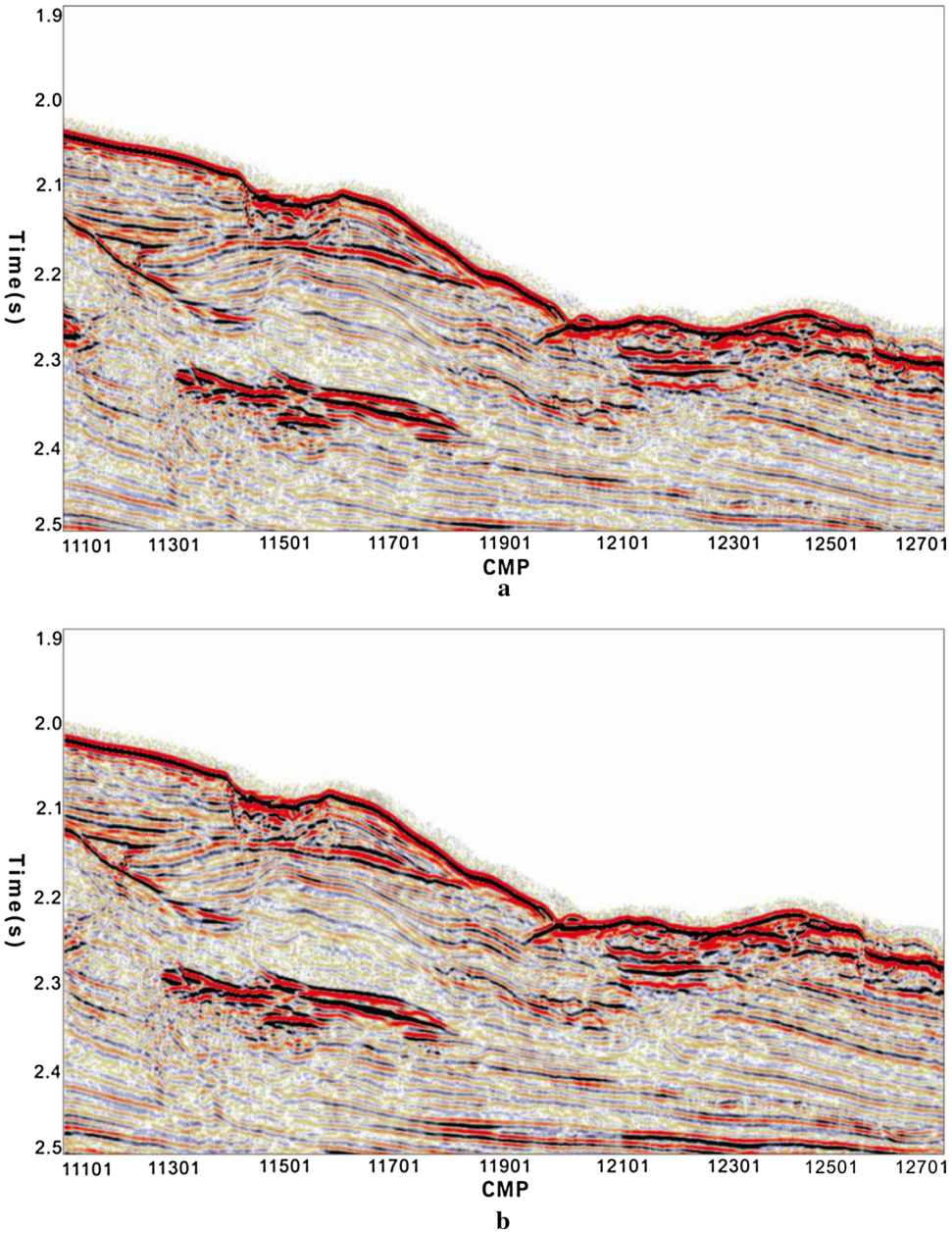


Fig. 12 Close-up view on the migration sections. **a** As obtained by the optical fiber hydrophones' data and **b** by the piezoelectric hydrophones' data

time migration sections obtained via the two systems exhibited little difference. The small differences between the two sections data may be explained by the 160 m interval between the two sections, which should induce small differences in the geological structures observed. Yet, the geological structure seen by both systems was consistent overall, and the resolution and SNR of the optical fiber system were both very high. These results thus validated the feasibility of data acquisition, and the quality of the optical fiber cable and hydrophone system for marine seismic surveying. For the optical fiber system needs little electric energy, in the future the optical fiber cable can be designed longer than the electric cable.

Acknowledgements This research was funded by the national key research and development program of China (2017YFC0307405), the National Natural Science Foundation of China (Grant No. 41874131), the NSFC-Shandong Joint Fund for Marine Science Research Centers (U1606401) and the Fundamental Research Funds for the Central Universities of China.

Compliance with ethical standards

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

References

Hao X, Zhang H, Wei C, Wang C, Zhang H (2018) Sea trial for fiber-optic hydrophone array used in marine geophysical exploration. J Trop Ocean 37(3):93–98

Hart PE (2007) High-resolution marine seismic reflection profiling: systems and capabilities. J Acoust Soc Am 122(5):P2982

Keith A (2018) Wear, considerations for choosing sensitive element size for needle and fiber-optic hydrophones I: spatiotemporal transfer function and graphical guide. IEEE Trans Ultrason Ferroelectr Freq Control 66:340–347

Luo H, Xiong S, Hu Y et al (2012) A fiber optic hydrophone to be used in towed line array. Appl Acoust 25(2):65–68

Ni M, Li X, Zhang R (2004) Sea tests of an all-optical fiber-optic hydrophone system. Acta Acust 29(6):539–543

Peng C, Zhang X (2018) A dynamic depth estimation method for towed optical fiber hydrophone array. J Acoust Soc Am 143:EL399

Shao M, Qiao XG, Fu HW et al (2011) Applications of fiber optical sensing in seismic exploration. Prog Geophys 26(1):342–348

Wang F, Wang L, Wang L et al (2010) Sea-trial comparison between a fiber-optic hydrophone array and a conventional piezoelectric array. Appl Acoust 29(2):109–114

Yun Z, Luo H, Hu Z et al (2012) Fiber optic hydrophone used for thin line towed array. Acta Opt Sin 32(12):1206004

Zhang R, Ni M (2004) Principle and applications of the fiber optic hydrophone. Physics 33(7):503–507

Acta Geophysica (2019) 67:1413–1418
https://doi.org/10.1007/s11600-019-00346-7

RESEARCH ARTICLE - APPLIED GEOPHYSICS

Feasibility of sparker source in marine seismic exploration: data analysis and processing

Xiangchun Wang[1] · Bin Qi[1] · Qingxian Zhao[2] · Weiwei Wang[2] · Kaiben Yu[3]

Received: 21 June 2019 / Accepted: 19 August 2019 / Published online: 23 August 2019

Abstract
Sparkers are the sound source widely used in marine seismic exploration to provide high-resolution vertical sections. Sparkers are relatively simple, inexpensive, high-frequency sources. In this study, the types of noise occurring in sparker source data were analyzed and attenuated by a processing technique. Frequency–wave number (*f*–*k*) filtering is used for attenuating the linear noise. Predictive deconvolution is used for attenuating the ghost waves and bubble effects. A complete processing workflow was designed for processing the data, and the migration section was obtained. The results show that the sparker source data are capable of achieving vertical sections with very high resolution. It is suggested as a necessary technique for high-accuracy gas hydrate exploration in the South China Sea.

Keywords Sparker · Marine seismic · Deconvolution · Migration · Gas hydrate

Introduction

Ice-like gas hydrate is formed by methane and water molecules with a clathrate structure and is stable under conditions of high pressure and low temperature. It is widely distributed in marine sedimentary layers at depths of several hundred meters beneath the seafloor on most continental slopes (Kong et al. 2012). Its thickness varies greatly in the horizontal direction. As a result, high-resolution seismic exploration is a necessary requirement for gas hydrate identification.

The use of relatively short, high-resolution digital hydrophone streamers with sparker sources provides many operational advantages. Sparkers use an impulsive electrical discharge to produce a sound pulse that, in theory, can produce very high-resolution images of the shallow subsurface because of its higher frequencies and bandwidth.

However, many marine seismic surveys that use sparker sources often apply standard processing flows that fail to take advantage of the broadband characteristics of the sparker source, thus compromising the solution (Kluesner et al. 2018). Maximizing the data resolution is critical in studies of near-surface processes such as Quaternary geology, tectonic geomorphology, substrate fluid flow, and submarine landslide generation (Jones 2013; Liberty et al. 2013; Brothers et al. 2018; Haeussler et al. 2015; Johnson et al. 2017; Hill et al. 2017; Maier et al. 2018; Beeson et al. 2017; Conrad et al. 2018).

In this study, data from sparker sources were analyzed and processed to confirm their feasibility for high-resolution gas hydrate exploration. Wavelet stability was analyzed, together with the types of noise in the data and attenuated them using a processing technique that employs *f*–*k* filtering to attenuate the linear noise. Application of predictive deconvolution is useful for attenuating the ghost waves and bubble effects. A complete processing flow was designed, and the data were processed to obtain stack and migration sections. The results show that sparker source data can produce very high-resolution vertical sections.

Data acquisition

An acquisition test of sparker-sourced seismic exploration was completed for the northern slope of the South China Sea by the Guangzhou Marine Geological Survey Bureau

✉ Xiangchun Wang
wangxc@cugb.edu.cn

[1] School of Geophysics and Information Technology, China University of Geosciences (Beijing), Beijing 100083, China

[2] Guangzhou Marine Geological Survey of China Geological Survey, Guangzhou 510760, China

[3] National Deep Sea Center, Ministry of Natural Resources, Qingdao 266237, China

in 2015. The depth of the sea in the test area is about 600–2000 m. One shot line was deployed on the sea surface, with shot intervals of 25 m and 120 traces per shot. The record length was 5 s; sample interval was 0.25 ms. The source was a 50 kJ sparker array. Receiver interval was 6.25 m. The depth of the sources and streamer was all 5 m. Offset was between 100 m (minimum) and 850 m (maximum).

Data analysis

Direct-arrival wave analysis

The direct-arrival wave quality is an indicator of the quality of the spark source. The same traces of two different sources were selected. Figure 1 shows that the waveforms are almost identical. This confirms that the spark sources are stable; therefore, the deconvolution technique is applicable for processing the data and improves the vertical resolution.

Noise analysis

Noise suppression is a very important step in the processing of marine seismic data. To effectively suppress noise, it is necessary to recognize the noise and its characteristics. Four categories of noise can be clearly recognized in the sparker source seismic data. The first is turbulence noise (Fig. 2a). The energy of turbulence noise is very strong and its main frequency is very low, with amplitude peaks at 3 Hz and 8 Hz. The second kind is random noise (Fig. 2b), evident in all of the shot gather data. The third kind is abnormal

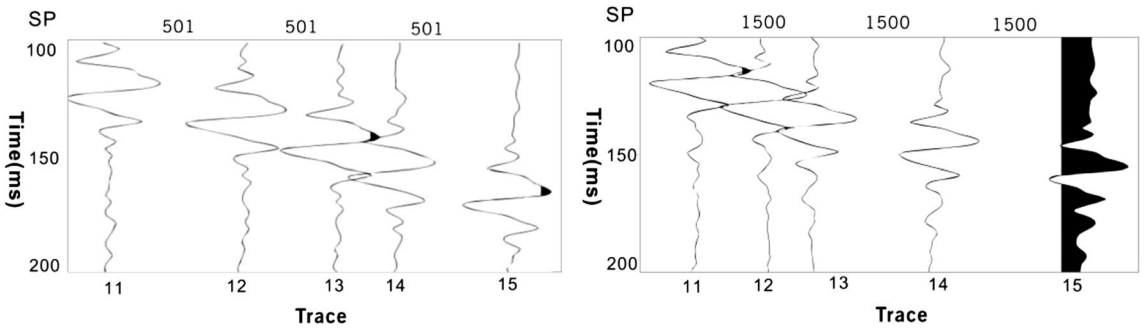


Fig. 1 Direct-arrival waves from two sources are consistent

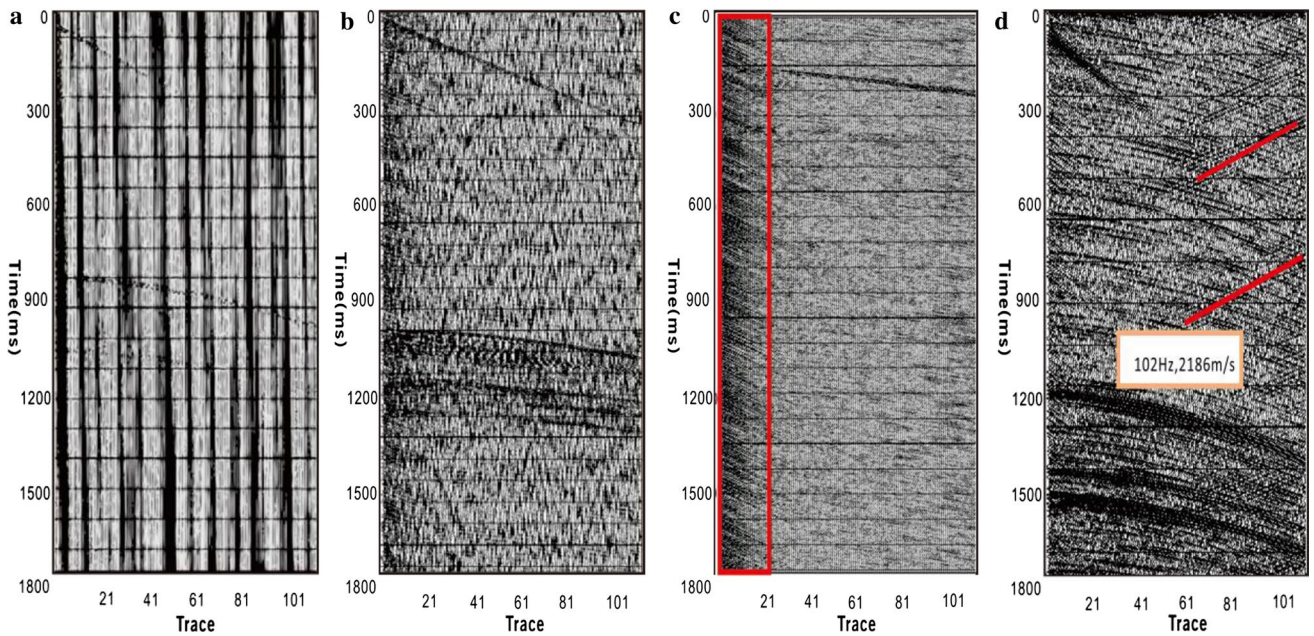


Fig. 2 Noise analysis: **a** turbulence noise; **b** random noise; **c** abnormal amplitude noise; and **d** linear noise

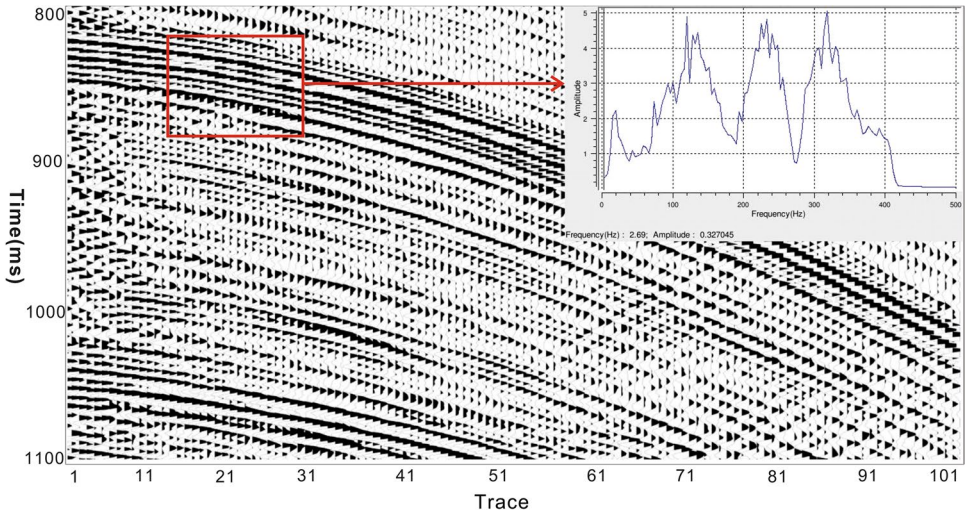


Fig. 3 Ghost waves follow the primary pulse in time. Periodic notch frequency points in the frequency–spectrum graph severely reduce vertical resolution of the image

amplitude noise on the shot side. Figure 2c shows abnormal amplitude noise (which is linear noise) throughout the whole record length. The fourth kind is linear noise (velocity about 2100 m/s and main frequency about 102 Hz) on the receiver side (Fig. 2d).

Ghost wave analysis

The reflections of the layers near the seafloor are very consistent (Fig. 3); however, they are not all true reflections of the layers. Some of the events are ghost waves, due to seismic energy that travels upward from the sparker source and is reflected at the seawater/air interface, which is an impedance boundary with reflection coefficient approximately equal to −1 (Baldock et al. 2013). This type of air–sea reflection ('source ghost') closely follows the primary pulse in time. Periodic notch frequency points occur in the frequency–spectrum figure, which may seriously reduce the vertical resolution.

Data processing

Routine marine seismic data processing was adopted in this study. The complete processing workflow is shown in Fig. 4. The main processing techniques are:

- pre-stack noise attenuation to improve the signal-to-noise ratio (SNR)
- amplitude compensation to improve the energy of deep reflection events
- predictive deconvolution to improve vertical resolution
- seismic-multiples attenuation to eliminate the multiples
- velocity analysis for achieving the velocity model

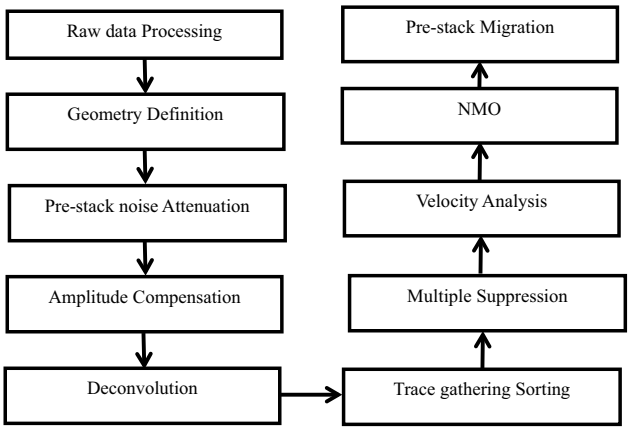


Fig. 4 Seismic data processing workflow. Main processing techniques are pre-stack noise attenuation, amplitude compensation, predictive deconvolution, multiples attenuation, velocity analysis, and normal moveout (NMO)

Finally, the pre-stack migration processing was conducted and the final migration section was obtained.

Noise attenuation

Random noise in sparker source data is more serious than from an air gun source, so attenuation of the random noise is a key step in improving SNR. Figure 5 shows that SNR was clearly improved after processing.

There is an abnormal linear noise in the shot side and the receiver side, which needs to be attenuated. Here, the seismic data are firstly separated in different frequency ranges, then the linear noise is attenuated in the different frequency ranges, and finally the seismic data are integrated. Figures 6 and 7 clearly show improved SNR after processing.

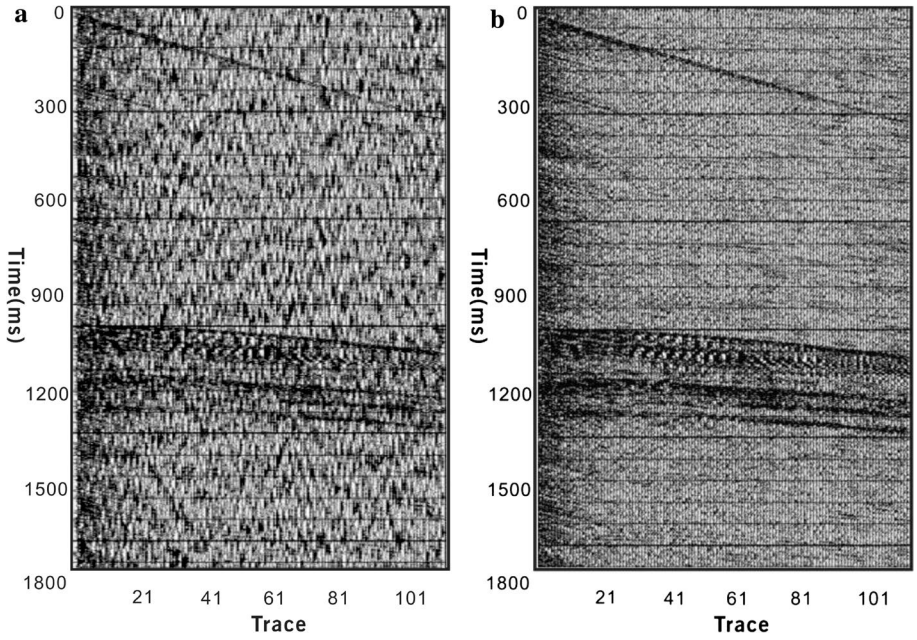


Fig. 5 Random noise. Shot gather **a** before suppressing random noise and **b** after suppressing random noise

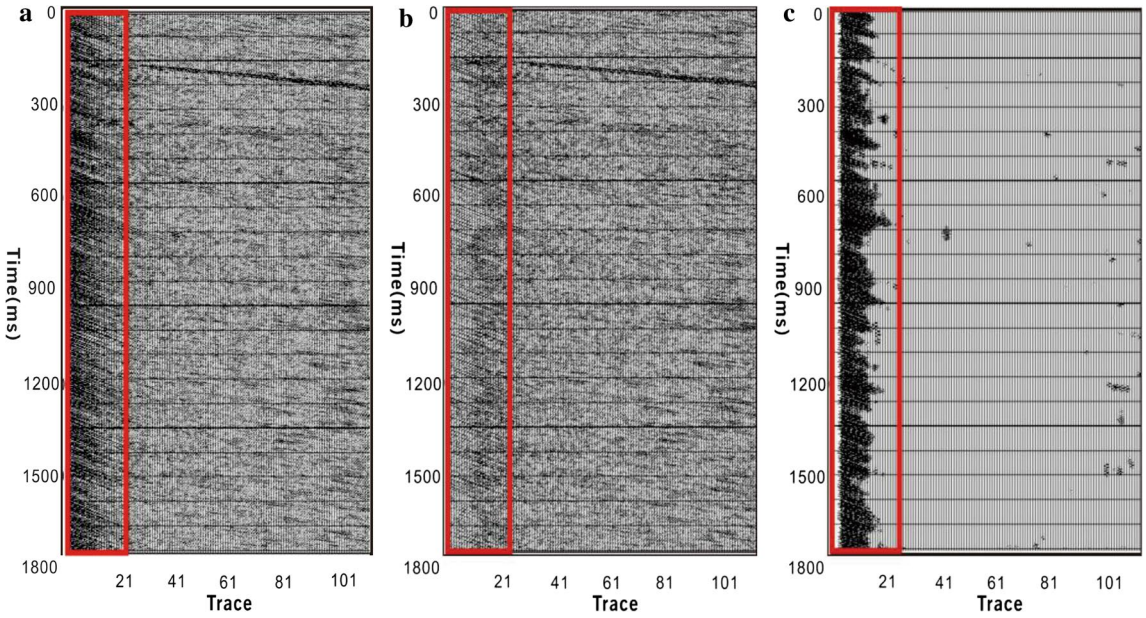


Fig. 6 Suppression of abnormal linear noise. Shot gather **a** before suppression; **b** after suppression; and **c** the suppressed linear noise

Ghost attenuation

In an effort to increase resolution and better isolate the Earth's impulse response, geophysicists commonly use processing techniques such as deconvolution to reduce the energy of multiples and sharpen the source signature on seismic reflection records (Yilmaz 2001; Duchesne et al. 2007; Sheriff 2005; Crocker and Fratantonio 2016; Scheuer and Oldenburg 1988). In the present study, the predictive deconvolution technique was used to attenuate the ghost wave and bubble effects. Figure 8 shows that after processing, the amplitudes at the notch frequency points were clearly improved. The spectrum becomes flat, and the vertical resolution is improved accordingly.

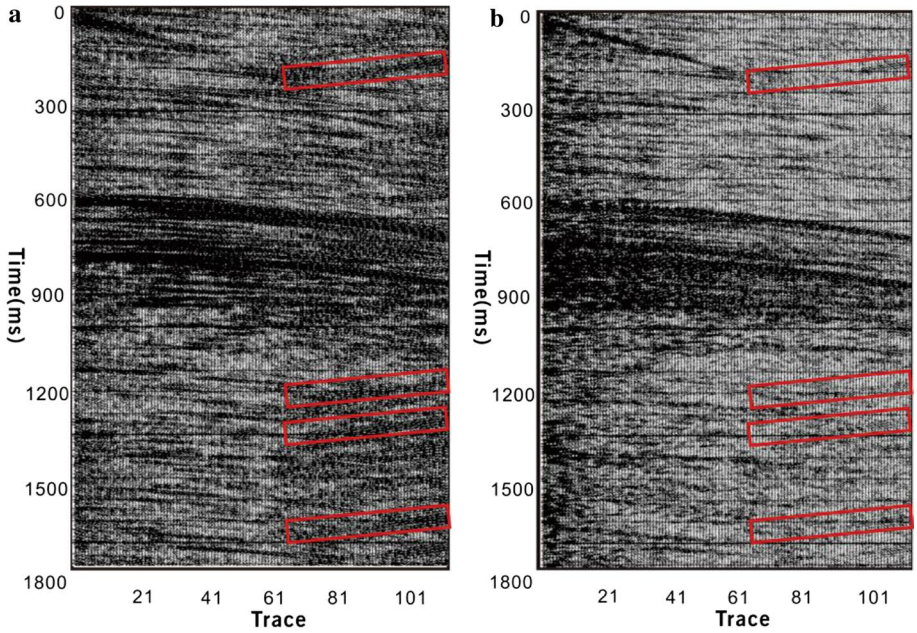


Fig. 7 Attenuation of abnormal linear noise. Shot gather **a** before suppression and **b** after suppression

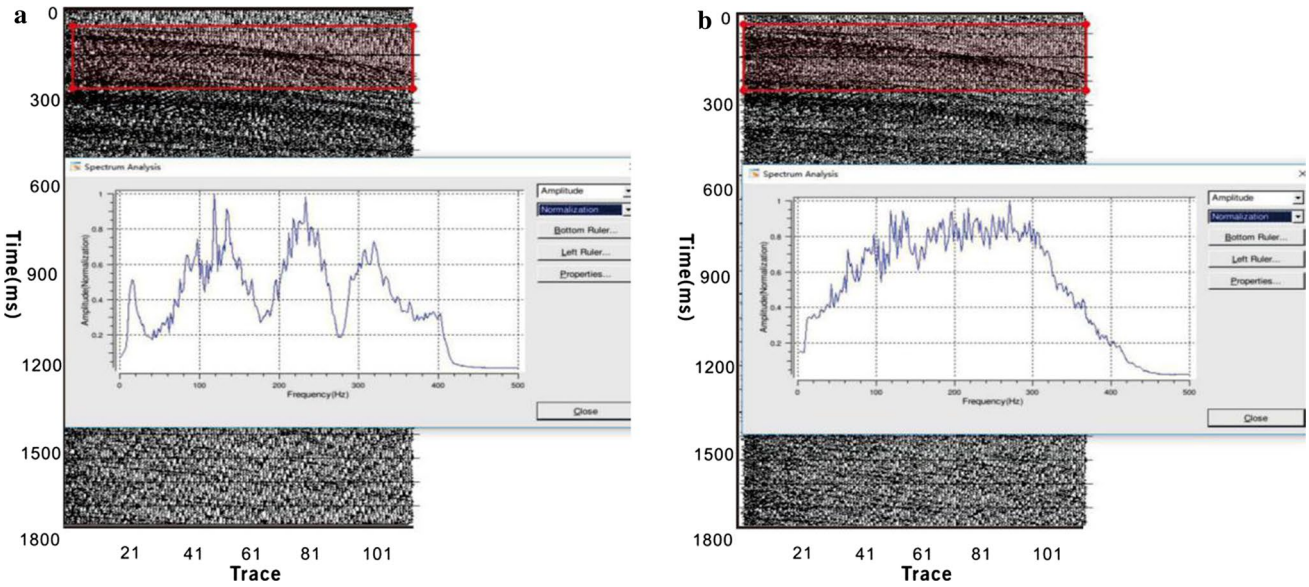


Fig. 8 Attenuation of ghost wave by predictive deconvolution. Shot gather **a** before suppression and **b** after suppression

Migration section

Figure 9 shows the final migration section. The seafloor and the bottom-simulating seismic reflector (BSR) can be clearly seen in the migration section. The reflection blank zone between seafloor interface and the BSR is very clear. The frequency bandwidth is about 10–400 Hz, indicating that the vertical resolution of the migration is very high.

Conclusion

The analysis showed that there are four kinds of noise in the sparker source seismic data: random noise, turbulence noise, linear noise on the receiver side, and abnormal linear noise on the shot side. These were attenuated by processing.

The migration section shows that the data provided very high vertical resolution sections. This is suggested as a very necessary technique for high-accuracy gas hydrate exploration in the South China Sea.

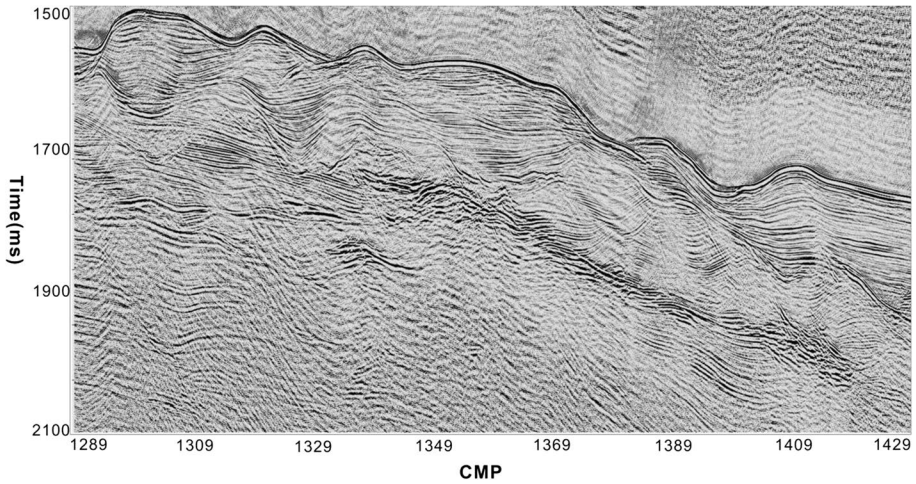


Fig. 9 Migration section showing high vertical resolution

Acknowledgements This research was funded by the National Key Research and Development Program of China (2017YFC0307405), the NSFC-Shandong Joint Fund for Marine Science Research Centers (U1606401), the National Natural Science Foundation of China (No. 41874131), and the Fundamental Research Funds for the Central Universities of China.

Compliance with ethical standards

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

References

Baldock S, Masoomzadeh H, Woodburn A et al (2013) Increasing the bandwidth of marine seismic data. Petroleum Exploration Society of Australia New Resources, Melbourne, pp 55–57

Beeson JW, Johnson SY, Goldfinger C (2017) The transtensional offshore portion of the northern San Andreas fault: fault zone geometry, late Pleistocene to Holocene sediment deposition, shallow deformation patterns, and asymmetric basin growth. Geosphere 13:1173–1206

Brothers DS, Andrews BD, Walton M et al (2018) Slope failure and mass transport processes along the Queen Charlotte Fault, southeastern Alaska, in subaqueous mass movements and their consequences: assessing geohazards, environmental implications. In: Campbell C, Lintern G, Bobrowski P (eds) Geological Society of London Special Volume, Special Publication, p 477. https://doi.org/10.1144/sp477.30

Conrad JE, Brothers DS, Maier KL, Ryan HF, Dartnell P, Sliter RW (2018) Right-lateral fault motion along the slope-basin transition, Gulf of Santa Catalina, southern California. In: Marsaglia KM, Schwalbach R, Behl RJ (eds) From the Mountains to the Abyss: The California Borderland as an Archive of Southern California Geologic Evolution, Special Publication 110: SEPM (Society for Sedimentary Geology), Tulsa, Oklahoma, p 110.07. https://doi.org/10.2110/sepms

Crocker SE, Fratantonio FD (2016) Characteristics of sounds emitted during high-resolution marine geophysical surveys. Naval Undersea Warfare Center Division-Newport, Technical Report 12,203

Duchesne MJ, Bellefleur G, Galbraith M et al (2007) Strategies for wave form processing in sparker data. Mar Geophys Res 28:153–164

Haeussler PJ, Armstrong PA, Liberty LM et al (2015) Focused exhumation along megathrust splay faults in Prince William Sound, Alaska. Quat Sci Rev 113:8–22

Hill JC, Brothers DS, Craig BK et al (2017) Geologic controls on submarine slope failure along the central U.S. Atlantic margin: insights from the Currituck Slide Complex. Mar Geol 385:114–130

Johnson SY, Cochrane GR, Golden NE (2017) The California seafloor and coastal mapping program—providing science and geospatial data for California's State Waters. Ocean Coast Manag 140:88–104

Jones LE (2013) High frequency enhancement of sparker sub bottom profiles with multichannel reflection processing. In: International geophysical conference and exhibition (ASEG-PESA), pp 1–4

Kluesner J, Brothers D, Hart P et al (2018) Practical approaches to maximizing the resolution of sparker seismic reflection data. Mar Geophys Res. https://doi.org/10.1007/s11001-018-9367-2

Kong F, Tao HE, Spence GD (2012) Application of deep-towed multichannel seismic system for gas hydrate on mid-slope of northern Cascadia margin. Sci China Earth Sci 55(5):758–769

Liberty LM, Flinn SP, Haeussler PJ et al (2013) Mega thrust splay faults at the focus of the Prince William Sound asperity, Alaska. J Geophys Res Solid Earth 118:5428–5441

Maier KL, Roland EC, Walton MA et al (2018) The tectonically controlled San Gabriel channel-lobe transition zone, Catalina Basin, southern California borderland. J Sediment Res 88:924–959

Scheuer T, Oldenberg DW (1988) Aspects of time-variant filtering. Geophysics 53:1399–1409

Sheriff RE (2005) Encyclopedic dictionary of applied geophysics. Society of Exploration Geophysicists, Tulsa

Yilmaz O (2001) Seismic data analysis: processing, inversion, and interpretation of seismic data. Society of Exploration Geophysicists, Houston

Acta Geophysica (2019) 67:1419–1433
https://doi.org/10.1007/s11600-019-00327-w

REVIEW ARTICLE - HYDROLOGY

The Pearson Type IV distribution function employed to describe the parametric flow hydrograph

Beata Baziak[1] · **Wiesław Gądek**[1]

Received: 15 March 2019 / Accepted: 8 July 2019 / Published online: 17 July 2019

Abstract
The gamma distribution functions with one shape parameter, employed to describe the parametric hydrograph, proved ineffective for the upper Vistula River and the middle Oder River water regions. It was therefore necessary to find a different function. The Pearson Type IV distribution functions proposed by Strupczewski with one and two shape parameters were analyzed for their applicability based on the data acquired from 60 water gauges, 30 of which were located on the Vistula River and the other 30 were on the Oder River. The shape parameter (parameters) and the time of rising limb were optimized based on the calculated hydrograph widths at 50% and 75% of peak flow (*W50* and *W75*) as well as on the skewness coefficient *s*. The calculated parametric hydrographs were compared with the nonparametric input hydrographs with regard to the closeness of their volumes and the position of their centers of gravity. Both Pearson Type IV distribution functions proved to fit well. However, the function with two shape parameters did not yield the exact solution since the condition of the assumed objective function was met by a very large group of pairs of *m* and *n* shape parameters. It was therefore assumed that the recommended function is the Pearson Type IV distribution with one shape parameter. This function has an additional advantage of having an inflection point located between the *W50* and *W75*, which allows to use the exponential function for the rising or recession limb that better describes either part of the hydrograph.

Keywords Parametric hydrograph · Strupczewski's methods · Gamma distribution function · Archer's method · Pearson Type IV distribution function

Introduction

Both climate change (Hattermann et al. 2013) and the effect of anthropopressure enforce the use of hydrological methods to assess the scale of threats and to predict their occurrence that have been neglected so far. Hydrological methods attempt to describe extreme phenomena. Increasing attention is paid to their definition in a time-variant system, i.e., focusing not only on extreme values but also on the time course of these phenomena, which is related to the determination of the shape of a flood wave.

The knowledge of the theoretical shape of a flood wave and the possibility of its definition using its basic parameters is very much needed and desired in a number of design tasks in the field of water management, hydraulic engineering (Mioduszewski 2014), water and sewage management, spatial management (Zevenbergen et al. 2011) as well as forest management. In contrast to the commonly used design flows, the hypothetical waves expand the range of usable data, e.g., by the volume of a flood wave with a given exceedance probability and the variation of the flow rate for the rising and falling limbs. Therefore, the design can take into account the flow in the form of a hydrograph with a given exceedance probability (Ciepielowski 1987, 2001).

The hypothetical hydrograph is understood as such theoretical hydrograph that demonstrates representative flood wave form, which may occur under specific conditions at a selected location, for a given maximum (design) value (Gądek and Środula 2014). It is being increasingly utilized in the widely understood flood risk assessment (Apel et al. 2006; Vrijling et al. 1998; Zeleňáková et al. 2017) and in

✉ Beata Baziak
beata.baziak@iigw.pk.edu.pl

Wiesław Gądek
wieslaw.gadek@iigw.pk.edu.pl

[1] Department of Water Engineering and Water Management, Cracow University of Technology, Warszawska 24, 31-155 Kraków, Poland

the estimation of loss of lives and property (Ernst et al. 2010; Jonkman and Vrijling 2008).

These hydrographs are presented in an analytical form, using a variety of functions, or in a synthetic form, which uses two-dimensional statistical analysis (De Michele et al. 2005; Serinaldi and Grimaldi 2011). In some countries, analytical hydrographs are called parametric flow hydrographs. Their main advantage is that they can be determined at any cross section of the river, with the influence of climate change on their course taken into consideration (O'Connor et al. 2014; Bayliss 1999; Mills et al. 2014). In order to describe the course of the parametric design hydrograph, it is necessary to use the appropriate mathematical function. The most common one is the gamma distribution function, which was proposed by Nash in 1957 (Nash 1957).

The function gamma describes the rising limb very well with large flow heights (above 50% of the maximum flow Qmax), but in the lower part of the recession (falling) limb, large discrepancies occur. For this reason, in Ireland, the exponential function known as the UPO-ERR-Gamma (unit-peak-at-origin gamma curve coupled with an exponential replacement recession curve) has been introduced for the recession limb (O'Connor et al. 2014).

It is more reasonable to use homogeneous functions instead of spline functions for the needs of analytic hydrographs. The attempt to use the Hayashi distribution (Hayashi et al. 1986; Aziz et al. 2006), the negative binomial distribution, the inverse Gaussian distribution and the gamma distribution with algebraic replacement recession curve was considered unconvincing (O'Connor et al. 2014). The authors of this manuscript also verified the applicability of the three-parameter Pearson Type III distribution function with two shape parameters (Gądek et al. 2017b). Although the proposed method yielded positive results, it could not be recommended due to the very large number of solutions for the parameters t_p, m and n (where m and n represent the shape parameters). The function describing a hydrograph must not only be adapted to the time course of flow variations but also yield the unique solution. Such rigorous assumption allows to determine the parametric design hydrograph in any section of the river, which has only been possible so far using hydrological models (Ozga-Zielińska et al. 2002; Wałęga 2013; Pietrusiewicz et al. 2014), being a rather cumbersome process and not always yielding unambiguous results, mainly due to the lack of procedures to determine the course of a hyetograph or a possibility to assess the moisture conditions in the catchment.

In the design hydrology, parametric hydrographs may be determined in any cross section of the river. This is in line with the expectations regarding this type of solutions and the idea originating in the 1930s associated with the isochrones theory developed by Dubelir, Boldakov and Čerkašin. This theory is based on the genetic flood wave equation which is given by:

$$Q_t = \int_{\tau=0}^{\tau=t} h_{t-\tau} b_\tau v_\tau \mathrm{d}\tau \tag{1}$$

where Q_t is the outflow rate from the catchment at time t, $h_{t-\tau}$ the thickness of water layer discharged by the catchment in the time unit $t-\tau$, b_τ the average width of the partial runoff area, v_τ the runoff velocity, t the time of discharge from the catchment, and τ the time needed for water to reach the cross section.

This method was used until the mid-1960s and resulted in the creation of hydrographs presented in the form of a triangle or trapezium. Its advantage was the ability to determine the hydrograph in a selected cross section, which was not possible later as a result of the use of the so-called hypothetical hydrographs determined by the Reitz and Kreps method (Reitz and Kreps 1945), the Warsaw University of Technology method (Gądek et al. 2017b), the Hydroprojekt method (Gądek and Środula 2014) or the Krakow method (Gądek and Tokarczyk 2015).

The modified Pearson Type III distribution function, consistent with the nonparametric hydrograph with one shape parameter m, is given by (Gądek et al. 2017b):

$$q_t = \left(1 + \frac{t}{t_p}\right)^m \exp\left(-m\frac{t}{t_p}\right) 100\% \tag{2}$$

and with two shape parameters:

$$q_t = \left(1 + \frac{t}{t_\mathrm{p}}\right)^m \exp\left\{\frac{m}{n}\left[1 - \left(1 + \frac{t}{t_\mathrm{p}}\right)^n\right]\right\} 100\% \tag{3}$$

where q_t is the percentage of peak flow at time t [%], t_p the time to peak [h], t the time from the beginning of rising limb [h], and m, n the shape parameters [−].

Similar solutions with one shape parameter were presented in the USA (McEnroe 1992) and in Ireland under the name UPO gamma (unit-peak-at-origin gamma) (O'Connor et al. 2014).

The authors of this research paper propose to use the Pearson Type IV distribution with one shape parameter and two shape parameters for the description of the analytical hydrograph. This distribution was tested for the analytical description of hydrographs in the 1960s, which was then considered to be less accurate than the Pearson Type III distribution with two shape parameters (Strupczewski 1964; Strupczewski et al. 2013). Currently, this distribution is practically disused. The authors modified the Pearson Type IV distribution functions proposed by Strupczewski to match the analytical hydrograph to the nonparametric hydrograph determined by the Archer's

method. The objective of this paper is to prove that the modified Pearson Type IV distribution functions are well suited for describing a parametric hydrograph based on three parameters: hydrograph widths at 50% and 75% of peak flow (*W50* and *W75*) and the skewness coefficient *s*. Innovative research has been carried out for two water regions of Poland: the upper Vistula River and the middle Oder River. Thirty gauged cross sections were included in the calculations for each of these regions. To determine the parameters *W50*, *W75* and *s,* nonparametric hydrographs were developed in each of these cross sections. The developed method ought to have a universal character; it should enable determining parametric hydrographs in any cross section of any river. In order to prove the universality of the proposed distribution functions, a large number of catchments with diverse hydrological regime were adopted for the calculations.

Materials and methods

Study area

The research studies were carried out based on the recorded hydrographs in 60 measurement cross sections, located in the upper Vistula River and the middle the Oder River water regions (Fig. 1). The selected catchments represented the areas of various types of hydrograph formation. The selection was made so that they represented different types of geographic areas: mountain, highland as well as lowland catchments. Eight unimodal hydrographs with the highest flow values Qmax selected from the period 1960–2014 were adopted. Table 1 illustrates the gauged stations systematized from 1 to 30 for the Vistula River, and from 31 to 60 for the Oder River. Some gauged stations are located downstream of the water reservoirs, but the distances from the reservoirs are so large that no influence of the reservoirs on the hydrographs in gauged stations could have been assumed.

Methods

Parametric flow hydrographs can be determined in any cross section of the river regardless of the size of the catchment. It is made possible thanks to the Archer's method of determining nonparametric hydrograph (i.e., the median of recorded hydrographs). The nonparametric hydrograph determined by the Archer's method is used only to determine the value of hydrograph width at 50% (*W50*) and 75% (*W75*) of peak flow and the skewness coefficient *s* (Archer et al. 2000). The Archer's method uses *W50* and *W75* similarly to the Snyder method (1938) where with similar parameters characterizing the Synthetic Unit Hydrograph (Snyder 1938; Challa 1997).

According to this method, the nonparametric hydrograph has an independent rising limb and an independent recession limb (Fig. 2). The flows are presented as

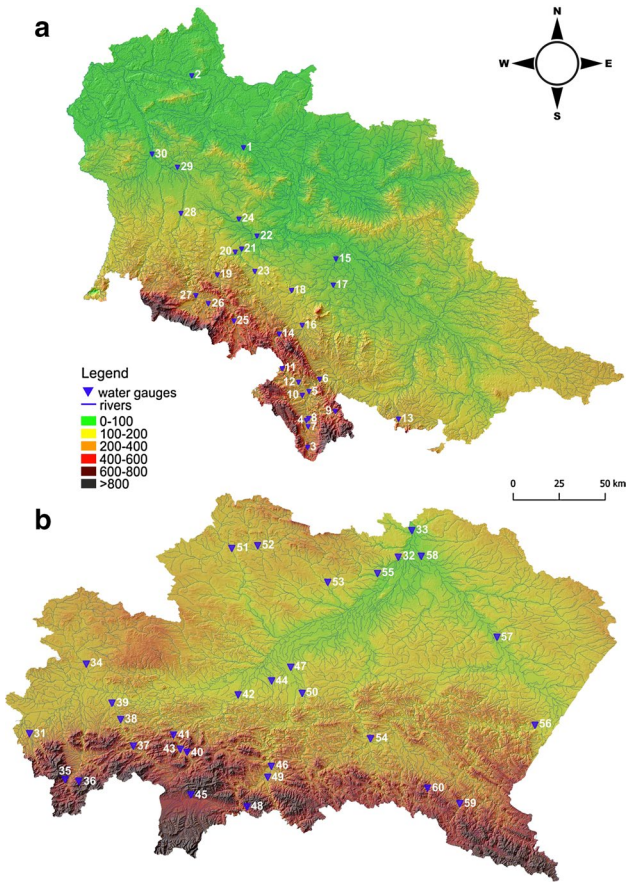


Fig. 1 Location of water gauge in: **a** the middle Oder River water region and **b** the upper Vistula River water region (see Table 1)

percentages of peak flow. The horizontal axis indicates the duration of percent flow exceeding the given value. The time for the rising limb of the hydrograph is expressed in negative values, and for the recession limb in positive values. At the time $t = 0$ there is a maximum percentage of peak value $q = 100\%$. The time t of the individual percent flows is the median of the durations of percent flow of the recorded hydrographs, separately for the rising limb and separately for the recession limb (O'Connor et al. 2014; Gądek et al. 2017a). Such a nonparametric hydrograph is determined based on the recorded hydrographs. The applied methods of analytical hydrographs determination based on nonparametric hydrographs assume various numbers of unimodal flow hydrographs. The Warsaw University of Technology method uses six unimodal flow hydrographs (Gądek et al. 2016), the Hydroprojekt method—one (Gądek and Środula 2014), and the Krakow method—eight (Gądek and Tokarczyk 2015; Gądek et al. 2016). The authors applied the maximum number of hydrographs from used methods, i.e., eight unimodal flow hydrographs.

In 1964, Strupczewski proposed to use the Pearson Type III distribution function with one shape parameter *m* and

Table 1 Water gauges in the hydrologic order for the Oder River and Vistula River water region

Nos.	River	Water gauge	A (km^2)	Nos.	River	Water gauge	A (km^2)
1	Odra	Głogów	36,403	31	Wisła	Skoczów	296
2	Odra	Cigacice	39,900	32	Wisła	Sandomierz	31,847
3	Nysa Kłodzka	Międzylesie	49.7	33	Wisła	Zawichost	50,732
4	Nysa Kłodzka	Bystrzyca Kłodzka	260	34	Przemsza	Jeleń	2006
5	Nysa Kłodzka	Kłodzko	1084	35	Bystra	Kamesznica	48.2
6	Nysa Kłodzka	Bardo	1744	36	Żabniczanka	Żabnica	22.8
7	Wilczka	Wilkanów	35.1	37	Skawa	Sucha Beskidzka	468
8	Bystrzyca	Bystrzyca Kłodzka	64	38	Skawa	Wadowice	835
9	Biała Lądecka	Lądek Zdrój	166	39	Wieprzówka	Rudze	154
10	Bystrzyca Dusznicka	Szalejów Dolny	175	40	Raba	Kasinka Mała	353
11	Ścinawka	Tłumaczów	256	41	Raba	Stróża	644
12	Ścinawka	Gorzuchów	511	42	Raba	Proszówki	1 470
13	Biała Głuchołaska	Głuchołazy	283	43	Lubieńka	Lubień	46.9
14	Bystrzyca	Jugowice	122	44	Uszwica	Borzęcin	265
15	Bystrzyca	Jarnołtów	1721	45	Dunajec	Nowy Targ-Kowaniec	681
16	Piława	Mościsko	292	46	Dunajec	Nowy Sącz	4341
17	Czarna Woda	Gniechowice	251	47	Dunajec	Żabno	6735
18	Strzegomka	Łażany	362.3	48	Grajcarek	Szczawnica	73.6
19	Kaczawa	Świerzawa	133.7	49	Poprad	StarySącz	2071
20	Kaczawa	Rzymówka	313.7	50	Biała Tarnowska	Koszyce Wielkie	957
21	Kaczawa	Dunino	774	51	Nida	Brzegi	3359
22	Kaczawa	Piątnica	1807	52	Czarna Nida	Morawica	755
23	Nysa Szalona	Jawor	298	53	Czarna	Staszów	571
24	Czarna Woda	Bukowna	430.5	54	Jasiołka	Jasło	164
25	Bóbr	Kamienna Góra	190	55	Koprzywianka	Koprzywnica	498
26	Bóbr	Wojanów	535.2	56	San	Przemyśl	3686
27	Bóbr	Jelenia Góra	1047	57	San	Rzuchów	12,180
28	Bóbr	Dąbrowa Bolesławiecka	1713	58	San	Radomyśl	16,824
29	Bóbr	Szprotawa	2879	59	Osława	Szczawne	302
30	Bóbr	Żagań	4255	60	Wisłok	Puławy	131

with two shape parameters m and n as well as Type IV with one shape parameter to describe the parametric hydrographs (Strupczewski 1964; Ciepielowski 1987, 2001). The solutions proposed by Strupczewski concerned the methods based on the traditional presentation of nonparametric and parametric hydrographs. The authors of this manuscript adapted the function notation to the description consistent with the properties of the nonparametric hydrograph (median of the hydrographs), developed using the Archer's method. A parametric hydrograph is created from the parameters *W50*, *W75* and *s* determined of the nonparametric hydrograph developed using the Archer's method (see Fig. 3).

The parametric hydrograph is developed on the basis of the Archer's nonparametric hydrograph. For analytical description, two versions: the first with one shape parameter and the second with two shape parameters, of the Strupczewski Pearson Type IV distribution function were adopted. The first function is defined as follows

$$q_t = \left(1 + \frac{t_\mathrm{p}}{t}\right)^m \exp\left(-m\frac{t_\mathrm{p}}{t}\right)100\%. \tag{4}$$

The second function is given by:

$$q_t = \left(1 + \frac{t_\mathrm{p}}{t}\right)^m \exp\left\{\frac{m}{n}\left[1 - \left(1 + \frac{t_\mathrm{p}}{t}\right)^n\right]\right\}100\% \tag{5}$$

The authors modified Eqs. (4) and (5):

$$q_t = \left(\frac{t_\mathrm{p}}{t + t_\mathrm{p}}\right)^m \exp\left[m\left(1 - \frac{t_\mathrm{p}}{t + t_\mathrm{p}}\right)\right]100\% \quad \text{for } t > -t_\mathrm{p} \tag{6}$$

$$q_t = \left(\frac{t_\mathrm{p}}{t + t_\mathrm{p}}\right)^m \exp\left[\frac{m}{n}\left(1 - \frac{t_\mathrm{p}}{t + t_\mathrm{p}}\right)^n\right]100\% \quad \text{for } t > -t_\mathrm{p} \tag{7}$$

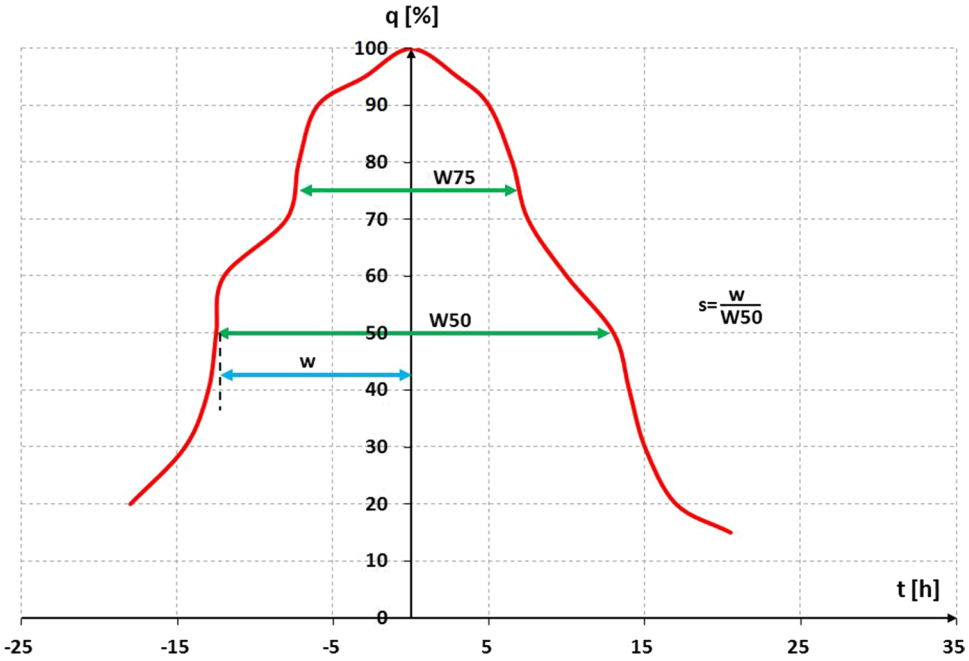


Fig. 2 Exemplary nonparametric hydrograph according to Archer

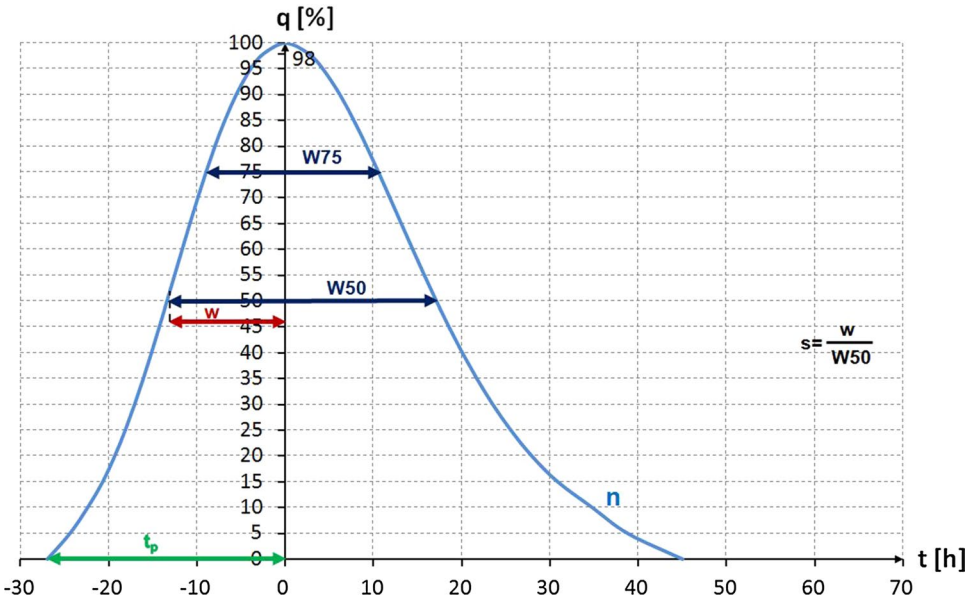


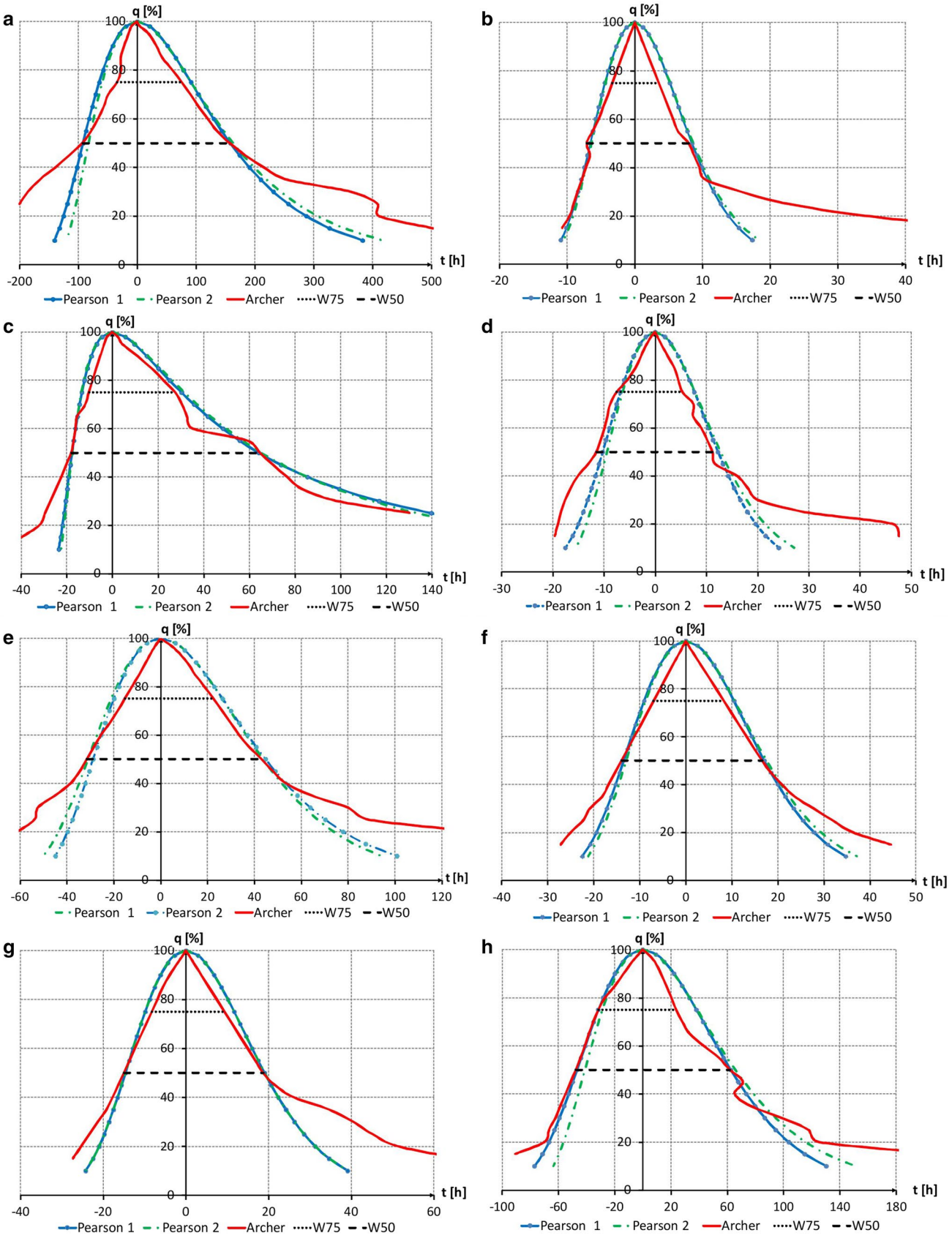
Fig. 3 Exemplary parametric design hydrograph

The optimization of the shape parameters and the time to peak t_p in all formulas was carried out based on the values *W50* and *W75* of the Archer hydrograph and the skewness coefficient *s*, determined for the hydrograph width *W50* (see Fig. 2). It was also assumed that the shape parameters were positive values to enable application of empirical formulas. The descriptors and the skewness coefficient *s* were calculated based on the median of hydrographs for 30-year data sequences for both catchments.

The smallest deviation of the values calculated from the given values of hydrograph width at 50% and 75% of peak flow was adopted as the selection criterion (the objective function) in accordance with the following dependence:

$$S = (W75 - \hat{a})^2 + \left(b - \hat{b}\right)^2 + (c - \hat{c})^2 = \min \tag{8}$$

where *W75* is the hydrograph width at 75% of peak flow determined by the nonparametric hydrograph [h], $\hat{a}$ the hydrograph width at 75% of peak flow *W75* calculated from one of the formulas (6) and (7) [h], *b* the duration of the percent flow exceeding 50% for the rising limb of the nonparametric hydrograph, $b = s \cdot W50$ [h], $\hat{b}$ the duration of


a
b
c
d
e
f
g
h
q [%]
t [h]
Pearson 1
Pearson 2
Archer
W75
W50

◂**Fig. 4** Parametric hydrographs calculated using the Pearson Type IV distribution with one shape parameter (Pearson 1) and two shape parameters (Pearson 2), and nonparametric hydrographs determined by the Archer's method for the following cross sections: **a** Odra–Cigacice (2), **b** Nysa Kłodzka–Kłodzko (5), **c** Kaczawa–Piątnica (22), **d** Bóbr–Kamienna Góra (25), **e** Wisła–Sandomierz (32), **f** Lubieńka–Lubień (43), **g** Koprzywianka–Koprzywnica (55), **h** San–Radomyśl (58)

the percent flow exceeding 50% calculated from one of the formulas (6) and (7) [h], c the duration of the percent flow exceeding 50%, for the recession limb of the nonparametric hydrograph [h], and $\hat{c}$ the duration of the percent flow exceeding 50% calculated from one of the formulas (6) and (7) [h].

Results

The calculations consisted of:

1. Determination of Archer's nonparametric hydrographs for 60 water gauges.
2. Determination of parameters based on the Archer's nonparametric hydrographs: hydrograph width at 50% of peak flow (*W50*), hydrograph width at 75% of peak flow (*W75*) and skewness coefficient *s*.
3. Definition of hydrograph shape parameters (m and n) and the rising time t_p for each water gauge cross section according to the selection criterion (Eq. 8).
4. Determination of the Pearson Type IV parametric hydrographs with one and two shape parameters for the calculated parameters: m, n and t_p.
5. Determination of the *W50*, *W75* and *s* parameters of the Pearson Type IV parametric hydrographs.

Specific steps of calculation were adopted for the calculated shape parameters m and n (0.01) and for the time of rising limb t_p (1 h).

The analytical hydrographs calculated using the Pearson Type IV function with one and two shape parameters exhibit similarity.

Figure 4 shows the values of *W50*, *W75* and *s* of selected Archer's hydrographs calculated using the Pearson Type IV distribution function with one and two shape parameters. Table 2 shows the hydrograph parameters for all 60 water gauges.

Figure 4 confirms that parametric hydrographs (Pearson 1 and Pearson 2) deviate from nonparametric hydrographs determined by the Archer's method. Much better fit occurs in the upper parts of the hydrographs (above *W50*). The fit in the lower parts is much worse which can be expected because of the assumption that the hydrographs are adjusted based on the *W50* and *W75* parameter values.

Analysis and discussion

Several types of quality measures for matching nonparametric and parametric hydrographs were adopted for the analysis. Relative error (RE) and mean relative error (MRE) are criteria recommended in Technical Research Report Volume III Hydrograph Analysis (O'Connor et al. 2014) to assess the compliance of the parametric and nonparametric hydrographs (Fig. 5).

Relative error of hydrograph width was calculated from the following formula:

$$\mathrm{RE}_p = \frac{\left|W_p - \widehat{W}_p\right|}{W_p} \tag{9}$$

where RE_p is the relative error of hydrograph width W_p, $p=50\%$, $p=75\%$ [−], W_p the hydrograph width at $p=50\%$, $p=75\%$ determined from nonparametric design hydrograph [h], and $\widehat{W}_p$ the hydrograph width at $p=50\%$, $p=75\%$ determined from parametric hydrograph for specific formulas which were used (gamma and Strupczewski) [h].

To analyze the calculated values of relative errors of hydrograph width W_p, the following quality assessment measures for *W50* and *W75* were adopted (Table 3).

More stringent criteria were adopted for the *W50* due to the objective function used in the optimization process. The best possible adjustment of the parametric hydrograph to nonparametric for this value was the main assumption of the objective function.

With the adopted criteria, the match quality of the *W50* value of the parametric hydrograph to nonparametric is very good (see Fig. 6a), while for the *W75* value is good (see Fig. 6b), which confirms the correctness of the objective function adopted in the study (Fig. 7).

Mean relative error (Elshorbagy et al. 2000) was calculated for the p percent flow, $p=75\%$ and $p=50\%$, using the following definition:

$$\mathrm{MRE}_p = \frac{1}{N_p}\sum_{i=1}^{N_p}\mathrm{RE}_i \tag{10}$$

where MRE_p is the mean relative error for the p percent flow $p=75\%$ and $p=50\%$, N_p the number of percent flows exceeding p percent flow, 6 for $p=75\%$ and 11 for $p=50\%$, RE_i the relative error of percent flows, $p_1=98$, $p_2=95$, $p_3=90$, $p_4=85$, $p_5=80$, $p_6=75,\ldots,p_{11}=50$ (see Fig. 3) [−], and i the percent flow number.

To analyze the calculated values of mean relative errors for the p percent flow, the following quality assessment measures for *W50* and *W75* were adopted (Table 4).

Figure 8 shows that the mean relative error criterion for the p percent flow for evaluating the fit of the parametric

Table 2 Values of parameters *W50*, *W75* and *s* of the Archer's nonparametric hydrographs and the Pearson Type IV parametric hydrographs with one shape parameter (Pearson 1) and two shape parameters (Pearson 2)

Water gauge nos.	Archer			Pearson 1			Pearson 2		
	W75	*W50*	*s*	*W75*	*W50*	*s*	*W75*	*W50*	*s*
1	107.4	268.9	0.439	171.9	268.8	0.437	163.3	260.3	0.379
2	109.8	250.3	0.371	155.8	250.2	0.368	151.0	245.7	0.336
3	3.9	9.1	0.453	5.7	9.1	0.395	5.7	9.1	0.395
4	3.3	11.0	0.552	7.0	10.9	0.456	7.0	10.9	0.455
5	6.9	15.1	0.473	9.6	15.1	0.437	9.5	15.0	0.424
6	10.7	26.3	0.417	16.6	26.3	0.402	16.7	26.3	0.401
7	7.3	14.7	0.440	9.3	14.7	0.406	9.3	14.7	0.406
8	11.9	24.3	0.185	12.2	25.0	0.170	11.4	24.4	0.165
9	11.4	25.7	0.348	15.6	25.8	0.329	15.7	25.7	0.329
10	6.3	10.1	0.371	6.0	10.1	0.319	6.1	10.1	0.322
11	11.5	23.4	0.522	15.0	23.3	0.456	14.8	23.1	0.459
12	11.0	24.5	0.392	15.2	24.5	0.373	15.3	24.4	0.371
13	6.4	15.9	0.291	8.9	15.7	0.254	9.1	15.9	0.262
14	8.7	18.2	0.531	11.5	18.0	0.456	11.5	18.0	0.450
15	36.3	77.5	0.241	42.4	77.5	0.233	43.4	77.4	0.234
16	34.5	68.1	0.326	40.9	68.0	0.319	41.3	68.1	0.319
17	39.9	75.3	0.386	47.2	75.2	0.383	47.3	75.2	0.383
18	38.0	57.8	0.282	33.3	57.8	0.272	33.7	57.7	0.272
19	8.6	17.9	0.377	11.0	17.9	0.350	11.0	17.9	0.349
20	15.5	28.0	0.305	16.4	28.0	0.289	16.6	28.0	0.286
21	14.2	32.8	0.316	19.6	32.8	0.307	19.7	32.8	0.305
22	37.9	82.7	0.218	43.7	81.9	0.216	45.1	82.8	0.213
23	13.5	25.9	0.429	16.4	25.8	0.410	16.4	25.9	0.409
24	100.2	167.1	0.202	72.7	158.5	0.146	89.0	167.1	0.198
25	12.9	22.7	0.505	14.4	22.5	0.456	14.1	22.2	0.424
26	15.7	38.2	0.398	24.0	38.2	0.386	24.1	38.2	0.387
27	23.4	44.6	0.443	28.5	44.6	0.432	28.5	44.6	0.432
28	27.0	78.4	0.254	43.9	78.6	0.247	44.3	78.2	0.242
29	35.3	98.4	0.194	44.9	94.5	0.159	50.6	98.0	0.184
30	45.2	114.8	0.247	63.4	113.7	0.244	65.1	114.5	0.245
31	12.8	24.6	0.386	15.3	24.6	0.369	15.5	24.6	0.390
32	52.9	95.1	0.308	56.7	95.4	0.304	56.0	90.6	0.332
33	48.4	143.8	0.294	83.7	143.2	0.286	85.3	144.0	0.286
34	62.2	131.7	0.398	83.1	131.8	0.396	80.6	129.2	0.365
35	22.2	44.8	0.469	28.7	44.8	0.456	28.4	44.4	0.432
36	10.2	16.0	0.377	10.9	17.5	0.371	10.9	17.5	0.371
37	13.2	26.5	0.490	19.0	29.8	0.436	16.8	26.2	0.448
38	11.2	19.4	0.388	12.0	19.4	0.364	12.0	19.3	0.362
39	10.1	22.3	0.427	14.1	22.3	0.408	14.0	22.0	0.410
40	10.2	20.5	0.432	13.0	20.5	0.408	13.0	20.5	0.407
41	11.2	23.8	0.498	15.2	23.6	0.456	15.1	23.6	0.450
42	21.6	44.6	0.510	28.4	44.2	0.456	28.1	44.0	0.443
43	15.3	30.5	0.457	19.5	30.5	0.440	19.4	30.4	0.424
44	17.4	34.2	0.423	21.6	34.2	0.405	21.3	33.7	0.407
45	11.1	23.3	0.427	14.8	23.4	0.406	14.8	23.3	0.405
46	18.9	40.1	0.417	25.4	40.1	0.404	25.4	40.1	0.404
47	27.9	57.2	0.576	36.1	56.2	0.456	35.8	55.8	0.452
48	13.3	30.1	0.479	19.3	30.0	0.456	19.1	29.8	0.445
49	27.9	54.8	0.512	34.9	54.3	0.456	34.3	53.7	0.434

Table 2 (continued)

Water gauge nos.	Archer			Pearson 1			Pearson 2		
	W75	*W50*	*s*	*W75*	*W50*	*s*	*W75*	*W50*	*s*
50	9.3	18.6	0.472	11.9	18.6	0.443	11.9	18.6	0.443
51	32.3	69.2	0.303	40.9	69.5	0.293	41.2	69.3	0.293
52	30.2	54.2	0.337	34.5	53.8	0.456	33.9	53.1	0.434
53	21.2	46.7	0.470	30.0	46.7	0.456	30.0	46.7	0.456
54	15.9	31.9	0.515	20.4	31.8	0.456	20.2	31.5	0.449
55	17.6	33.7	0.446	21.5	33.7	0.435	21.5	33.7	0.435
56	34.2	58.3	0.623	36.4	56.7	0.456	36.6	56.9	0.462
57	53.7	94.2	0.571	59.3	92.5	0.456	59.0	92.0	0.451
58	55.2	109.4	0.433	69.7	109.3	0.428	67.2	106.9	0.384
59	14.4	29.4	0.455	18.6	29.4	0.411	18.6	29.4	0.410
60	10.4	24.3	0.466	15.6	24.3	0.445	15.4	24.2	0.424

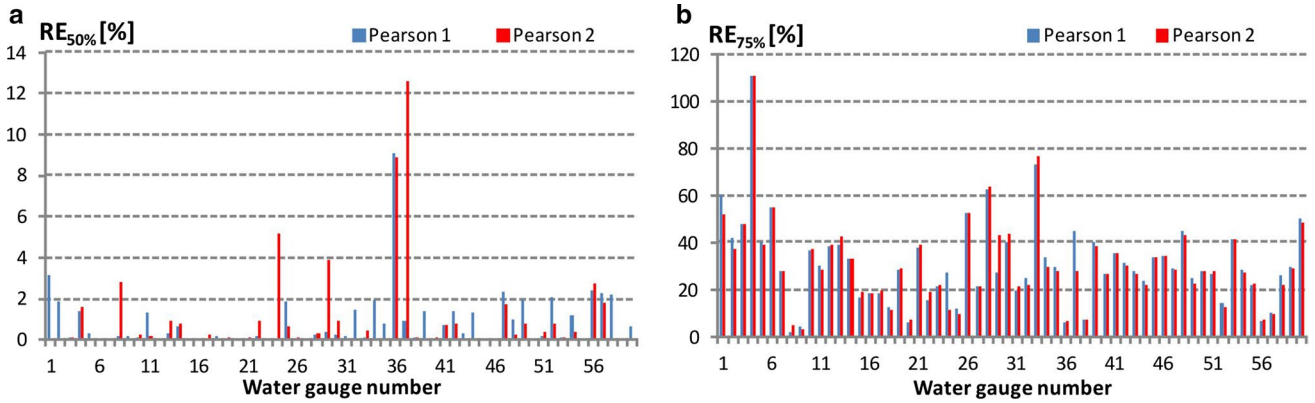


Fig. 5 Relative errors of hydrograph width W_p: **a** for $p=50\%$; **b** for $p=75\%$

Table 3 Quality measures for relative errors of hydrograph width W_p

Quality	*W50*	*W75*
Very good	< 1%	< 20%
Good	< 1%, 2%)	< 20%, 40%)
Weak	< 2%, 4%)	< 40%, 60%)
Very weak	≥ 4%	≥ 60%

hydrograph to nonparametric one is weak. The visual evaluation of the hydrographs shown in Fig. 4 suggests much smaller matching errors.

The REp and MREp measures do not answer unambiguously as to whether the functions used should be recommended for the Vistula or the Oder water regions, or not. A similar observation was reported by Chai and Draxler (2014).

Therefore, two other measures were assumed to assess the similarity of the parametric and nonparametric hydrographs: the volume of hydrograph, V, and the center of gravity time coordinate, r_p.

The volume of hydrograph was determined above the p percent flow, $p=50\%$ and $p=75\%$, using the following definition (see Fig. 9):

$$V_p = \sum_{i=1}^{N_p} V_{p,i} \tag{11}$$

where V_p is the volume of hydrograph above the p percent flow, $p=50\%$, $p=75\%$, N_p the number of percent flows exceeding p percent flow: 6 for $p=75\%$ and 11 for $p=50\%$, and $V_{p,i}$ the partial volume of the hydrograph between successive p percent flows.

The center of gravity time coordinate was determined for the hydrograph part above the p percent flow, $p=50\%$ and $p=75\%$ (see Fig. 9).

$$r_p = \frac{\sum_i^{N_p} V_{p,i} l_i}{\sum_i^{N_p} V_{p,i}} \tag{12}$$

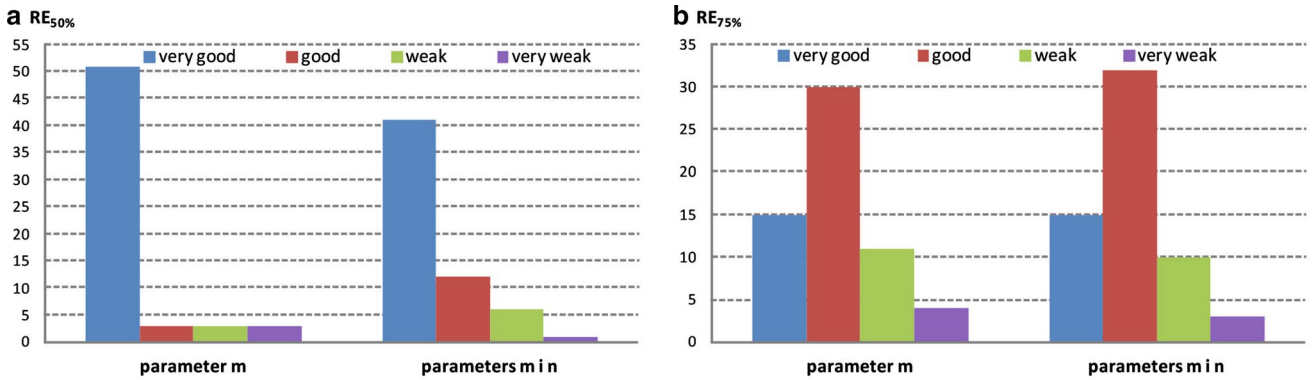


Fig. 6 Quality measures for relative errors of hydrograph width W_p: **a** $p=50\%$; **b** $=75\%$

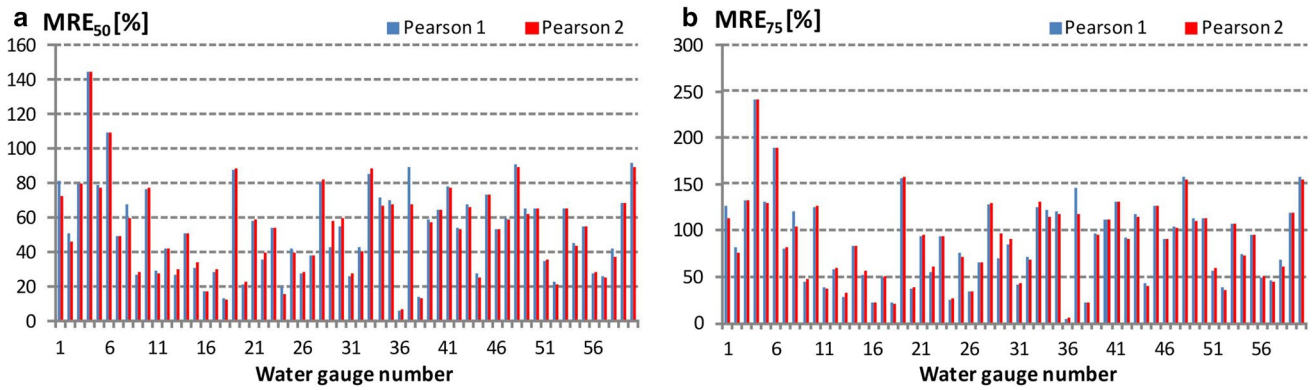


Fig. 7 Mean relative errors for the p percent flow: **a** $p=50\%$; **b** $p=75\%$

Table 4 Quality measures for mean relative errors for the p percent flow

Quality	*W50*	*W75*
Very good	< 10%	< 10%
Good	< 10, 50%)	< 19%, 50%)
Weak	< 50%, 100%)	< 50%, 100%)
Very weak	≥ 100%	≥ 100%

where r_p is the time coordinate of the center of gravity of the hydrograph above the p percent flow, $p=50\%$ and $p=75\%$ [h], N_p the number of percent flows exceeding p percent flow, 6 for $p=75\%$ and 11 for $p=50\%$, $V_{p,i}$ the partial volume of the hydrograph between successive p percent flow [h], l_i the time coordinate of the gravity center r_i of the partial volume [h], and r_i the gravity center of the partial volume.

The analysis involved the assessment of the conformity between the centers of gravity of the parametric hydrographs relative to the flow axis for the percentage of peak $p=75\%$ and higher, and for the percentage of peak $p=50\%$ and higher (Fig. 10). The position of the center of gravity indicates the proportion between the rising limb volume and the recession limb volume of the hydrograph. The slope coefficient of the trend line represents the relationship between the position of the center of gravity of the parametric hydrograph r_p and the nonparametric hydrograph r_{ar}. Slope coefficient values below 1 indicate that the center of gravity of the nonparametric hydrograph r_{ar} is located further away from the q axis than the center of gravity of the parametric hydrograph r_p. Figure 10 shows that the position of the centers of gravity of both hydrographs is better in case of distribution with two shape parameters m and n than with one parameter.

The analysis of the volume of the parametric hydrographs V_p compared to the nonparametric V_{ar}, for the percentage of peak $p=75\%$ and higher, and for the percentage of peak $p=50\%$ and higher (Fig. 11), shows a better fit above *W50*. This analysis confirms that the fit of parametric hydrographs to nonparametric ones above *W75* is weak and the volume of parametric hydrographs is about 30% larger than that of the nonparametric ones.

The proposed criteria (13) and (14) offer a possibility to evaluate the determined parametric hydrograph when

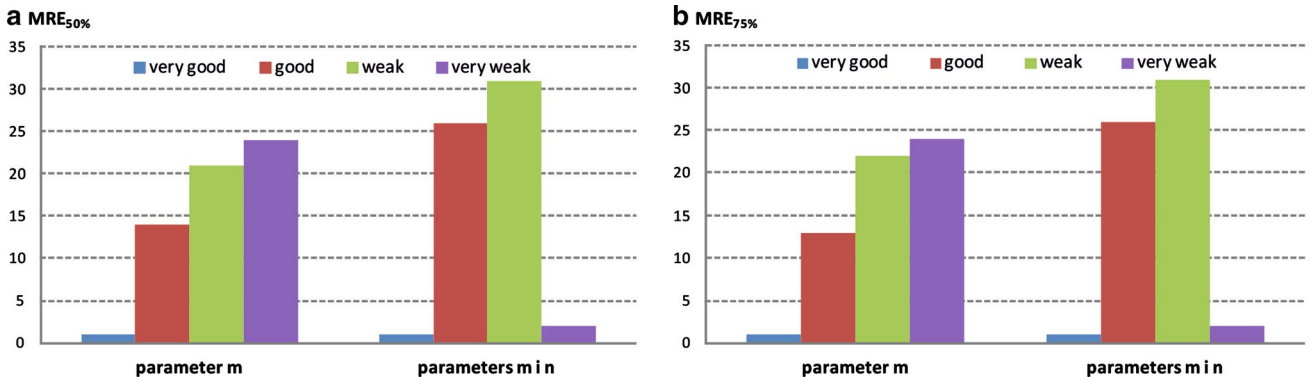


Fig. 8 Quality measures for mean relative errors for the p percent flow: **a** $p=50\%$; **b** $p=75\%$

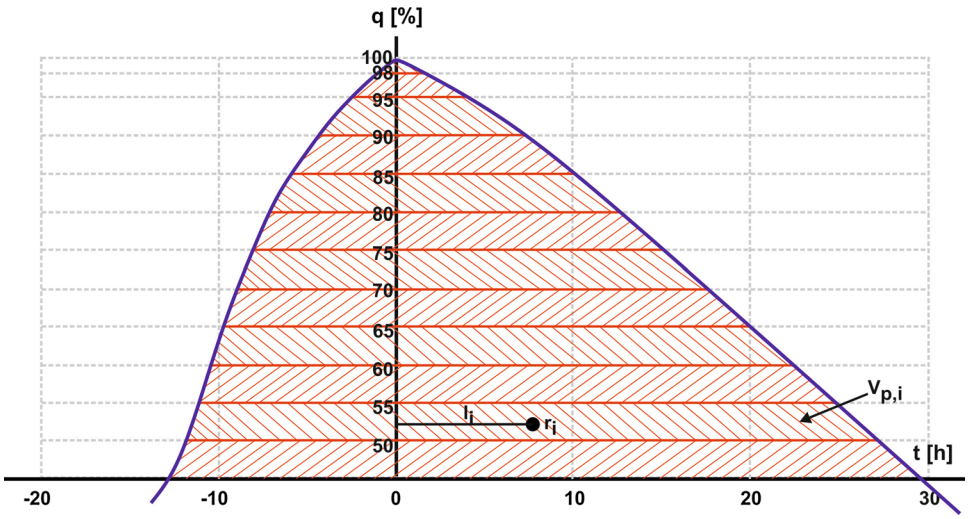


Fig. 9 Sketch for determining partial volume of hydrograph (trapezoidal area) and the center of gravity time coordinate used in Eqs. 11 and 12

compared to the input (nonparametric) hydrograph. In addition, an analysis of the absolute deviation S_s of the values of the calculated hydrograph width at 50% (*W50*) and 75% (*W75*) of peak flow, depending on the skewness coefficient s, was carried out.

$$S_s = \sqrt{(a-\hat{a})^2 + \left(b-\hat{b}\right)^2} \tag{13}$$

where a is the duration of the percentage of peak flow $p=50\%$ or $p=75\%$, or higher, for the rising limb of the nonparametric hydrograph, $a=s\ W50$ or $a=s\ W75$ [h], $\hat{a}$ the duration of the percentage of peak flow $p=50\%$ or $p=75\%$, or higher, calculated from one of the formulas (8) and (9) for the rising limb [h], b the duration of the percentage of peak flow $p=50\%$ or $p=75\%$, or higher, for the recession limb of the nonparametric hydrograph [h], and $\hat{b}$ the duration of the percentage of peak flow $p=50\%$ or $p=75\%$, or higher, calculated from one of the formulas (8) and (9) for the recession limb [h].

Figures 12 and 13 demonstrate the relationship between the absolute deviation S_s and the skewness coefficient s for the hydrograph widths *W50* and *W75*, respectively. This analysis is used to determine the possibility of using Pearson Type IV distribution in both considered water regions. The skewness coefficient of the hydrograph characterizes the proportion of the rising limb of the hydrograph to the recession limb. The smaller the value of the skewness coefficient s, the larger the share of the recession limb. The analysis shows that for hydrographs with values of the coefficient s about 0.2 and above 0.5, the compliance of parametric hydrographs above *W50* described with the Pearson Type IV distribution with one shape parameter m with nonparametric hydrographs is smaller (see Fig. 12a). In the case of two shape parameters m and n (see Fig. 12b) fit differences of hydrographs are already visible for the value of the skewness coefficient $s>0.3$. For hydrographs above *W75*, both for one and for two shape parameters, the fit for values of $s<0.4$ is less than for values $s>0.4$ (see Fig. 13). In the whole range of variation the skewness coefficient of the hydrograph s for the *W75* fitting of the both hydrographs is much worse than for the *W50*.

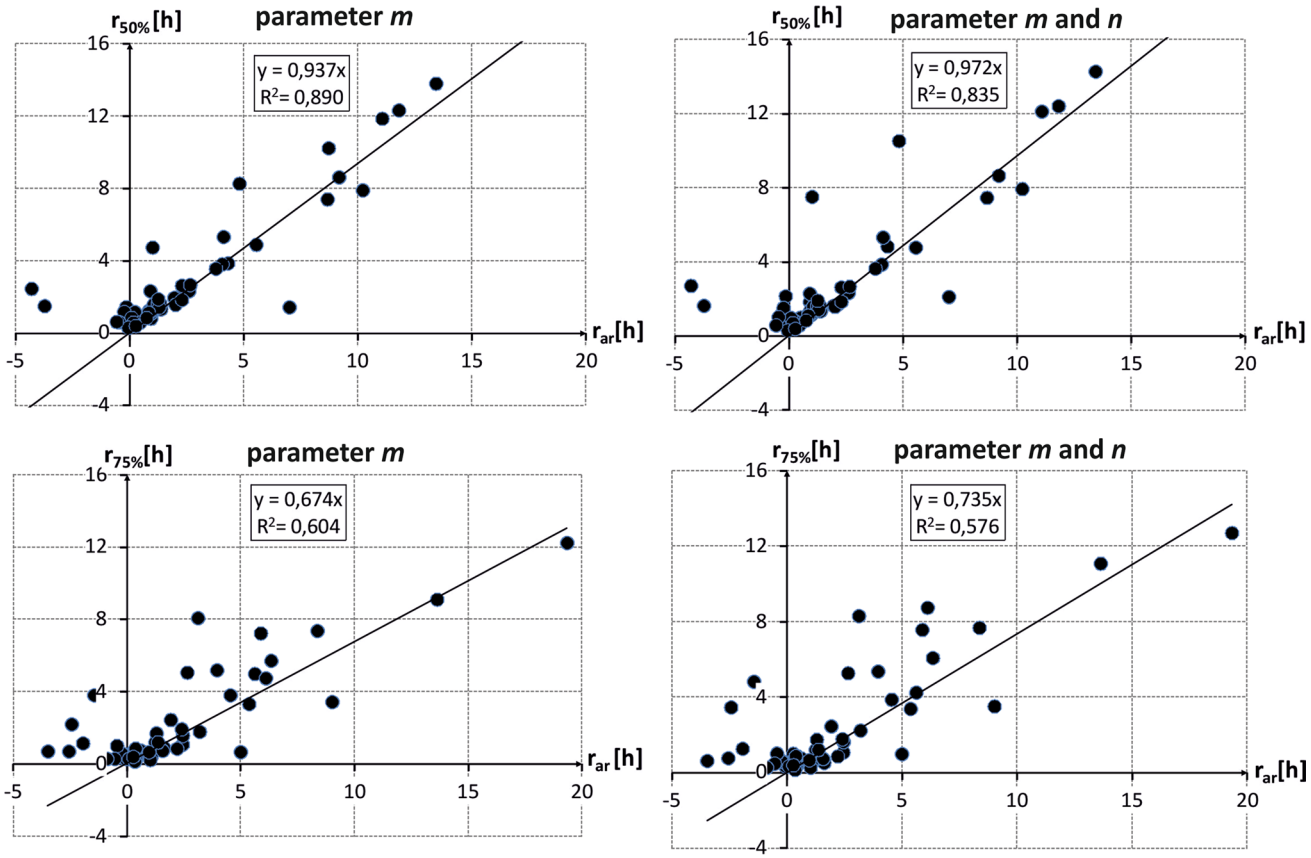


Fig. 10 Relationships between centers of gravity for parametric hydrographs determined by the Person Type IV (r_p) and the centers of gravity for Archer's nonparametric hydrographs (r_{ar})

Summary and conclusions

The gamma distribution function, i.e., Pearson Type III distribution function with one shape parameter, is the most often used function for parametric hydrographs description in the relevant literature. Authors of such publications (for example, O'Connor et al. 2014) indicate the imprecise fit of the recession limb of parametric hydrograph to the nonparametric one. One of the proposed solutions is to use a spline function consisting of two different functions describing independently two parts. The upward part of the recession limb to the inflection point, which is located between the parameters *W75* and *W50*, is described by the gamma function. The recession limb below the inflection point is described by the exponential function. In Ireland, this spline function is known as UPO gamma (unit-peak-at-origin gamma) (O'Connor et al. 2014). The research conducted for the Ireland area showed that this solution is not universal. This prompted the authors of this manuscript to find one function that would allow us to describe both the rising and recession limbs of a nonparametric hydrograph in any water gauge.

The Pearson Type IV distribution function proposed by Strupczewski concerned the description of a nonparametric flow hydrograph obtained as a medium hydrograph from unimodal recorded hydrographs. Strupczewski claimed that this distribution function is recommended to use only when the duration of the recession limb of the hydrograph is six times longer than the rising limb duration.

Current trends in hydrology recommend the use of the Archer's method for the nonparametric hydrographs description. This hydrograph represents the median durations of a given percent flow independently for rising and falling limbs. It is used to determine the value of the hydrograph width at 50% (*W50*) and 75% (*W75*) of peak flow and the skewness coefficient *s* (Archer et al. 2000). The parameters are used to determine the shape of a parametric hydrograph from *W50* to peak flow. The Archer's method allows to use the Pearson Type IV distribution function under conditions other than those considered by Strupczewski.

In this paper, the authors modified the formulas for the Pearson Type IV distribution with one and two shape parameters proposed by Strupczewski.

The analyses were conducted for the two water regions: the upper Vistula River and the middle Oder River. For each

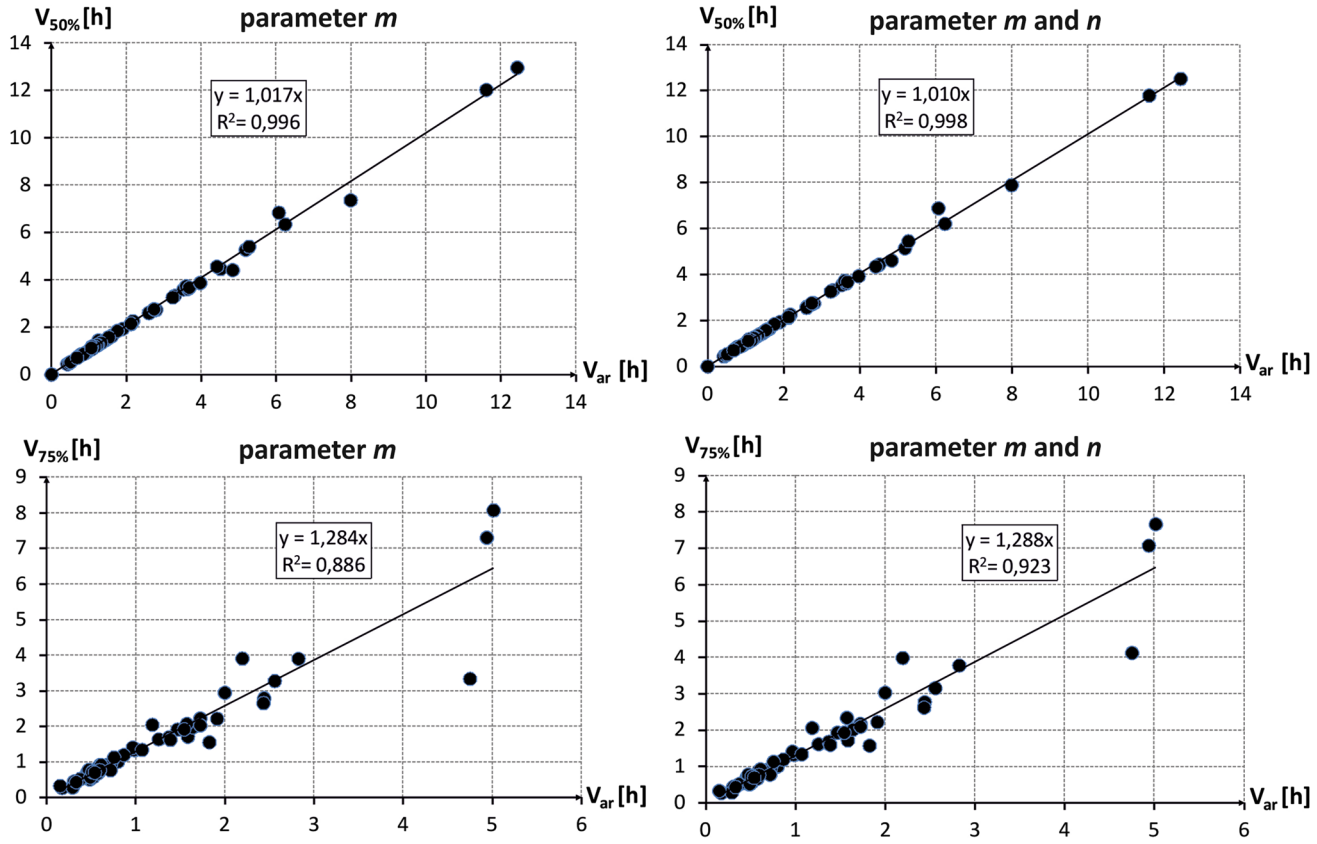


Fig. 11 Relationships between volumes of parametric flow hydrographs determined by the Person Type IV (V_p) and Archer's nonparametric hydrographs (V_{ar})

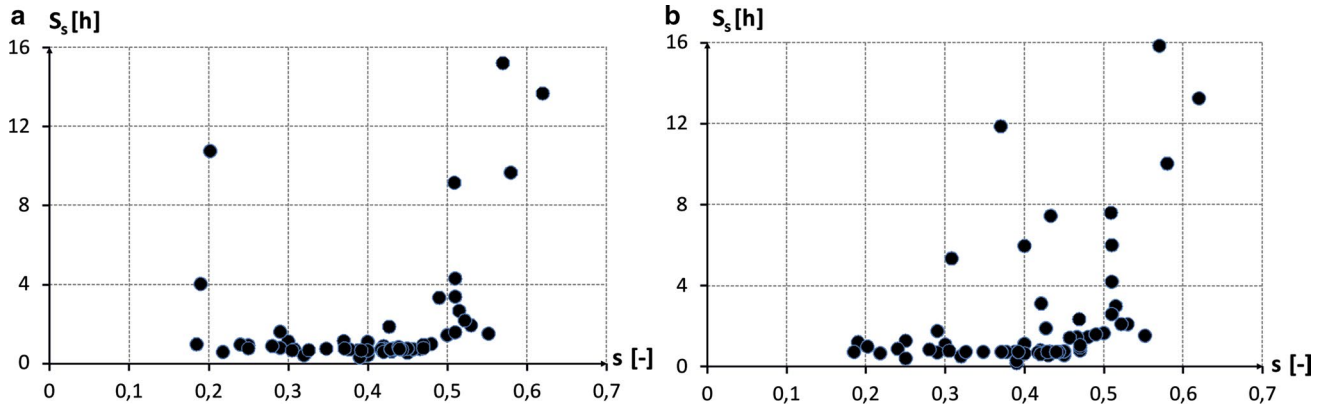


Fig. 12 Absolute error values S_s versus values of skewness coefficient s for the Pearson distribution function Type IV: **a** one shape parameter m; **b** two shape parameters m and n for *W50*

of these regions, 30 catchments were selected, for which multiannual flow records were available. In these areas, large floods occurred. The flood hydrographs were highly variable, with variable ratio of the rising/recession limb duration.

The RE_p and MRE_p measures do not answer unambiguously as to whether the functions used should be recommended for the Vistula or the Oder water regions, or not. That is why the three independent methods were used to verify the obtained results: How the absolute error changed in relation to the skewness coefficient of the hydrograph (Figs. 12, 13); what relationships are between the calculated and the input hydrographs related to the

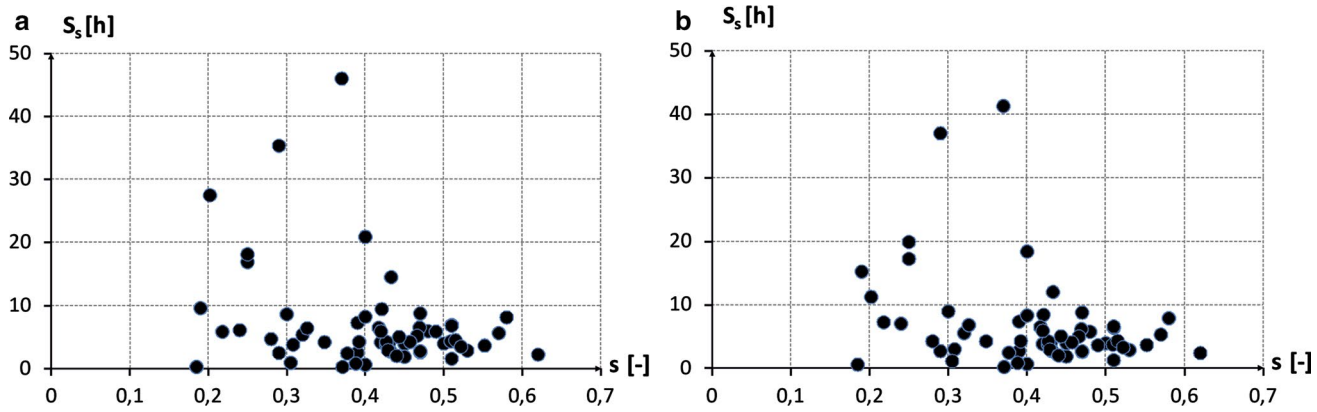


Fig. 13 Absolute error values S_s versus values of skewness coefficient s for the Pearson distribution function Type IV: **a** one shape parameter m; **b** two shape parameters m and n for *W75*

changes in the position of the centers of gravity (Fig. 10) and the volume of the hydrographs (Fig. 11). The results of relative and average relative error analysis do not allow unambiguous application or rejection of the Pearson Type IV distribution function to describe the parametric hydrograph. The remaining three analyses confirmed that both Pearson Type IV distribution functions could be used to describe the parametric hydrograph. They confirm a good fit for the recession limb of the hydrograph. In the upper part of the parametric hydrograph above the 75% percent flow (*W75*), a relatively weak fit is observed, but it does not affect either the values of the volume in this part or the position of the center of gravity of the hydrograph. The applied measures of the volume and the position of the center of gravity of the hydrograph are more objective than the relative error (RE) and mean relative error (MRE) recommended in Technical Research Report Volume III Hydrograph Analysis (O'Connor et al. 2014). When using the Pearson Type IV distribution it is difficult to state clearly what effect the skewness coefficient s has on the function's fit for the given input parameters (*W50* and *W75*) (Figs. 12, 13).

As a result of additional tests performed, it was observed that there exist many potential pairs of shape parameters m and n for Pearson Type IV distribution function satisfying the objective function criterion (Eq. 8). This is a major inconvenience because the shape parameters are to be determined for ungauged cross sections, based on the physical catchment descriptors. Thus, the development of empirical formulas using the physical catchment descriptors to determine the parameters *W50*, *W75* and s is impossible. Therefore, the Pearson Type IV distribution with a single shape parameter to describe the parametric hydrograph well enough is recommended.

References

Apel H, Thieken AH, Merz B, Blöschl G (2006) A probabilistic modelling system for assessing. Flood Risks Nat Hazards 38:295–308

Archer D, Foster M, Faulkner D, Mawdsley H (2000) The synthesis of design flood hydrographs. In: Proceedings of flooding: risks and reactions CIWEM/ICE conference, London, pp 45–57

Aziz MA, Shalaby NA, El-Shafie OM, Mahdy AT, Nishida A (2006) Comparison between the shapes of lactation curve of Egyptian buffalo milk yield estimated by the incomplete gamma function and a new model. Livestock Res Rural Dev 18(5):59

Bayliss AC (1999) Catchment descriptors volume 5 of the flood estimation handbook. Center for Scology and Hydrology, p 130

Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. Geosci Model Dev 7(3):1247–1250

Challa SNM (1997) Water resources engineering: principles and practice. New Age International (P) Limited, New Delhi, p 174

Ciepielowski A (1987) Statistical Methods of determining typical winter and summer hydrographs for ungauged watersheds. In: International symposium on hood frequency and risk analyses. Department of Civil Engineering, Louisiana State University, Baton Rouge, pp 117–124

Ciepielowski A (2001) Relationships between selected elements of the flood hydrographs in rivers. J Water Land Dev 5:89–105

De Michele C, Salvadori G, Canossi M, Petaccia A, Rosso R (2005) Bivariate statistical approach to check adequacy of dam spillway. J Hydrol Eng 10:50–57

Elshorbagy A, Simonovic SP, Panu US (2000) Performance evaluation of artificial neural networks for runoff prediction. J Hydrol Eng 5:424–427

Ernst J, Dewals BJ, Detrembleur S, Archambeau P (2010) Micro-scale flood risk analysis based on detailed 2D hydraulic modelling and high resolution geographic data. Nat Hazards 55:181–209

Gądek W, Środula A (2014) The evaluation of the design flood hydrographs determined with the hydroproject method in the gauged catchments. Infrastruktura I Ekologia Terenów Wiejskich Nr 2014/IV [3 (Dec 2014)], pp 1355–1366

Gądek W, Tokarczyk T (2015) Determining Hypothetical floods in the odra basin by means of the Cracow method and by volume formula infrastructure and ecology of rural areas Nr IV/4 2015, pp 1507–1519

Gądek W, Tokarczyk T, Środula A (2016) Estimation of parametric flood hydrograph determined by means of Strupczewski method in the Vistula and Odra catchments. J Water Land Dev 31(X–XII):43–51

Gądek W, Baziak B, Tokarczyk T (2017a) Nonparametric design hydrograph in the gauged cross sections of the Vistula and Odra catchment. Meteorol Hydrol Water Manag 5(1):53–61

Gądek W, Baziak B, Tokarczyk T (2017b) Strupczewski method for parametric design hydrographs in ungauged cross-sections. Arch Hydro-Eng Environ Mech 64(1):49–67

Hattermann FF, Kundzewicz ZW, Huang S, Vetter T, Gerstengarbe F-W, Werner P (2013) Climatological drivers of changes in flood hazard in Germany. Acta Geophys 61(2):463–477

Hayashi T, Nagamine Y, Nishida A (1986) A vibration model to describe the lactation curve of a dairy cow. Jpn J Zootech Sci 57:471–478

Jonkman SN, Vrijling JK (2008) Loss of life due to floods. J Flood Risk Manag 1(1):43–56

McEnroe BM (1992) Sizing stormwater detention reservoirs to reduce peak flow W: hydraulic engineering: saving a threatened resource. In: Search of solutions conference proceeding paper Reston, VA. ASCE, pp 719–724

Mills P, Oliver Nicholson O, Reed D (2014) Physical catchment descriptors based on research reports by compass informatics. DWRconsult and OPW, p 90

Mioduszewski W (2014) Small (natural) water retention in rural areas. J Water Land Dev 20:19–29

Nash JE (1957) The form of the instantaneous unit hydrograph In: Proceedings of Toronto General Assembly, IASH Publ No 45, pp 114–121

O'Connor K, Goswami M, Faulkner D (2014) Flood studies update. Technical Research Report, Volume III, Hydrograph Analysis, p 186

Ozga-Zielińska M, Gądek W, Książyński K, Nachlik E, Szczepanek R (2002) Mathematical model of rainfall-runoff transformation. In: Singh VP, Frevert DK (eds) WISTOO mathematical models of large watershed hydrology. Water Resources Publications, LLC, Littleton, pp 811–860

Pietrusiewicz I, Cupak A, Wałęga A, Michalec B (2014) The use of NRCS synthetic unit hydrograph and wackermann conceptual model in the simulation of a flood wave in an uncontrolled catchment. J Water Land Dev 23:53–59

Reitz W, Kreps H (1945) Näherungsverfahren zur Berechnung des erforderlichen. Stauraumes für Zwecke des Hochwasserschutzes Deutsche, Wasserwirtschaft, p T1

Snyder FF (1938) Synthetic unit-graphs. Trans Am Geophys Union 19:447–454

Serinaldi F, Grimaldi, S (2011) Synthetic design hydrographs based on distribution functions with finite support. J Hydrol Eng 101061/(ASCE) HE1943-55840000339, 434–446

Strupczewski W (1964) Równanie Fali Powodziowej (equation of flood crest). Wiadomości Służby Hydrologicznej i Meteorologicznej 2(57):35–58

Strupczewski W, Bogdanowicz E, Kochanek K (2013) Discussion of "Synthetic design hydrographs based on distribution functions with finite support" by Francesco Serinaldi and Salvatore Grimaldi. J Hydrol Eng 18(1):121–126

Vrijling JK, Van Hengel W, Houben RJ (1998) Acceptable risk as a basis for design. Reliab Eng Syst Saf 59:141–150

Wałęga A (2013) Application of HEC-HMS programme for the reconstruction of a flood event in an uncontrolled basin. J Water Land Dev 18:13–20

Zeleňáková M, Gaňová L, Purcz PM (2017) Mitigation of the adverse consequences of floods for human life, infrastructure, and the environment. Nat Hazards Rev. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000255

Zevenbergen C, Cashman A, Evelpidou N, Pasche E, Garvin S, Ashley R (2011) Urban flood management. CRC Press, London

Acta Geophysica (2019) 67:1435–1449
https://doi.org/10.1007/s11600-019-00342-x

RESEARCH ARTICLE - HYDROLOGY

Assessment of flood hazard mapping in urban areas using entropy weighting method: a case study in Hamadan city, Iran

Mehdi Sepehri[1] · Hossein Malekinezhad[1] · Seyed Zeynalabedin Hosseini[1] · Ali Reza Ildoromi[2]

Received: 10 April 2019 / Revised: 8 August 2019 / Accepted: 9 August 2019 / Published online: 21 August 2019

Abstract

Flood is one of the major natural disasters which cause enormous casualties and damages particularly in urban areas. In urban areas, studies on flood hazards have been accompanied by tensions for various reasons, including complexity of urban levels, different spatial modeling indices, lack of accurate hydrological data, and precise modeling of land surface simulations. This paper used a Constrained Delaunay Triangular Irregular Network to model fine urban surfaces (based on the detailed ground sampling data), and subsequently discusses five indicators regarding the dangers of flood, namely (1) elevation, (2) slope, (3) distance to discharge channels, (4) index of development and persistence of the drainage network (IDPR), and (5) infiltration rate. In the next step for flood hazard mapping, the combination of geographical information systems and the entropy weight method as the multi-criteria decision analysis was used to combine the indicators. The proposed methodology was used for Hamadan city that is located in the central part of Hamadan Province in Iran where several floods occur annually. The flood hazard mapping indicates that approximately 15.83% of the total study area is classified as very highly hazardous, 31.72% as hazardous, 20.11% as moderate, 16.02% as minor, and 16.32% as the least hazardous. Finally, superimposition and receiver operating characteristic (ROC) curve methods were used to verify the accuracy of the obtained flood hazard map. In terms of superimposition and ROC curve, the accuracy of the model was approximately 70% and 73%, respectively.

Keywords Flood · CD-TIN · Hamadan · GIS · Entropy weighting method

Introduction

Flood as a major natural hazard, whose impact is immeasurable, affects almost 170 million people annually (Kowalzig 2008; Kazakis et al. 2015; Wang et al. 2018). Therefore, the present study makes the flood risk management a lever to eliminate some limitations such as geographic location, socio-economic topics, and borders between nations (Degiorgis et al. 2012). Given that total elimination of flood risk is impossible and inefficient, flood risk management is divided into flood risk assessment and flood risk mitigation (Schanze 2006; Tehrany et al. 2013). Certainly, on the region scale, there are some actions such as flood hazard mapping to prepare warning systems, improve quick response and reduce the impact of flood events; the use of these actions can reduce the flood impacts (Skilodimou et al. 2019).

Several evaluation tools and indices have been developed to assess the vulnerability of urban areas from different natural disasters such as flood, earthquake, and pollution. Although these indices and tools are widely used in academic and practical environments to evaluate and monitor the urbanization process, using them can have some limitations (Tan et al. 2018). Huang et al. (2009) in their study concluded that in urban flood hazard, there are several variables preventing these indices from reflecting the systematic interactions between these variables. Therefore, in urban flood hazard studies, we cannot define the normative indications of flood hazard risk. In another study, Uwasu and Yabar (2011) asserted that in urban basins, flood indices may overlook some of the important flood variables and not explicitly identify the behavior of these variables.

Degree of urbanization is one of the most significant issues which can enhance the sophistication of environmental studies. Urbanization is a method of urban growth

✉ Hossein Malekinezhad
hmalekinezhad@yazd.ac.ir

[1] Department of Rangeland and Watershed Management, Faculty of Natural Resources, Yazd University, Yazd, Iran

[2] Department of Watershed and Range Land Management, Faculty of Natural Resources, Malayer University, Hamadan, Iran

which increase built-up land uses. These land uses in hydrology studies are mainly called hard features (the term hard features is pertinent to features changing the flow direction such as curbs, road camber, walls, and buildings) based on the studies by Li et al. (2014), Rodríguez et al. (2017) and Sepehri et al. (2018), which are defined as Fine Constrained Features (FCFs). Obviously, these built-up areas transform the surface features and flood propagation. Therefore, these FCFs and their properties must be considered to increase the accuracy degree of flood hazard mapping in urban surfaces (Li et al. 2014).

After simulating the urban surface feature and extracting the flood indicators from simulated surface, weighting the indices and combining them should be carried out.

The existence of nonlinear and complex relations between these assessment indices and risk levels creates a serious challenge in confirming the accuracy of the obtained maps. In recent decades, various systematic methods including the analytic hierarchy process (AHP) method (Fernández and Lutz 2010; Stefanidis and Stathis 2013; Yang et al. 2013), set pair analysis (SPA) method (Zou et al. 2013; Guo et al. 2014), and entropy (Kawachi et al. 2001; Lotfi and Fallahnejad 2010; Mishra et al. 2009) and fuzzy method (Lohani et al. 2014; Li 2013) have been developed to overcome this problem. Although these methods are to analyze flood hazard risk as an efficient tool, they are also associated with weaknesses and uncertainties due to their complex and difficult structures.

Fernández and Lutz (2010) in their study expressed urban flood hazard mapping in the Tucuman Province using the integration of five factors into AHP method, from which they concluded that the distance to discharge channel has a substantial role in the determination of susceptibility zones. Elkhrachy (2015) investigated flood sensitivity in Najran city in Saudi Arabia. In this study, he used the AHP method to determine the flood susceptibility map for weighting and the relative importance of effective indices in flood hazard risk. He concluded that runoff index and soil type had the most important roles in the spatial distribution of flood hazard mapping, compared to other indices such as macro-roughness, slope, land use, and drainage network. These two indices alone were involved approximately 35.5% to the final flood hazard map.

Bathrellos et al. (2016) delineated flood hazard mapping by integrating AHP method and GIS in the basin of Athens Metropolitan city, Greece.

Gigović et al. (2017) reviewed and compared three scenarios for flood hazard mapping in Palilula Municipality, Belgrade, Serbia. In the first scenario as the basic scenario, authors considered flood hazard mapping using AHP method as the common and classic method of multi-criteria decision making. In the second scenario, the authors used a method called "the interval rough numbers (IR'AHP)" to reduce the uncertainty of the AHP method. In the last scenario, they used the combination of the fuzzy and the AHP method to reduce uncertainty of AHP. The results of this study indicated that the second scenario had the highest accuracy compared to other scenarios.

Xu et al. (2018) simulated the rainfall runoff using SWMM and flood hazard mapping to increase the accuracy of the mentioned model and combined the results of flood hazard mapping considering two popular multi-criteria decisions, i.e., entropy weighting and AHP method. The results showed that the combination of these models had higher accuracy than their single usage.

Lee et al. (2018) used the ratio frequency and the logistic regression method for mapping the 2011 flood Seoul, South Korea. At first, they selected and optimized flood indices. Finally, flood hazard mapping showed that both methods had the accuracy of 79.6 and 79.05% for the ratio frequency and the logistic regression method, respectively.

Mahmoud and Gan (2018) surveyed and zoned flood hazard in Riyadh, Saudi Arabia. At first, they provided ten effective flood hazard indices. Then, they weighted and combined these indices to produce a flood hazard map using the AHP method. Finally, to measure the accuracy of the obtained map, they used flood data that occurred in the past.

There are few studies considering a differentiation aspect with natural areas for urban flood hazard mapping.

Li et al. (2014) investigated the depth of flooding at the enclosure of Beijing Normal University. To do this, they first simulated surface features such as buildings. Then, they estimated the depth of flooding in different areas using mass conservation law and rainfall data, runoff and drainage network capacity. The comparison between the final map of flooding and field observations indicated the high accuracy of the method used in simulation of flooding. Hence, they suggested that the used method could be used as a powerful tool in simulating the drainage network and installing urban flood hazard warning systems.

Rodríguez et al. (2017) simulated the hydrological behavior of sub-urban basins based on morphological features. So, they applied three methods. The first method known as the "traditional method" is based on the digital elevation model (DEM). The second method is pertinent to urban database analysis; it is based on the vector data. The third and final method is CD-TIN which can preciously describe flow direction from a slope. The results showed that the DEM method had the lowest accuracy in simulating the hydrological behavior. Two remaining methods have better accuracy to simulate the hydrological behavior for their ability in representing the areas with high drainage density. Finally, they asserted that CD-TIN for its surplus ability in the representation of topological discontinues (streets, curbs and so on) could be a more appropriate method for urban areas.

Despite the few studies into urban flood hazard mapping investigating the effect of urbanization degree, it has recently become popular among researchers.

From the past century until now, the study area has experienced various floods. Early studies of archeology in Hegmataneh, Hamadan Province, indicated that destruction of this historic region is relevant to the floods in the past centuries. Hamadan province has a population of approximately 554,406 people, which is distributed in an area of 73.5 Km^2. From the social and geographical point of view, Hamadan is one of the most vulnerable cities and the second most populated city in the western region of Iran (Sepehri et al. 2019).

Therefore, it is necessary to develop flood hazard mapping in Hamadan city in order to flood mitigation, land-use planning, and urban construction. This study aims at: (1) development of a simulation model for displaying urban characteristics, (2) definition and elicitation of effective indices in urban flood hazard mapping from step one, (3) application of the entropy method as the multiple criteria decision method for weighting indices and combining them for final flood hazard mapping, and (4) investigation and evaluation of the accuracy of the obtained flood hazard map.

Finally, it is expected that the knowledge will be useful for public policymakers and locals who are unaware of the benefits of flood hazard mapping.

Materials and method

The study area

The study area with an area of over 73.5 Km^2 is located on the northern hill slopes of Mount Alvand, the central part of Hamadan Province in central Zagros. According to the Hamadan Meteorological Station reports in the Hamadan metropolitan area, the average annual temperature is approximately 9.6 °C, varying between + 36.8 °C (summer) and − 29.6 °C (winter). Moreover, the average annual rainfall is 343.11 mm. According to the Ombrothermic chart, the driest months are from May to September. Furthermore, based on the Ambreget technique, local climate is cold, semiarid, and sub-humid.

Historically, Hamadan's urban population growth was largely limited to the north and northwest of the region. However, in recent years, due to various reasons such as luxury orientation and urban development, it has also rapidly expanded to the southern parts of Hamadan as today the city is divided into two parts, the north and the south of the city or into the upper part of the city and the lower part of the city.

Surface water collection and disposal system of Hamadan are generally defined in three sections, including networks, main transmission lines, and drainage disposals (natural rivers). A network drawing water from the level of streets and alleys leads to the main lines. The transmission lines that usually run on main roads are responsible for transferring the water of their basin into disposal areas (which they have received through the network). In most cases, rivers, streams, lakes, or seas are considered as disposal areas. In this study, the disposal areas are Abbas Abad Rivers, Dewin, Moradbeik Valley, and Khezr, which are the natural trough line of the city and direct input runoff from the south to the north. These drainage systems worked well in the past; however, due to excessive urban development and rainfall intensity increase in recent years, there has been an increase in the rate of stronger flows, which is a superfluent on the drainage capacity of the system.

The results of intra-urban hydrology and network hydraulic studies indicate that a large portion of the Khezr River, a part of Dewin, and the end of Moradbeik are devoid of a sufficient capacity for flood safety passing with a return period of 50 years. As a result, the superfluous flow on drainage capacity moves toward downstream areas or more vulnerable areas, which can cause economic, psychological, and even life-threatening damage. According to the reports, the storm that occurred in November 1993 at 53 mm in intensity caused high economic damage to downstream areas as well as substructures. Considering that in the upper regions of the study area, there is a sudden increase in the slope and elevation from the southern side of the city to the north of the city, the flood characteristics are different in these two regions. In the upstream areas, due to the precipitous slope, the runoff is formed more quickly and moves rapidly toward the downstream areas. However, these areas have become susceptible to stagnation after a number of occurrences for various reasons such as arrival flows from the upstream areas, rising rainfall intensity in recent years, and the lack of an appropriate drainage system (Fig 1).

Flood hazard index (FHI)

In this study, the information was collected from different sources and ultimately extended to a database in a spatial environment. Therefore, an index model was established to identify the susceptible flood areas with a local focus in the GIS environment. The proposed model performs the flood hazard mapping using the multi-criteria analysis. In general, the objective of flood hazard mapping is to identify flood-prone points and perform comparative analyses between different basins. Figure 2 shows the proposed method. Initially, information from different databases was imported to the GIS environment. Next, the processed data, along with the definition of the entropy weighting method, the flood hazard mapping, were prepared.

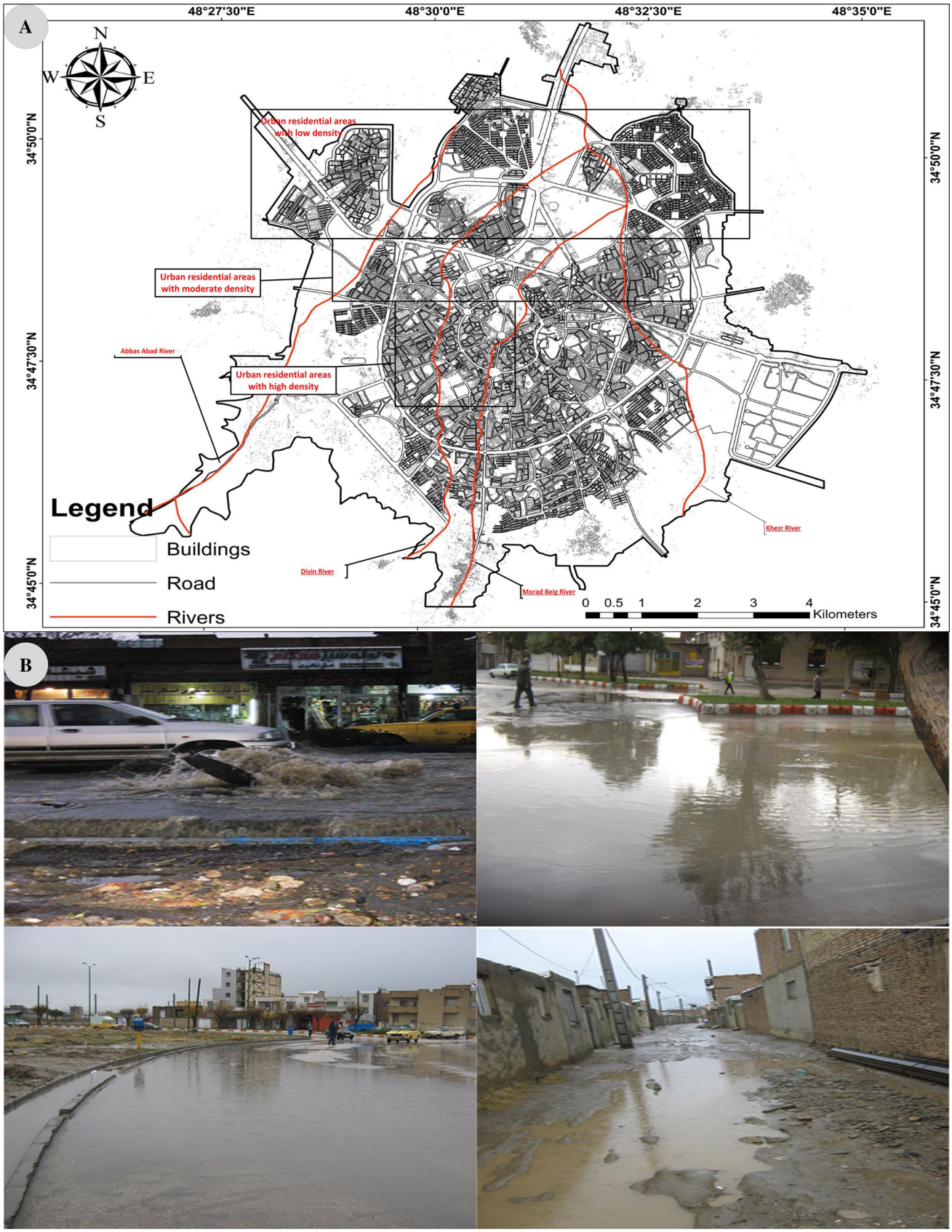


Fig. 1 **a** Schematic of the study area. **b** Photographs of flood taken on the case study (https://www.mashreghnews.ir/photo/832347)

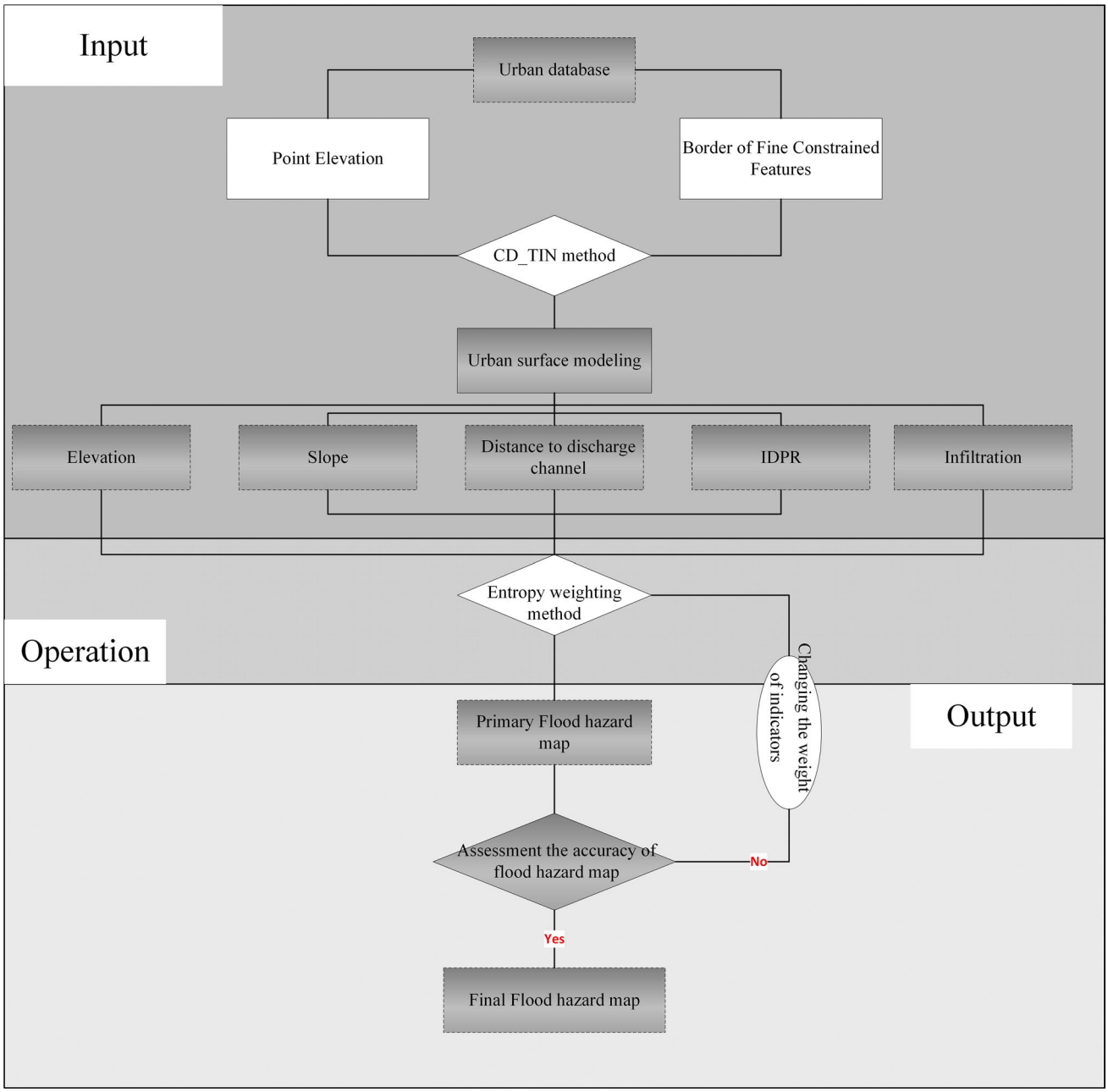
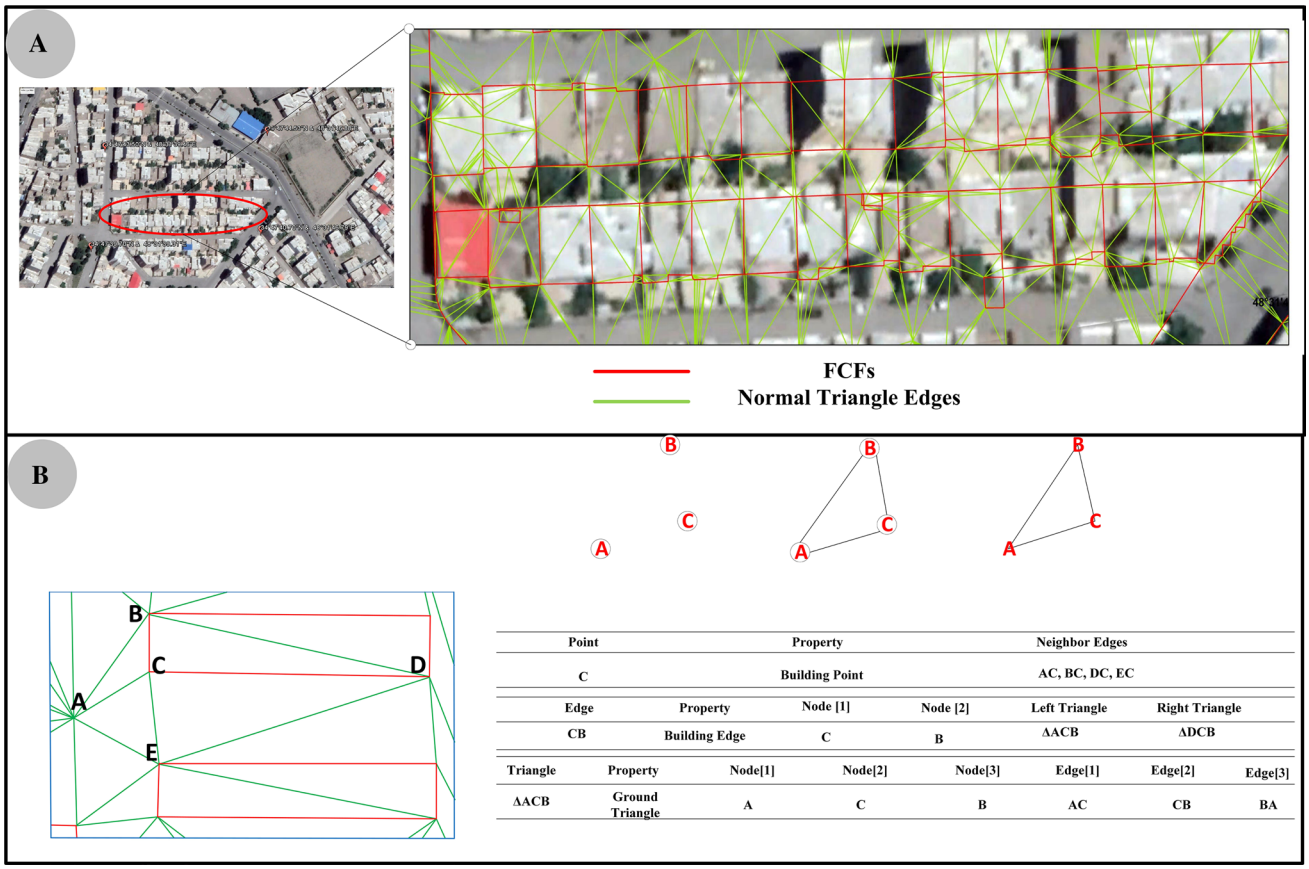


Fig. 2 Flowchart for the preparation of urban flood hazard map

Parameters included in the FHI

Modeling urban surface with CD-TIN

In this study, a Constrained Delaunay Triangle Irregular Network (CD-TIN) representation based on Bernal and Sloan's method is used to model the complexity of urban surface and their details. The CD-TIN not only considers the shape or boundary of the FCFs but also enters the properties of FCFs. These features can significantly affect the flow propagation on urban surfaces (Li et al. 2014) (Fig 3). For example, the flow direction changes when it encounters a building or other rigid walls; reduction in water runoff volume when it moves across a grass land is different from the time this flow moves across a playground. To simulate urban surface and their details by CD-TIN, a high-precision DEM is also required in addition to the properties of FCF_s. In this study, elevation maps with a scale of 1/2000 were used for the metropolitan area. In the 1/2000 maps of Hamadan, two groups of elevation digital models are presented as a group of digits taken from the floor of the road and the roof of buildings. Using these digits, a DEM of metropolitan Hamadan was prepared. On the other hand, The SRTM



Point	Property	Neighbor Edges
C	Building Point	AC, BC, DC, EC

Edge	Property	Node [1]	Node [2]	Left Triangle	Right Triangle
CB	Building Edge	C	B	ΔACB	ΔDCB

Triangle	Property	Node[1]	Node[2]	Node[3]	Edge[1]	Edge[2]	Edge[3]
ΔACB	Ground Triangle	A	C	B	AC	CB	BA

Fig. 3 **a** Demonstration of urban surface using CD-TIN topology. **B** The red lines are related to constrained features such as buildings and street curbs. The normal triangle edges are shown with green lines (Li et al. 2014)

DEM (90 m) maps were used in sub-urban and upstream areas, which are substantially pertinent to agricultural and natural tourist lands. Finally, by combining these two elevation maps, a DEM was prepared from the city of Hamadan. To prepare FCFs, the database of the municipality administration of Hamadan province was used (Table 1).

Figure 3b presents the properties of triangle ABC and its elements (point A and edge BC) to assess their impacts on urban surface runoff. In topology of CD-TIN, the properties of neighbor triangles with their elements will be also considered for future confluence calculation (Li et al. 2014).

Table 1 shows the types of FCFs in shape file format (ARC GIS software prevalent format) added to the CD-TIN algorithm.

Definition of indices

To perform flood susceptibility, it is essential to first determine the flood conditioning indices (Ghiglieri et al. 2014). An acceptable flood hazard map is highly dependent on the quality of the spatial and temporal data required to be taken from the case study. Unfortunately, many case studies, particularly in developing countries, are ungauged or poorly gauged (Sivapalan 2003). In some cases, the number of the existing gauging stations has decreased. Furthermore, the existence of multivariate and nonlinear relationships between indices and risk levels is a major intrinsic challenge for flood hazard risk assessment (Wang et al. 2015). Therefore, providing a flood hazard map in

Table 1 FCFs in the study area and their entry type to the CD-TIN algorithm (Li et al. 2014)

Constrain features	Data type	Data organization
Road cambers, street curbs, road isolation strip, walls, etc.	Constrained polyline	Shape file polyline
Buildings (or other man-made structures), grass lands, playgrounds, lakes (or other water bodies), etc.	Constrained polyline	Shape file polygon

these areas is a serious challenge. The first step in this regard is to select appropriate indices.

In most studies on flood hazard mapping using multiple criteria decision, researchers focus on discrete or integrated indices. In discrete indices, only the characteristics of the desired cell (such as the elevation map) or adjacent cells (such as slope map) are investigated. In integrated indices, only the integrated characteristics of the upstream (such as flow accumulation) of the desired cell are examined. However, one of the most important indices that can be mentioned in flood hazard studies or in other hydrological studies is associated with connectivity indices (Mayor et al. 2008; Gay et al. 2016; Covino 2017). In these indices, the characteristics of upstream and downstream of the desired cell are checked. Heckmann et al. (2015) in one study, conducted in 2015 about effective indicators of flood and erosion hazard mapping, concluded that if we can fully model reality, then there is no need for connectivity indices. However, the indices used for flood hazard mapping are different in all studies, and the choice of these indices is most arbitrary without adherence to specific regulations (Shadman Roodposhti et al. 2016; Patra et al. 2018).

After modeling urban surface using CD-TIN, a composite flood hazard index based on five causal indices, such as elevation, slope, distance to discharge channels, IDPR (connectivity index), and infiltration extracted from surface modeled, was used in this study (Fig. 4). These indices were selected based on the information of different case studies with similar characteristics.

The method of weighting the indicators based on the entropy weight method

In the evaluation system, determining the weights for all indices is an important process that can measure the impact of the indices. When the weight of an index is high, it considerably impacts the ability (of intended target); unless, the effect is less. In the information theory, information entropy is an important concept that can measure the amount of useful information produced in a system. The main factors for the entropy weight method are as follows: When the data of evaluated multiple objects on an index show a large difference, the entropy value of this index should be low in accordance with the information theory. This indicates that this index can have more useful information; therefore, the weight of this index should be set high; otherwise, when the entropy value of an index is high, it may limit useful information with respect to the information theory, and its weight should be very low. The methods of weighting the indices are as follows (Singh 1997; Kawachi et al. 2001; Liu et al. 2010):

1. The main data of all indices should be normalized, and this can eliminate the impact of the dimension. In a profitability index, the higher the value of the index is, the higher the amount of flood hazard risk will be. Equation 1 is used to normalize the indices, and for any index whose value and the amount is low, the flood hazard risk will be higher. Equation 2 is used to normalize the data.

$$P\left(x_{ij}\right) = \frac{x_{ij} - \min\left\{x_{ij}\right\}}{\max\left\{x_{ij}\right\} - \min\left\{x_{ij}\right\}} \tag{1}$$

$$P\left(x_{ij}\right) = \frac{\max\left\{x_{ij}\right\} - x_{ij}}{\max\left\{x_{ij}\right\} - \min\left\{x_{ij}\right\}} \tag{2}$$

where X_{ij} corresponds to the values of the $i = 1, 2, 3, \ldots, m$, index, which are used in the flood hazard mapping (subscript of $j = 1, 2, 3, \ldots, n$ is considered to be number 1, due to the objective of the study in this paper that is concerned with floods hazard mapping).

2. To evaluate the issue with m index and the n estimated target value, the entropy value Pi for the ith index can be defined as:

$$H(X) = -\left(\log_2 m\right)^{-1} \cdot \sum_{j=1}^{n} f(x_{ij}) \log_2\left[f\left(x_{ij}\right)\right] \tag{3}$$

where $f_{ij} = \frac{p_{ij}}{\sum_{i=1}^{m} p_{ij}}$ (where $f_{ij} = 0$; it is assumed that the value $f_{ij} \ln f_{ij} = 0$).

The estimation of the accuracy of the flood hazard map

In this study, superimposition and ROC curve methods were used to evaluate the accuracy of the final flood hazard map. In the superimposition method which is simple and precise, the flooding points of the study area are superimposed on the final flood hazard map. Eventually, the accuracy of the final flood hazard map is calculated based on the percentage of points that are in very high flood classes (Machiwal et al. 2011; Thapa et al. 2017). The ROC curve method which is applied to evaluate the accuracy of the simulation model is a graphical representation of the equilibrium between the positive rate (plotted on the Y-axis) (Eq. 4) and the negative rate (plotted on the X-axis) (Eq. 5) of the error for each of the possible values of the cut-point between the cases and controls (Unal 2017).

$$r_{tp} \approx \frac{\text{Positive correctly classified}}{\text{Total positives}} \tag{4}$$

$$r_{fp} \approx \frac{\text{Negative sin correctly classified}}{\text{Total negatives}} \tag{5}$$

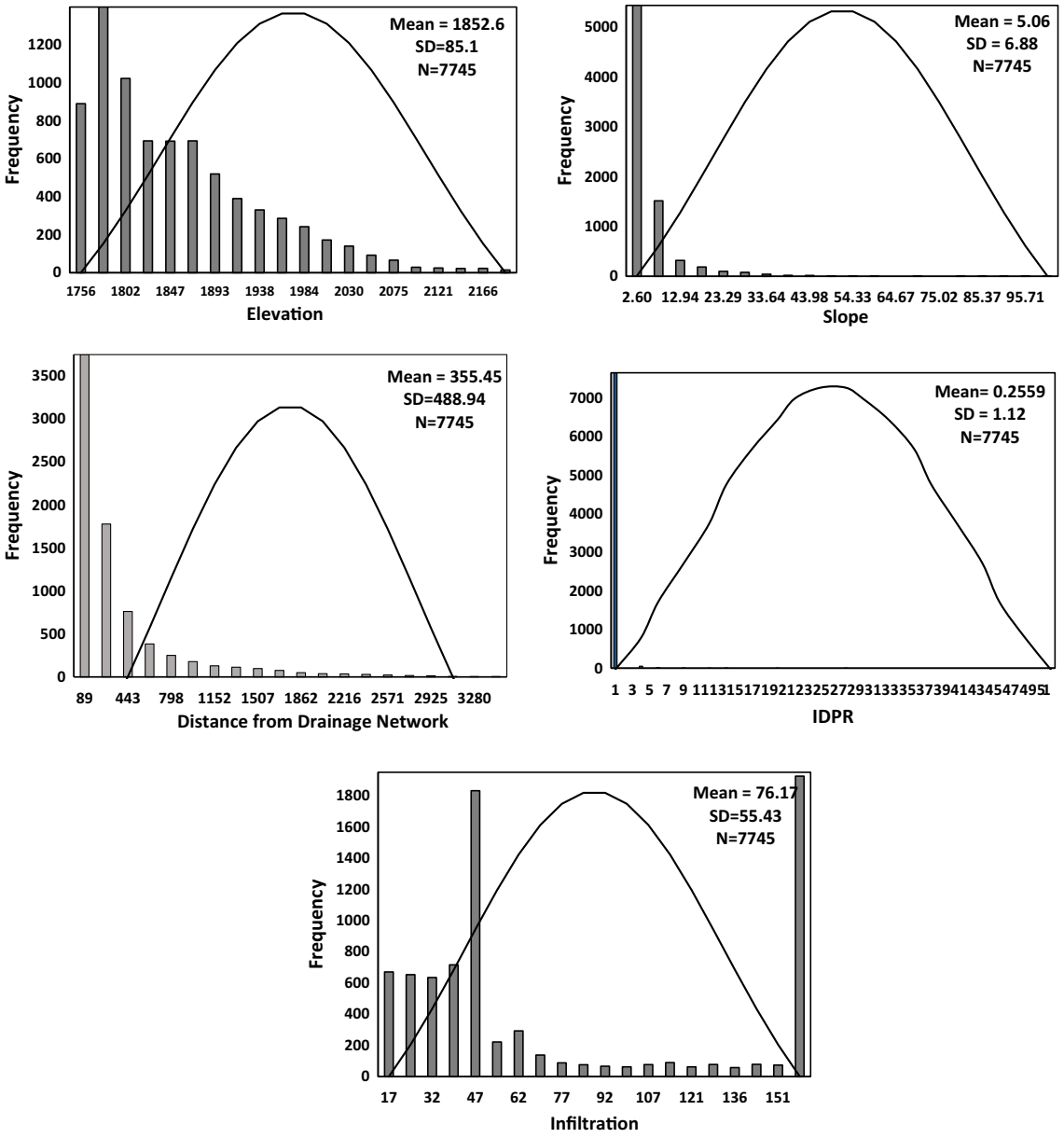


Fig. 4 Image showing the indices incorporated within the model as GIS layers and their classification

If the value of the ROC curve index, which is the area under the ROC curve (AUC), is equal to 1, it indicates the full fit of the model. And if this value is equal to 0.5, it indicates the inadequacy of the simulation model. The quantitative and qualitative correlations of the area under the ROC curve and evaluation of estimation are as follows:

1 – 0.9 complete. 0.9 – 0.8 very good. 0.8 – 0.7 good. 0.7 – 0.6 moderate, and 0.6 – 0.5 low.

Results and discussion

Development of flood susceptibility factors database

Assessment and zoning of flood risk is essential for sustainable development of human settlements in the urban

environment discussions due to their development on the margins of rivers, margins of flood plains, attention toward the hydrological and dynamic conditions of rivers, and upper parts of the basin, which increases the risk of flood and damage life, financial, and infrastructure. Therefore, it is necessary to identify the causes of flood and predict the extent of damage by which it is caused (Bathrellos et al. 2017; Dysarz et al. 2019; Radwan et al. 2019). To do this, the surface map of the studied area was simulated using the model in Sect. 2.3.1. After preparing this map, all the effective flood indices were extracted and then the entropy weighting method was used for weighting indices and combining them for final flood hazard mapping.

Elevation

The elevation index of a region can be a direct criterion for macro-roughness. This index plays a vital role in the hydrological response of different areas in the study. Areas with lower elevation (downstream areas) tend to be more likely to hold water. However, in areas with higher elevation, large amounts of precipitation change into runoff and small quantities penetrated into the earth (Fernández and Lutz 2010; Mahmoud and Gan 2018; Lyu et al. 2019).

The elevation values from the southern part to the northern have a decreasing rate of 2204.25–1744.58, since the highest degree of flooding occurs in lower elevation and vice versa. Therefore, Eq. (3) is used to examine the experimental probability of this index, which ultimately amounts to the entropy of this index that also varies from 0.367 for elevation of 2204.25 to ≈ 0 for elevation of 1744.58 (Fig. 5).

Slope

Water flow moves from a higher to a lower elevation. Therefore, this index plays an important role in the direction of water flow and the depth of the water table as well as the determination of the flooding potential of different regions of the study area. The southern and middle parts of the study area have the most precipitous slope compared to other regions (Fernández and Lutz 2010; Mahmoud and Gan 2018; Lyu et al. 2019).

Since the areas with a mild slope affect the flood hazard drastically, Eq. (2) is used to study the probability of the slope values. Hence, this index has entropy values in the range of 0.36- ≈ 0, with high entropy values for low slopes (Fig. 5).

Distance to drainage network

In addition to areas where the flow of water is directly concentrated, river flooding can be considered an important issue as it may strike the surrounding lands, as well. Therefore, the role of the bed decreases with increasing the distance from the evacuation channels. According to local reports, most flood-susceptible areas are adjacent to the rivers where excessive water flows to these areas. Investigating the relationship between the distance to discharge channel and flood sensitivity can be mentally considered. However, the ideal solution to investigate this relationship should be established on historical records. For example, Samanta et al. (2016), in a survey on the relationship between the distance from river index and the flood hazard risk, placed the areas located in less than 100 m from the drainage network at a very high flood hazard level, and on the opposite side, the areas located in more than 2000 m were placed at a very low flood hazard level. In one similar study conducted by Pradhan (2009), areas located in less than 90 m from the drainage network were placed at a very high flood hazard level.

For this index, similar to the slope and elevation map, Eq. (2) is used to test the probability of this index. The areas near the drainage network have the highest entropy (0.36) reduced by increasing the distance from the drainage network to 0.001 (Fig. 5).

IDPR

One of the major effects of urbanization is the alteration of natural stream channels and installation of sewerages. These changes can affect the hydrological connectivity. Hydrologic connectivity is a broad term used in various contexts by numerous researchers though its meaning is often different among different fields of study (Hall and Ellis 1985; Hernandes et al. 2018).

Hydrology connectivity as one of the key characteristics in runoff distribution in the region means the location connection of the site of runoff production in the upstream region to the receipted region of the downstream part. In this study, the hydrological connectivity coefficient, called IDPR, was used by Gay et al. 2016 Indeed, the IDPR index, which suffers from range of 0 to $+\infty$, states that runoffs passing through the hill slopes reach theoretical drainage network before reaching the actual drainage networks (Eq. 6). Obviously, the larger IDPR is, the more dominant is the runoff phenomenon to infiltration, and naturally has a high hydrological connectivity (Fig. 5).

$$IDPR = \frac{\text{The least cumulative cost distance for each cell to the nearest theoretical water cource over the slope surface}}{\text{The least cumulative cost distance for each cell to the nearest real water cource over the slope surface}} \quad (6)$$

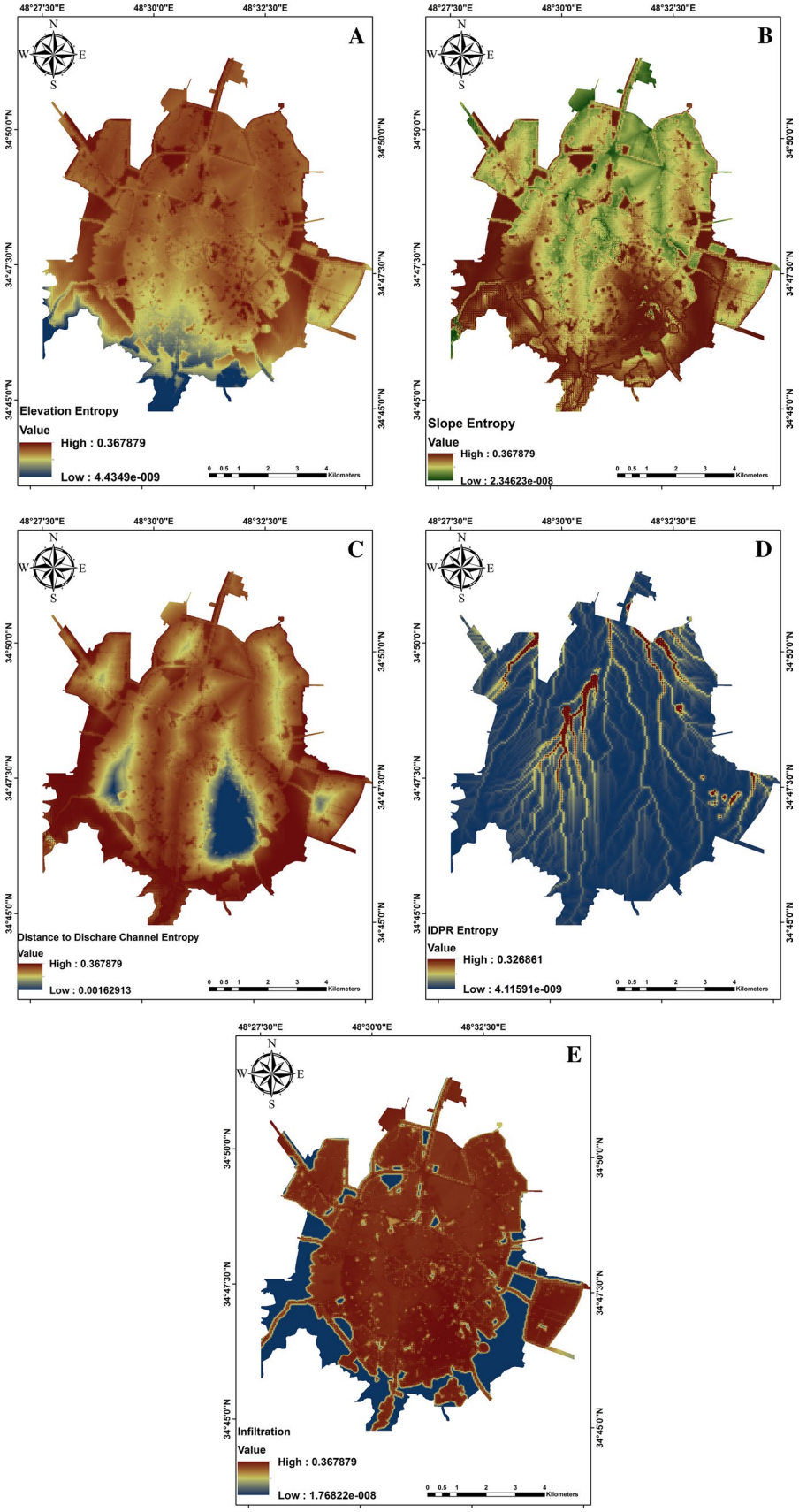


Fig. 5 The entropy value of the used indices. **a** Elevation. **b** Slope. **c** Distance to drainage network. **d** IDPR. **e** Infiltration

Table 2 Land use and their percentage of area in the study area (Sepehri et al. 2018).

Land use	Area covered (%)
Residential buildings	32
Roads and side walks	33
Parks and gardens	14
Higher education, culture	6
Industry and workshop	3
Storehouse	2
Commercial	2
Urban Facilities	1
Health and treatment	1
Sports	1
Other land uses	5

Table 3 CN values for different land uses in the study area (Sepehri et al. 2018)

Land use	CN
Garden, park, forest park	61
Higher education, culture, sports, military, cemetery	69
Health, primary, secondary, high school, technical–professional, other educational, office, hotel, residential low density, medium-density residential, high-density residential, religious, show home	85
Parking, industry and workshop, utility, commercial, municipal terminal, domestic terminal, passenger terminal, streets, side walk	92
	95

Infiltration

Infiltration is one of the most important characteristics of land use which considerably affects flooding, because land use absorbs water with its high infiltration rate, thereby producing fewer runoffs. In urban areas with increasing land-use varieties, the variety of infiltration rate also increases (Table 2). In this regard, some methods exist such as Horton, Green_Ampt and Soil Conservation Service (SCS) methods, among which the SCS method was applied because it was a popular method that is simple, flexible, and in need of merely one parameter called Curve Number (CN) (USDA 1986) (Eq. 7). CN has a range from 30 to 100 (Table 3); lower numbers indicate low runoff potential, while larger numbers are for increasing runoff potential. The lower the curve number, the more infiltrated the soil.

The relative weighting for this index varies from 0.36 to ≈ 0 for infiltration rate of 13.36 to 162.39mm (Eq. 2) (Fig. 5).

$$S = \left(\frac{25400}{\mathrm{CN}}\right) - 254 \tag{7}$$

S = Infiltration

Advantages and disadvantages of the approach

Generally, multiple criteria decisions prioritize options (Cox 2009; Malczewski 1999). Therefore, other multiple criteria decision capabilities determine the best and most effective options compared to other suitable options, determining the acceptable options or rejecting them. In this research, the entropy weighting method, known as an objective multiple criteria decision method, was used for flood risk mapping. This technique has been widely used to solve multiple criteria decision problems and is widely used in stability analysis and natural hazards. Two samples of the most important advantages and one of the most important disadvantages of using the above method are as follows:

Advantages

1. One of the most important constraints in multiple criteria decision is the uncertainty debate so that after the flood map is prepared, the uncertainty of the map could be examined. This uncertainty causes errors in the accuracy of the intended map due to the classification made by Smithson (1989). Unlike other methods introduced in the introduction, entropy method (such as AHP and SPA) employs the natural distribution of the values of these indices in the initial weighting for the indices. However, in other methods such as AHP classified in the category of subjective methods, experts' opinions should be deemed in initial weighting of the indices; these opinions, in addition to the impact of the scientific level of individuals, are affected by other conditions, such as the experts' mental conditions. This issue creates a high degree of uncertainty and thus reduces the accuracy of the final map.
2. In this method, each point of the study area (pixels) can have a primary weight. However, in the other methods mentioned in the introduction, there is a limitation in the initial weighting, so that in the AHP it takes place using a limited 9-point table. Therefore, in these methods, each of the indices should be out of the pixel mode and classified. Then, a specific weight according to their role and importance in the target should be given to each class. The question arising here is how this classification is conducted and how many classes should be chosen; this topic causes uncertainty in the index and the intended target. Furthermore, this classification method also creates unusual and unconventional flood hazard zones (e.g., the flood hazard zone with a very high

degree of risk has an over-real area), resulting in higher false precision in flood hazard zoning map. Therefore, this case becomes more important, particularly in urban basins where any management decisions are costly due to the high cost of urban use.

Disadvantages

In the entropy method, the basic assumption is that all indices have the same importance in the intended target. However, in other methods such as the AHP, the significance of the indices is also measured in relation to each other and due to the intended target. According to Huang et al. (2009), Uwasu and Yabar (2011) and Tan et al. (2018), urban basins have a complex mechanism reducing the relative importance of the indices in relation to each other.

The areas with low and very low flood hazard risks, with an area of 23.77 km^2, account for approximately 33.34% of the total area under study. The major part of this class is located in the margin and upstream of the study area. In the marginal areas, the main land use is relevant to agricultural lands and the upstream area is pertinent to natural and touristic areas. Inside the studied area, this degree of flood hazard risk is sporadic, as seen with small areas that are mainly pertinent to agricultural lands, parks and natural spaces. The areas with moderate flood hazard risk account for approximately 14.78 km^2, or 20.11% of the study area. This class of flood hazard risk is located in the middle of the upstream region of the studied area. In this section, we see a sharp increase in the slope degree and elevation from the downstream side of the study area to the upstream (dotted line A in Fig. 6). As noted in Sects. 3.1.1 and 3.1.2, with increasing the slope and elevation, the risk of flood hazard decreases. However, these areas with this degree of flood hazard risk have a high ratio of impermeable surfaces to permeable surfaces, indicating the high importance of these two indices in flood hazard risk. The fourth class of the studied area is relevant to areas with high flood hazard risk. These regions with an area of 23.31 km^2 (31.72 % of the total studied area) have dedicated the highest proportion of flood hazard risk than other areas. Also, Sepehri et al. (2018) in their paper using

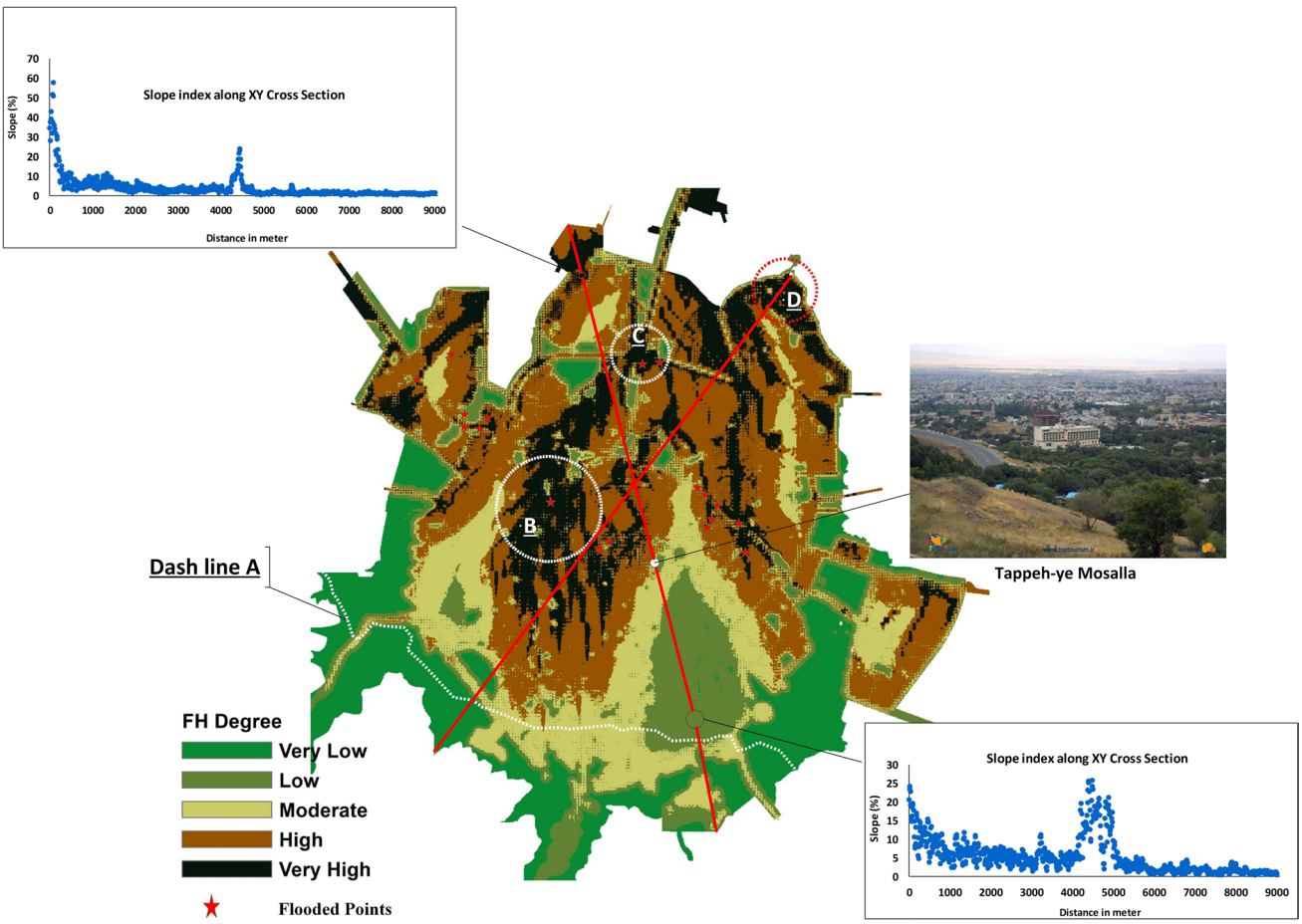


Fig. 6 Reclassification of flood hazard map

the HEC-HMS model investigated the impact of rainwater extract structures on the flood hazard risk of Hamadan. It was also pointed out that the largest proportion of the studied area was in the high-risk class, being related to indices such as low slope, high drainage density and a high ratio of impermeable to permeable surfaces.

The last and most important class of flood hazard risk having 15.83 % of the studied area (11.63 km^2) is related to the class with high flood risk.

To better characterize this degree of flood hazard risk, the three regions B, C, D in this class are discussed. According to Fig. 1, in B located almost in the middle of the study area, it has a high hydrological connectivity (IDPR index) in addition to its passing on the Moradbeik River; this is due to high density of residential areas. In C located in the downstream of B, residential houses have a moderate density. Furthermore, considering this area as the outlet of the communication passages of the study area, roads and routes of communication have a high density leading to an increase in the IDPR index in the area. D is related to a region called Madani town. According to the statistics provided by the Statistical Center of Iran, the region has a population of approximately 36 thousand people. According to Fig. 1, the residential home in this area has a low density; however, due to the fact that the area is the eastern, western and northern arterial highway of the capital of Hamadan Province, it has caused roads and high-density communication routes in the region. Furthermore, the IDPR index is also high in the area.

Therefore, according to the above-mentioned cases, it can be seen that the IDPR index as a criterion of hydrological connectivity plays a significant role in distribution of flood hazard risk.

One of the most important points distinguishing this study from others in the literature review is the study of the importance of the indices in flood hazard risk. In all previous studies conducted in urban basins, the distance to the discharge channels has been considered as the most important index in flood hazard risk. This study, however, indicates that the distance to discharge channel can be deemed as one of the most important parameters in flood hazard risk. Thus, this is consistent with the results obtained by Huang et al. (2009), Uwasu and Yabar (2011) and Tan et al. (2018), studies, stating that in urban basins, according to the increasing degree of urban complexity, the importance of indices decreases in relation to each other.

To check the accuracy of flood hazard mapping, there are different types of methods, among which two methods of superimposition and incorrect and correct percentage rate were used.

In the superimposition method, the most important flood areas in Hamadan, registered by the Municipality of Hamadan, are superimposed on flood hazard zoning map, and the percentage of points in different classes of flood hazard risk is computed. Thus, from the 20 important areas with flood hazard risk in the study area, 14 items were in a class with very high risk, 3 items were in a class with high risk, 2 items were in a class with moderate risk, and 1 was in a class with low risk. If the accuracy of the final map is based on the presence of flooding points in the class with very high risk, then the accuracy of the final map is approximately 70%, indicating the high accuracy of this study in selecting and preparing effective indices in flood hazard risk. Here, it should be noted that the percentage of areas with very high flood hazard risk was low.

Regarding the accuracy check of the flood hazard zoning map, using the ROC curve, the flooding points were superimposed onto a flood hazard zoning map, and to the points that

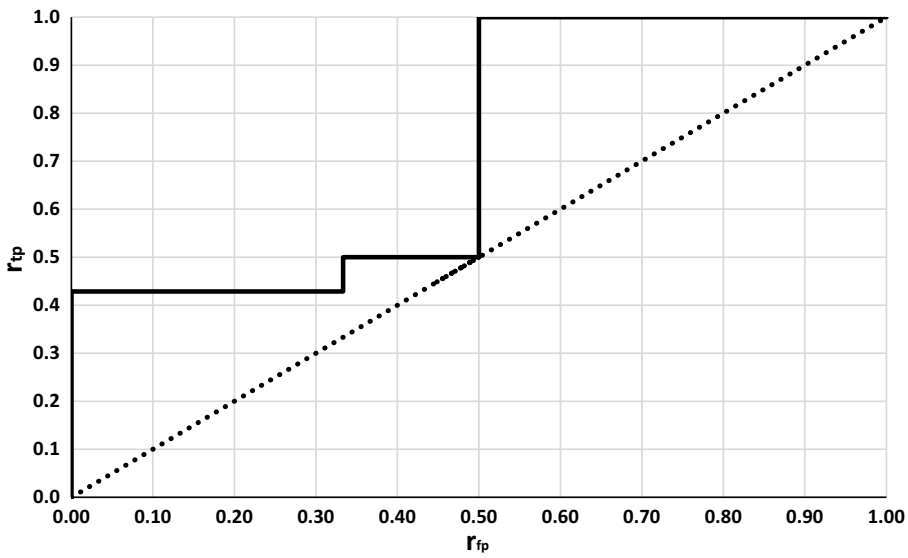


Fig. 7 The ROC curve for a flood hazard based on an entropy model in Hamadan city

were located in high flood risk areas, 1 was allocated, and to the points that were in other flood risk levels, 0 was allocated. After randomly placing the set of numbers, 1 and 0 together with the displacement of the location of cut-points from 1 to 19, the Cartesian chart with right rate and the negative rate was plotted.

According to the results of ROC curve, the AUC value of the study area was 73 %, representing a good evaluation of Shannon's entropy model (Fig. 7).

Conclusion

The main objective of this research was to develop a method identifying areas with flood hazard risk and can be used in other urban areas. This is important for decision making since it creates a roadmap to reduce the consequences of floods. One of the most important points not considered in most studies is the difference in the production of flood hazard risk maps in urban and natural areas. In urban areas, with progress of urbanization, FCFs also expand, causing changes in the land surface, and consequently disturbing the flow paths of runoff.

Therefore, this study comprises four parts to investigate flood hazard mapping in the studied area:

1. Simulation of the surface of the study area:
   The CD-TIN is one of the most important and powerful algorithms that can examine the effect of several features and elements on a surface map. What was introduced in this study was a digital elevation model with the scale of 1/2000, FCFs as an input to the CD-TIN algorithm, and the simulated surface of the study area.
2. Extraction of effective indices in flood hazard risk from the simulated ground surface
   After simulating the surface of the area, five effective indices of flood hazard risks such as slope, elevation, permeability, distance to discharge channel, and IDPR were extracted from the simulated surface.
3. Weighting of indices and preparation of flood hazard risk map
   The last but not least important step is the weighting of the indices. Following the entropy weighing method and considering the important role of the indices in the flood hazard risk, we evaluated and combined, and finally prepared the hazard zoning map.
4. Verification of the accuracy of flood hazard risk map
   In this study, the accuracy of the flood hazard risk map was obtained from using two methods of superimposition and ROC curve. In the superimposition method, by using 20 flood hazard risk observation points, the results indicate that 14 observation points or 70% of the points are located in the area with very high flood risk. Furthermore, in the ROC curve whose basis of accuracy is the surface below the curve, the results show that the surface under the rock curve is approximately 73%, indicating the ability of the model to simulate the flood hazard risk in the study area.

In future researches, other than flood hazard mapping as the first element of flood risk studies, economic aspects and social vulnerability must be also considered. Hence, this approach will be highly important for insurance companies and governments to plan and reduce their financial budget. Moreover, owing to environmental impacts of these incidents, urban decision makers and planners require further data to manage the risks of floods.

References

Bathrellos G, Karymbalis E, Skilodimou H, Gaki-Papanastassiou K, Baltas E (2016) Urban flood hazard assessment in the basin of Athens Metropolitan city. Greece Environ Earth Sci 75:319

Bathrellos GD, Skilodimou HD, Chousianitis K, Youssef AM, Pradhan B (2017) Suitability estimation for urban development using multi-hazard assessment map. Sci Total Environ 575:119–134

Covino T (2017) Hydrologic connectivity as a framework for understanding biogeochemical flux through watersheds and along fluvial networks. Geomorphology 277:133–144

Cox LA (2009) Limitations of risk assessment using risk matrices. In: Cox LA (ed) Risk analysis of complex and uncertain systems. Springer, Berlin, pp 101–124

Degiorgis M, Gnecco G, Gorni S, Roth G, Sanguineti M, Taramasso AC (2012) Classifiers for the detection of flood-prone areas using remote sensed elevation data. J Hydrol 470:302–315

Dysarz T, Wicher-Dysarz J, Sojka M, Jaskuła J (2019) Analysis of extreme flow uncertainty impact on size of flood hazard zones for the Wronki gauge station in the Warta river. Acta Geophysica 67:661–676

Elkhrachy I (2015) Flash flood hazard mapping using satellite images and GIS tools: a case study of Najran City, Kingdom of Saudi Arabia (KSA). Egypt J Remote Sens Space Sci 18:261–278

Fernández D, Lutz M (2010) Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. Eng Geol 111:90–98

Gay A, Cerdan O, Mardhel V, Desmet M (2016) Application of an index of sediment connectivity in a lowland area. J Soils Sedim 16:280–293

Ghiglieri G, Carletti A, Pittalis D (2014) Runoff coefficient and average yearly natural aquifer recharge assessment by physiography-based indirect methods for the island of Sardinia (Italy) and its NW area (Nurra). J Hydrol 519:1779–1791

Gigović L, Pamučar D, Bajić Z, Drobnjak S (2017) Application of GIS-interval rough AHP methodology for flood hazard mapping in urban areas. Water 9:360

Guo E, Zhang J, Ren X, Zhang Q, Sun Z (2014) Integrated risk assessment of flood disaster based on improved set pair analysis and the variable fuzzy set theory in central Liaoning Province, China. Nat Hazards 74:947–965

Hall M, Ellis J (1985) Water quality problems of urban areas. GeoJournal 11:265–275

Heckmann T, et al (2015) Indices of hydrological and sediment connectivity-state of the art and way forward. In: EGU general assembly conference abstracts

Hernandes TAD, Scarpare FV, Seabra JEA (2018) Assessment of the recent land use change dynamics related to sugarcane expansion and

the associated effects on water resources availability. J Clean Prod 197:1328–1341
Huang S-L, Yeh C-T, Budd WW, Chen L-L (2009) A sensitivity model (SM) approach to analyze urban development in Taiwan based on sustainability indicators. Environ Impact Assess Rev 29:116–125
Kawachi T, Maruyama T, Singh VP (2001) Rainfall entropy for delineation of water resources zones in Japan. J Hydrol 246:36–44
Kazakis N, Kougias I, Patsialis T (2015) Assessment of flood hazard areas at a regional scale using an index-based approach and analytical hierarchy process: application in Rhodope-Evros region, Greece. Sci Total Environ 538:555–563
Kowalzig J (2008) Climate, Poverty, and Justice: What the Poznan UN climate conference needs to deliver for a fair and effective global deal. Oxfam Policy Pract Clim Change Resil 4:117–148
Lee S, Lee S, Lee M-J, Jung H-S (2018) Spatial assessment of urban flood susceptibility using data mining and geographic information system (GIS) tools. Sustainability 10:648
Li Q (2013) Fuzzy approach to analysis of flood risk based on variable fuzzy sets and improved information diffusion methods. Nat Hazards Earth Syst Sci 13(2):239–249
Li Z, Wu L, Zhu W, Hou M, Yang Y, Zheng J (2014) A new method for urban storm flood inundation simulation with fine CD-TIN surface. Water 6:1151
Liu L, Zhou J, An X, Zhang Y, Yang L (2010) Using fuzzy theory and information entropy for water quality assessment in Three Gorges region, China. Expert Syst Appl 37:2517–2521
Lohani AK, Goel N, Bhatia K (2014) Improving real time flood forecasting using fuzzy inference system. J Hydrol 509:25–41
Lotfi FH, Fallahnejad R (2010) Imprecise Shannon's entropy and multi attribute decision making. Entropy 12:53–62
Lyu H-M, Shen S-L, Zhou A, Yang J (2019) Perspectives for flood risk assessment and management for mega-city metro system. Tunn Undergr Space Technol 84:31–44
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 Manag 25:1359–1386
Mahmoud SH, Gan TY (2018) Multi-criteria approach to develop flood susceptibility maps in arid regions of Middle East. J Clean Prod 196:216–229
Malczewski J (1999) GIS and multicriteria decision analysis. Wiley, Hoboken
Mayor ÁG, Bautista S, Small EE, Dixon M, Bellot J (2008) Measurement of the connectivity of runoff source areas as determined by vegetation pattern and topography: a tool for assessing potential water and soil losses in drylands. Water Resour Res. https://doi.org/10.1029/2007WR006367
Mishra AK, Özger M, Singh VP (2009) An entropy-based investigation into the variability of precipitation. J Hydrol 370:139–154
Patra S, Mishra P, Mahapatra SC (2018) Delineation of groundwater potential zone for sustainable development: A case study from Ganga Alluvial Plain covering Hooghly district of India using remote sensing, geographic information system and analytic hierarchy process. J Clean Prod 172:2485–2502
Pradhan B (2009) Flood Susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. J Spatial Hydrol 9:1–18
Radwan F, Alazba A, Mossad A (2019) Flood risk assessment and mapping using AHP in arid and semiarid regions. Acta Geophysica 67:215–229
Rodríguez LR, Nouvel R, Duminil E, Eicker U (2017) Setting intelligent city tiling strategies for urban shading simulations. Sol Energy 157:880–894
Samanta S, Koloa C, Kumar Pal D, Palsamanta B (2016) Flood risk analysis in lower part of Markham river based on multi-criteria decision approach (MCDA). Hydrology 3:29
Schanze J (2006) Flood risk management–a basic framework. In: Schanze J (ed) Flood risk management: hazards, vulnerability and mitigation measures. Springer, Berlin, pp 1–20
Sepehri M, Malekinezhad H, Ilderomi AR, Talebi A, Hosseini SZ (2018) Studying the effect of rain water harvesting from roof surfaces on runoff and household consumption reduction. Sustain Cities Soc 43:317–324. https://doi.org/10.1016/j.scs.2018.09.005
Sepehri M, Malekinezhad H, Hosseini SZ, Ildoromi AR (2019) Suburban flood hazard mapping in Hamadan city, Iran. In: Proceedings of the institution of civil engineers-municipal engineer. Thomas Telford Ltd, pp 1–13
Shadman Roodposhti M, Aryal J, Shahabi H, Safarrad T (2016) Fuzzy shannon entropy: a hybrid GIS-based landslide susceptibility mapping method. Entropy 18:343
Singh V (1997) The use of entropy in hydrology and water resources. Hydrol Process 11:587–626
Sivapalan M (2003) Prediction in ungauged basins: a grand challenge for theoretical hydrology. Hydrol Process 17:3163–3170
Skilodimou HD, Bathrellos GD, Chousianitis K, Youssef AM, Pradhan B (2019) Multi-hazard assessment modeling via multi-criteria analysis and GIS: a case study. Environ Earth Sci 78:47
Smithson M (1989) Cognitive science Ignorance and uncertainty: emerging paradigms. Springer, New York. https://doi.org/10.1007/978-1-4612-3628-3
Stefanidis S, Stathis D (2013) Assessment of flood hazard based on natural and anthropogenic factors using analytic hierarchy process (AHP). Nat Hazards 68:569–585
Tan Y, Jiao L, Shuai C, Shen L (2018) A system dynamics model for simulating urban sustainability performance: a China case study. J Clean Prod 199:1107–1115
Tehrany MS, Pradhan B, Jebur MN (2013) Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. J Hydrol 504:69–79
Thapa R, Gupta S, Reddy D (2017) Application of geospatial modelling technique in delineation of fluoride contamination zones within Dwarka Basin, Birbhum, India. Geosci Front 8:1105–1114
Unal I (2017) Defining an optimal cut-point value in roc analysis: an alternative approach. Comput Math Methods Med 2017:3762651. https://doi.org/10.1155/2017/3762651
USDA (1986) Urban hydrology for small watersheds. Tech Release 55:2–6
Uwasu M, Yabar H (2011) Assessment of sustainable development based on the capital approach. Ecol Indic 11:348–352
Wang Z, Lai C, Chen X, Yang B, Zhao S, Bai X (2015) Flood hazard risk assessment model based on random forest. J Hydrol 527:1130–1141
Wang M, Zhang DQ, Su J, Dong JW, Tan SK (2018) Assessing hydrological effects and performance of low impact development practices based on future scenarios modeling. J Clean Prod 179:12–23
Xu H, Ma C, Lian J, Xu K, Chaima E (2018) Urban flooding risk assessment based on an integrated k-means cluster algorithm and improved entropy weight method in the region of Haikou, China. J Hydrol 563:975–986. https://doi.org/10.1016/j.jhydrol.2018.06.060
Yang X-l, Ding J-h, Hou H (2013) Application of a triangular fuzzy AHP approach for flood risk evaluation and response measures analysis. Nat Hazards 68:657–674
Zou Q, Zhou J, Zhou C, Song L, Guo J (2013) Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy AHP. Stoch Environ Res Risk Assess 27:525–546

Acta Geophysica (2019) 67:1451–1469
https://doi.org/10.1007/s11600-019-00323-0

RESEARCH ARTICLE - HYDROLOGY

Development of a relationship between hydrometric and hydrographic observations to predict reservoir capacity loss

Y. C. Jabbar[1] · S. M. Yadav[1]

Received: 24 September 2018 / Accepted: 4 July 2019 / Published online: 17 July 2019

Abstract

Accuracy of reservoir capacity loss estimation on daily timescale is dependent on the certainty of sediment load prediction, density estimate and capacity observed by consecutive hydrographic surveys. Data-scarce and uncertain data conditions restrict the development of a relationship between hydrographic surveys and hydrometric observations. The present study has been carried for Ukai Reservoir, India. A novel sediment rating curve fitting approach by optimization technique has been proposed in order to accurately predict sediment load from low-frequency sampled discharge and sediment concentration observations. The study demonstrates the validation of the bulk density estimate using statistical hypothesis testing and identifies the correctness of the hydrographic survey results. Application of the developed hydrometric and hydrographic relationship indicated that about 50% of the capacity loss of a year might occur during a single extreme event. The proposed approach can serve as a decision support system to monitor and manage sedimentation for the reservoir having uncertain data conditions.

Keywords Hydrometric observations · Hydrographic survey · Sediment rating curve · Representative sediment density · Reservoir capacity loss

Introduction

Rivers are the carrier channels for both water and sediment transported by the flow. Impoundment of water in the reservoir is a prime objective, but in doing so silting of sediments carried with the flow needs to be checked. Understanding that reservoirs are non-renewable resources drives the research on their capacity loss (Kondolf et al. 2014). Initially, defining the rate of storage loss over a period of the dam's operation was the sole interest. However, of late, attention is also being paid to augment the life of the reservoirs despite the sediment inflow experienced by them (Chaudhuri 2006; Kondolf et al. 2014; Palmieri et al. 2001; Sumi and Hirose 2009). For achieving this goal, the inflowing sediment load has to be associated with reservoir capacity loss (RCL). Marineau and Wright (2017) quoted that a model that can relate the hydrological history to the reservoir sedimentation rates, at shorter timescales, can give precise estimates of the economic life of the reservoir.

The transportation and deposition of sediments in the reservoir can be studied by hydrometric observations and hydrographic surveys. Capacity loss noticed between two hydrographic surveys can be related to the suspended sediment concentration (SSC) observed at the hydrometric gauging station (Marineau and Wright 2017; Tebbi et al. 2012; Verstraeten and Poesen 2002); however, there are multiple practical limitations to it (Salas and Shin 1999).

If the bed load is unmeasured, then the use of empirical formulations and approximations may bring uncertainty to the predicted total sediment load (Swamee and Ojha 1991; Vanoni 1979). A low-frequency sampling (once a day) of the discharge and the SSC may generate an inaccurate sediment load estimate (Arabkhedri et al. 2010; Bussi et al. 2017; Harrington and Harrington 2013; Walling 1977a, b). In other words, hydrological sampling plan based on rising and falling limb of the hydrograph requires a couple of samples per day, while sampling once a day (time-based sampling) will bring inaccuracy in sediment load estimation. Statistically, a fixed time-based sampling plan to measure SSC will limit the sensitivity of RCL analysis, as

✉ Y. C. Jabbar
yazad.jabbar@gmail.com

[1] Department of Civil Engineering, S. V. National Institute of Technology, Surat, Gujarat 395007, India

the variation in the hydrograph and sediment graph is not captured adequately. In such instances, a sediment rating curve (SRC) can predict the continuous record of concentration to estimate the sediment load. Yet, a stable SRC relationship can only be developed from high-frequency sampled data, (Bussi et al. 2017; Crawford 1991; Horowitz 2003) and in its absence, different grouping and fitting procedures can be tested to identify a profound rating relationship. SRC can be obtained by grouping the data based on either time or the stage of the river (Walling 1977a, b). Seasonal variations in the SSC can be detected by time-based grouping, while grouping based on the stage of the river (rising and falling limb of the hydrograph) incorporates the fluctuations resulting from the hysteresis effect (De Girolamo et al. 2015).

The power form of relationship, as an SRC model, relates SSC and discharge by means of nonlinear curve fitting. A nonlinear (power) form of relationship can be converted to a linear form by logarithmic data transfer (Heidarnejad et al. 2006). Thus, ordinary least square (OLS) regression technique can be used to fit the linearized relationship of SSC and discharge. However, regression faces a limitation since the fitted model will possess the least square of residuals only for the concentration. Besides, regression between the suspended sediment load and discharge as an alternative to SSC and discharge is also considered to be a wrong practice. The suspended sediment load includes discharge in its computation and generates a nonexistent superficial correlation (Annandale et al. 2016). Hence, to achieve the minimum difference of load estimates and to fit the suspended SRC such that the value of the coefficient of determination for SRC remains high, an optimization model approach is developed which further calibrates the SRC.

In the absence of the observed trap efficiency (TE), its estimation using empirical equations and curves (Brune 1953) may cause uncertainty in the computation of reservoir-deposited sediments. The bulk density of the complete reservoir is hard to identify as no theoretical base has been established to obtain the reservoir density from the sample-point densities. Bussi et al. (2013) found the predictability of the Lane and Koelzer (1943) empirical formulation to be satisfactory. For a small check dam, the authors validated the predicted dry bulk density using five sampled measurements. Yet, the complex distribution of sediments over the reservoir and neglecting the organic matter content compromised its validity (Verstraeten and Poesen 2001). Small et al. (2003) collected over 30 sediment samples from the Crombie Reservoir to determine the wet and dry densities. The researchers documented that the basal region had high density values (up to 2200 kg/m^3) while the surface region had low density (500 kg/m^3). Thus, the bulk density sampling might not give a representative value for the whole deposit and is usually estimated and not sampled. Tebbi et al. (2012) estimated a typical value of dry bulk density (1400 kg/m^3) by considering the composition of the sediments.

The objective of this paper is to predict RCL (on a daily time step) from the hydrometric observations using low-frequency (once a day) suspended sediment sampled data. The impact of the grouping and fitting procedures on the SRC for estimating the sediment load has been assessed, and a novel SRC model fitting approach by optimization technique has been proposed. A matrix of RCL is computed, using the SRC models and bulk density estimates, to assess the uncertainty of the predictions.

Study area and data collection

Ukai Reservoir, India, is chosen for the study, as it is equipped with an upstream gauging station having a long period of SSC record, reservoir sediment sampling data are available (for density estimate) and multiple hydrographic surveys are carried out on the reservoir. Ukai Dam reservoir lies in the middle of the Tapi basin (Fig. 1). Tapi River originates near Multai, Betul district, Madhya Pradesh, India, at an elevation of 752 m and drains in the Arabian Sea. The length of the Tapi River is 724 km; Purna (length of 274 km) and Girna (length of 260 km) are two major tributaries of the river. Tapi basin lies between 72°33′ and 78°17′ east longitudes and 20°9′ to 21°50′ north latitudes. The basin is surrounded by the Satpura Range (from the north), Mahadev Hills (from the east) and Ajanta Range (from the south). Tapi basin is covered with agriculture, forest and water bodies by 66.19%, 25% and 2.99%, respectively, of the total area. The basin majorly consists of black cotton soil. The annual rainfall in the Tapi basin is 830 mm.

The first impoundment of the reservoir occurred in 1972. The gross reservoir capacity at the time of the first impoundment was 8510×10^6 m^3. The full reservoir level (FRL) is at 105.15 m from the mean sea level, and its water spread area is 520 km^2. The catchment area up to the dam wall is 62,225 km^2. As per the last hydrographic survey report (2003), the total loss observed in reservoir capacity is 1095.71×10^6 m^3, the existing storage capacity is reduced to 7414.29×10^6 m^3 and the distribution of loss is 51% in dead live and 49% in live storage.

Sarangkheda gauging station (21°25′55″N, 74°31′37″E) is the nearest upstream suspended sediment and discharge gauging station of the Central Water Commission (CWC). The discharge and sediment concentration data are available from 1984 at the gauging station. Discharge is measured at the station gauge line by the velocity-area method. The depth of flow in the cross section is measured at verticals of the segmented station gauge line, and observations of velocity are obtained with a current meter at 0.6 m depth point. The widths, depths and velocities observed

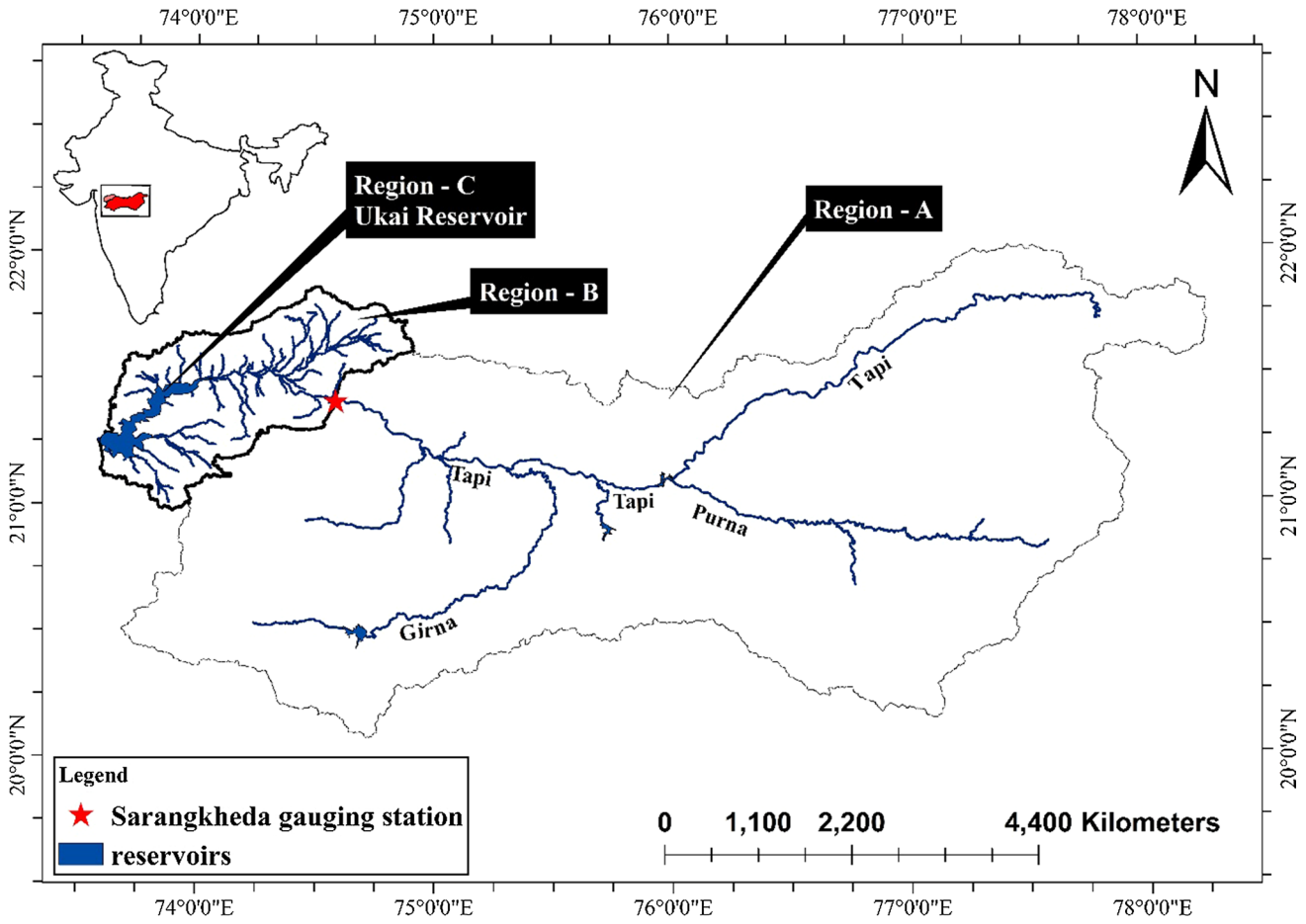


Fig. 1 Location of Ukai Reservoir in Tapi basin

are used to compute discharge for each segment of the cross sections. Summation of these segmental discharges is the total discharge observed at the station gauge line. The Punjab-type bottle sampler is used to collect suspended sediment at 0.6 m depth. The frequency of discharge and sediment measurement at station gauge is once per day, whereas the water surface elevation is measured every hour (CWC 2014).

Sedimentation surveys (hydrographic surveys) of the Ukai Reservoir were conducted in 1979, 1983, 1992, 2001 and 2003. A schematic representation of the Tapi basin up to the Ukai Dam is given in Fig. 2. Region A (58,400 m^2)

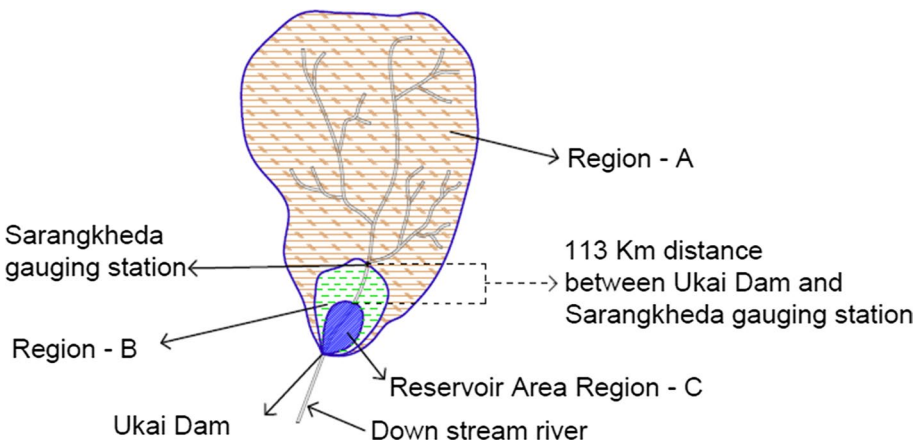


Fig. 2 Schematic representation of the reservoir its upstream gauging station and ungauged catchment

is the Tapi basin area up to the Sarangkheda gauging station. The ungauged region B (3825 m^2) is the area of the Tapi basin between the Ukai Dam and the gauging station. The distance between the reservoir's tail and Sarangkheda is 113 km. The reservoir area at its full level (105.15 m) is 520 m^2, as depicted in region C (Fig. 2). The mean of daily discharge and sediment concentration observed at the gauging station (period 1984–1992) is 276 m^3 and 0.45 kg/m^3, respectively, while the maximum value of discharge and sediment concentration observed at the gauging station (period 1984–1992) is 13,750 m^3 and 17.09 kg/m^3, respectively. Grain size analysis of the deposited sediment in the Ukai Reservoir at varying distances from the dam wall is provided in Table 1.

Methodology

Quantitative RCL is computed by converting the reservoir-inflowing sediment load to the deposited sediment volume. Sediment load inflowing the reservoir can be predicted using a stable SRC, which can be then converted to deposition volume using the estimated reservoir sediment density. The predictability of the capacity loss from different SRC models and density estimates is checked in three phases (Fig. 3). In the first phase, emphasis has been made for the precise prediction of the reservoir-inflowing sediment load. The second phase of the work deals with the estimation of sediment bulk density for the conversion of the inflowing sediment load to deposition volume. Four data grouping approaches and two fitting procedures are used to develop eight types of rating curve relationships. The predictions of these rating curves are subjected to three density estimates.

Table 1 Grain size analysis of Ukai Dam Reservoir

Sample no.	P_C^a	P_M^b	P_s^c
1	0.36	0.56	0.08
2	0.35	0.59	0.06
3	0.37	0.56	0.07
4	0.34	0.57	0.09
5	0.39	0.56	0.05
6	0.36	0.56	0.08
7	0.00	0.02	0.98
8	0.35	0.58	0.07
9	0.35	0.55	0.08
10	0.34	0.56	0.10
Average	0.32	0.51	0.17

The data were abstracted from the publication of the CWC (2015)

[a]P_C, Percentage of clay/100 (particle size less than 0.002×10^{-3} m), [b]P_M, Percentage of silt/100 (particle size 0.002 to 0.075×10^{-3} m), [c]P_S, percentage of sand/100 (particle size 0.075 to 4.75×10^{-3} m)

It is to be noted from phase three that if the RCL predicted is not equivalent to observed RCL, the best sediment load-estimating model (Phase 1) is to be used to fill the data gaps and inconsistencies. Then, an assessment of the estimated bulk density should be done to correct the predicted RCL. From the above process, the established temporal-lumped relationship is utilized to disintegrate the capacity loss on daily time step.

Computation of daily RCL is done from the daily trapped inflowing load (predicted using the SRC model) and estimated sediment densities. Gross reservoir capacity at the end of the period is obtained by deducting the inflowing sediment volume from the gross capacity at the beginning of the period (Eq. 1):

$$\mathrm{GRC}_t = \mathrm{GRC}_{t-1} - \left[\frac{L_t * \mathrm{TE}}{\mathrm{BD} * 10^6}\right] \quad (1)$$

where GRC_{t-1} = gross reservoir capacity at the beginning of period ($10^6\times m^3$), GRC_t = gross reservoir capacity at the end of period ($10^6\times m^3$), L_t = predicted sediment load inflowing the reservoir during period t (Kg), TE = trap efficiency of the reservoir, BD = bulk reservoir density of the reservoir (kg/m^3).

SRC model development

The gauged data were observed to be inconsistent with data gaps in the SSC measurements. In such a situation, the missing data were filled using suspended SRC (Walling 1977a, b). The available instantaneous daily time-stepped data may not reveal the hysteresis effect. Thus, in the present study, the SRCs are developed using time-based grouping, that is, daily, monthly average and yearly average data. Data grouping was carried out to reconnoiter a relationship between discharge and concentration, while the application of the developed relationship was utilized to predict the concentration on a daily time step. Fitting of the rating curve to the data was achieved by two methods, namely ordinary least square (OLS) linear regression and fitting using optimization.

Data grouping

For developing the daily SRC model, directly available data were utilized. The monthly model was developed by grouping the data for the months in different years and calculating the average value of discharge and concentration for each month's group in a year. A yearly average model was developed by categorizing the data according to the calendar year and then establishing the average value of discharge and concentration. Month-wise models were obtained by clustering the daily data of all the similar months.

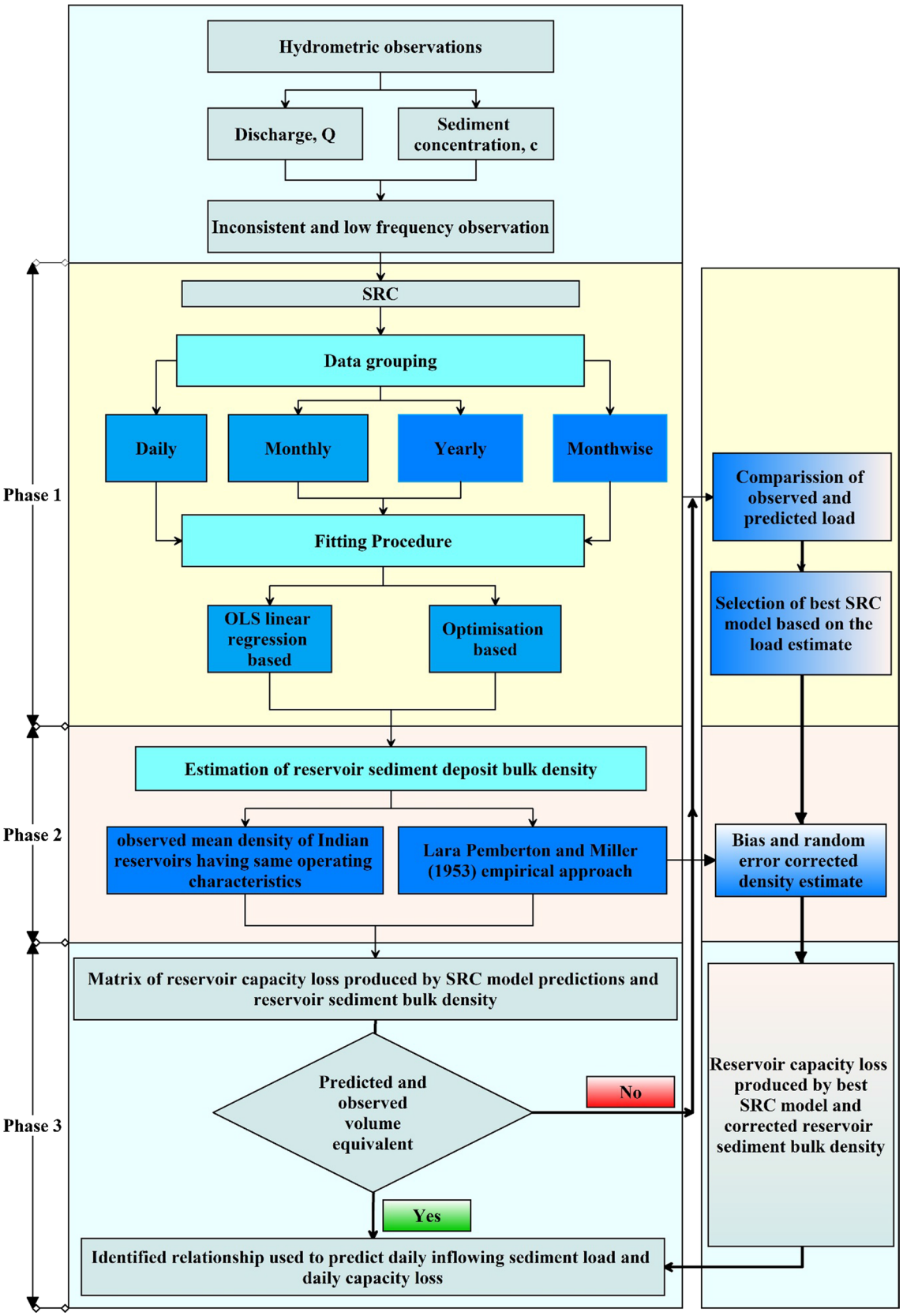


Fig. 3 Methodology adopted for identification of reservoir capacity loss

Curve fitting by OLS regression and proposed optimization

After grouping, the conversion of the values to logarithms of base ten was performed for transforming the model from the power form (Eq. 2) to the linear form (Eq. 3). The scaling coefficient, a, is transferred to the intercept ($\log a$), and the exponent, b, becomes the slope of the transferred linear relationship:

$$c = aQ^b \tag{2}$$

$$\text{Log}\, c = b * (\log Q) + (\log a) \tag{3}$$

Fitting a curve to the data is done by two methods, viz. ordinary least square (OLS) regression (linear fitting) and optimization technique. Both the fitting procedures are optimization techniques that minimize a particular statistical error function (objective) with respect to certain constraints. OLS linear regression is derived from calculus while optimization obtains solution numerically. The OLS regression would give a good relationship between discharge and concentration but may not necessarily produce a good sediment load estimate. On the other hand, the correlation between suspended sediment load and discharge is of interest; however, sediment load includes discharge in its computation and regression between them may generate a nonexistent superficial correlation (Annandale et al. 2016). The novel approach proposed in the present research is to calibrate the SRC coefficients in such a way that no significant change is observed in the coefficient of determination (obtained from OLS regression fitting procedure) between discharge and SSC relationship. Yet, the predictive accuracy of the accumulated suspended sediment load is increased. That is, the SRC models developed using OLS regression are further calibrated to have a minimum difference in load estimates. In order to fit the SRC models by optimization, the coefficient of determination (obtained using the OLS regression) between discharge and SSC relationship is considered as a benchmark. The lower bound of the coefficient of determination (for the SRC model fitted by optimization technique) is selected as 5% lower than the benchmark value, while the upper bound for the coefficient of determination is given as one. Similarly, the upper and lower bounds of the scaling coefficient, a, and the exponent, b were selected such that the range of bound was maintained between ± 5% of the SRC coefficients obtained using the OLS regression (Eq. 3). The objective of the daily, monthly average and yearly average SRC models fitted by optimization was to minimize the absolute percentage error in the suspended sediment load accumulated over a period (1984–1992) and is expressed mathematically in Eq. 4. On the other hand, the objective of the month-wise SRC was the minimization of the absolute percentage error in the load accumulated over the period of analysis for the respective month only:

$$\text{Percentage error} = \frac{L_\text{p} - L_\text{o}}{L_\text{o}} * 100 \tag{4}$$

The observed (L_o) and predicted (L_p) sediment loads (ton/day) were obtained using Eqs. 5 and 6, respectively, by summing the product of the SSC, c_i, with the daily mean discharge, Q_i, for the study period (n days):

$$L_\text{o} = \sum_{i=1}^{n} \left(c_{i\,\text{observed}} * Q_i * \frac{24 * 60 * 60}{1000} \right) \tag{5}$$

$$L_\text{p} = \sum_{i=1}^{n} \left(c_{i\,\text{predicted}} * Q_i * \frac{24 * 60 * 60}{1000} \right) \tag{6}$$

where average observed discharge Q_i (m^3/s) of a particular day is obtained from the gauge-discharge curve and the water surface elevation (observed every hour). The predicted concentration $c_{i\,\text{predicted}}$ (kg/m^3) is obtained as per Eq. 2, and observed concentration $c_{i\,\text{observed}}$ (kg/m^3) is the observed SSC for a given day. It is to be noted that observed sediment concentration does not represent the average sediment load of the day but is a single temporal-point measurement. Sediment transport rate throughout a day is majorly dependent on the variation in precipitation. Hence, the measured concentration value is not aligned with the hydraulic and hydrological factors. Using such measurement brings intrinsic uncertainty in the load prediction (Singh et al. 2013).

Optimization of the objective function for the constraints is done utilizing Excel Solver optimization tool. The solver is a spreadsheet-based optimization tool that provides nonlinear generalized reduced gradient (GRG) and evolutionary method to optimize nonlinear problems. The GRG method starts with an initial solution. It looks at the gradient of the absolute percentage error (Eq. 4) as the SRC coefficients are changed and stops as the first derivatives equal zero. The evolutionary method starts with random values of the coefficients (parent population) and evaluates them by a fitness function (Eq. 4). The population is mutated, a new set of SRC coefficients are created as offspring, the individual fitness of each SRC coefficient is evaluated and the least fit is replaced with new values.

Using both methods one after another, it is assured that a global optimal solution of the SRC coefficients is reached. In GRG method and evolutionary method, the convergence is given as 0.0001. In the evolutionary method, the population size is given as 100 and the mutation rate is given as 0.075. The limiting time bound in the evolutionary method is selected as 30 s; i.e., the optimization process is to be

stopped if maximum time without improvement of the solution is more than 30 s.

The statistical function of the Nash–Sutcliffe model efficiency factor *NSMEF* (Nash and Sutcliffe 1970) and the index of agreement *d* (Willmott 1981) signify the model efficiency and are used for comparing the observed and predicted suspended sediment loads. The *NSMEF* scores range from negative infinity to one. A value equal to or less than zero denotes that the developed model should be disregarded, while the value of unity indicates a perfect prediction. The index of agreement *d* is a non-dimensional and bounded measure. It is bounded from zero to one, with one suggesting a perfect match and zero connoting a lack of match between the observed and predicted values.

Computation of the sediment volume

Analyzing the observed SSC and discharge data obtained on daily time step (period 1983–1992), it has been found that about 33% of SSC data were not measured but can be predicted by SRC. By applying the SRC, the spatial and temporal data gaps were filled and the time series of daily suspended sediment load was obtained. The daily suspended sediment load records were summed up for the period between the two consecutive hydrographic surveys. The total load cannot be estimated if the bed load is not included in the predicted suspended load. The bed load was not measured during the study period; therefore, considering the grain size distribution of the sediment, it was assumed to be 20% of the measured load (BIS-12182 1987; Waikhom and Yadav 2017). Besides, the calculated sediment load was adjusted as per the reservoir TE, which was estimated from the Brune (1953) median curve. The mathematical relationship of Brune (1953) median curve (Eq. 7) proposed by Garg and Jothiprakash (2008) was employed:

$$\mathrm{TE} = \frac{\frac{C_o}{I}}{0.00013 + 0.01\left(\frac{C_o}{I}\right) + 0.0000166\sqrt{\frac{C_o}{I}}} \tag{7}$$

where TE = trap efficiency (%), C_o = storage capacity of the reservoir (10^6 m^3), I = inflow of water in the reservoir (10^6 m^3).

The total load (kg) can be converted to volume (m^3) based on the deposited sediment bulk density (hereafter merely referred to as density), which can be acquired by sampling the deposited sediment or by using empirical formulae. In the present analysis, sediment sampling was not a feasible approach due to the depth of the reservoir. Furthermore, point densities recorded in different locations need to be converted into representative density (Annandale et al. 2016). The sediment volume was computed from the density based on the Lara and Pemberton (1963) and Miller (1953) (i.e., empirical approach), the observed mean density of Indian reservoirs and the typical value of density as per Tebbi et al. (2012).

Reservoir-submerged sediment density

Lara and Pemberton (1963) and Miller (1953) empirical approach is used for the estimation of the submerged sediment bulk density. The initial density was calculated using the Lara–Pemberton method (Strand and Pemberton 1982) as shown in Eq. 8:

$$W_\mathrm{i} = W_\mathrm{c}P_\mathrm{c} + W_\mathrm{m}P_\mathrm{m} + W_\mathrm{s}P_\mathrm{s} \tag{8}$$

W_i = density in kg/m^3, P_c, P_m and P_s = clay, silt and sand percentages of the incoming sediment, respectively. W_c, W_m and W_S = clay, silt and sand coefficients of the incoming sediment, respectively.

Miller's (1953) approach was applied to determine the average sediment density deposited in T years of the reservoir's operation, as provided in Eq. 9:

$$W_T = W_i + 0.4343K\left[\frac{T}{T-1}\left(\log_e T\right) - 1\right] \tag{9}$$

W_T = average density in kg/m^3, after T years of reservoir operation, W_i = initial density in kg/m^3, as derived from Eq. 8, K = constant, based on compacting characteristics of sediment and reservoir operation.

The value of K relates the compacting characteristics of the sediment based on the sediment size analysis. The bulk densities obtained from the Lara and Pemberton (1963) and Miller (1953) (cited in Strand and Pemberton 1982) approach depend on the grain size analysis.

In addition to the density computed by Lara and Pemberton (1963) and Miller (1953) approach, the typical value of density (1400 kg/m^3; Tebbi et al. 2012) and the mean of observed sediment densities (i.e., 1191 kg/m^3; CWC 2015) were utilized for the computation of sediment volume. Natural variability of the arithmetic mean sediment densities observed from 21 Indian reservoirs ranged from 780 to 1555 kg/m^3.

Results and discussion

SRC models

An SRC can represent flow and sediment transport relationship for a location under a certain range of environmental, climatic and land use conditions. In other words, the relationship of sediment load and discharge together should be consistent during the period of analysis

(Asselman 2000; Warrick 2015). Before the development of SRCs, trend analysis was performed by nonparametric tests, which showed that discharges and suspended sediments have trend during the study period. However, the obtained trend between observed discharges is similar to the trend in observed concentrations; i.e., both the discharge and SSC are found consistent with each other. Hence, SRCs by different data grouping and curve fitting procedures were developed at Sarangkheda gauging station for the period 1984–1992.

SRC developed by OLS regression

Figure 4 illustrates the SRC fitted with daily, monthly and yearly groups of data utilizing the OLS regression fitting procedure. The month-wise data-grouped SRC models fitted by OLS regression are illustrated in Figs. 5 and 6. It is inferred from Figs. 5 and 6 that only the monsoon months' discharge and SSC are correlated, while the non-monsoon months' data exhibit no connection whatsoever. Statistical significance of the log of discharge and concentration data used in the regression model was checked by hypothesis testing (*P* values). It was found that all the SRCs produced in Figs. 4, 5 and 6 are statistically significant except the month-wise SRCs of the months February, April, May and

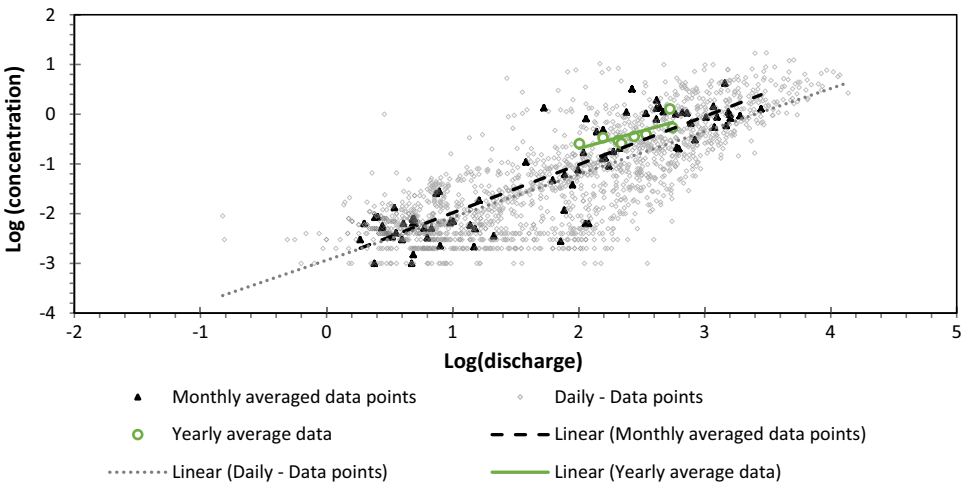


Fig. 4 SRC fitted using OLS regression

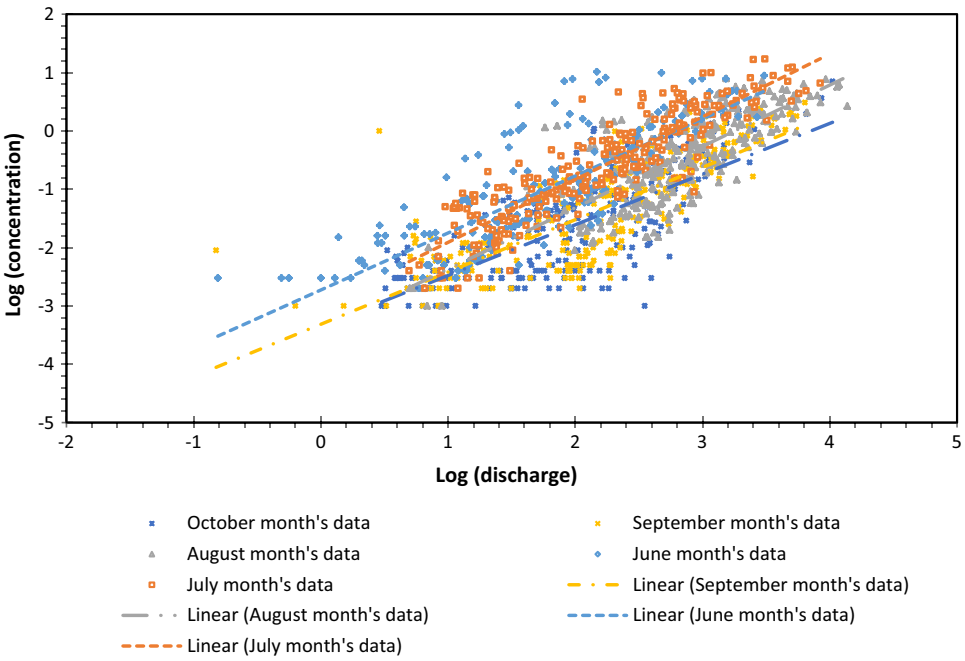


Fig. 5 SRC fitted using OLS regression for monsoon months (month-wise SRC)

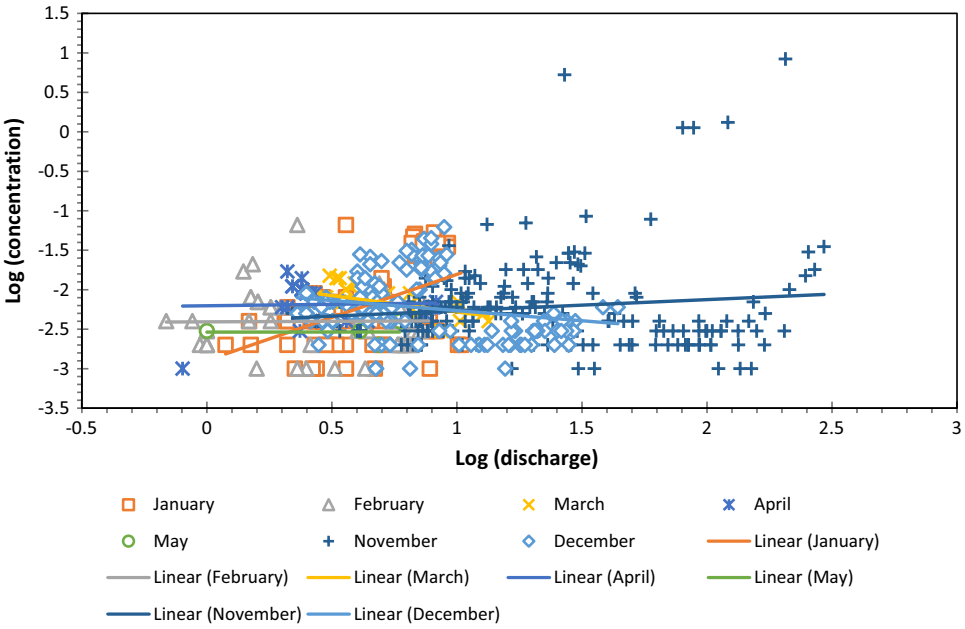


Fig. 6 SRC fitted using OLS regression for non-monsoon months (month-wise SRC)

November. It is to be noted that during these months, very lean flow is observed and the sediment load contribution is less than 0.2%. Scaling (intercept) and exponent (slope) coefficients along with the coefficient of determination for the SRCs which are shown in Figs. 4, 5 and 6 are presented in Table 2. By conversion of the intercept [Log(a)] (presented in Table 2) to scaling coefficient (a), it is understood that for one unit of discharge, the concentration (kg/m^3) will be its thousandth part.

SRC developed by optimization

The information of the relationship obtained between discharge and SSC was utilized to fit SRC models by optimization. Selected bounds for the SRC models to be fitted by optimization technique are listed in Table 3. The coefficient of determination (r^2) between SSC and discharge was considered as one of the constraining conditions, and the lower limit of r^2 was provided. The curves obtained for daily, monthly and yearly data by optimization are presented in Fig. 7. For the month-wise SRC model fitting by optimization, only the monsoon months were considered, as they exhibited a good correlation between the discharge and concentration. Fitted month-wise SRCs are portrayed in Fig. 8. By considering the objective function as per Eq. 4 along with the constraints as mentioned in Table 3, SRC fitted by the optimization technique is presented in Table 4.

Table 2 SRC models developed using OLS regression

SRC model		Log (a)	b	Coefficient of determination, R^2
Group of data	No. of data points			
Daily data-based SRC model	1889	−2.9349	0.8628	0.6476
Monthly averaged data-based SRC model	76	−2.9542	0.9701	0.7801
Yearly averaged data-based SRC model	9	−2.1072	0.7076	0.6122
Month-wise data-based month-wise SRC models				
June	217	−2.9349	0.8628	0.6476
July	260	−2.9862	1.0757	0.7900
August	279	−3.4197	1.0514	0.5947
September	268	−3.3142	0.8962	0.6430
October	240	−3.3467	0.8688	0.5916

Here, a, scaling coefficient of the sediment rating curve [(kg/m^3)/(m^3/s)]; b, exponent coefficient of rating curve (unitless)

Table 3 Range of upper and lower bounds

Group of data	Coefficient log (*a*)		Coefficient *b*		Coefficient of determination, R^2	
	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound
Daily data-based SRC model	−2.79	−3.08	0.91	0.82	1	0.62
Monthly averaged data-based SRC model	−2.81	−3.10	1.02	0.92	1	0.74
Yearly averaged data-based SRC model	−2.00	−2.81	0.74	0.67	1	0.58
Month-wise data-based month-wise SRC models						
June	−2.79	−3.08	0.91	0.82	1	0.62
July	−2.84	−3.14	1.13	1.02	1	0.75
August	−3.25	−3.59	1.10	1.00	1	0.56
September	−3.15	−3.48	0.94	0.85	1	0.61
October	−3.18	−3.51	0.91	0.83	1	0.56

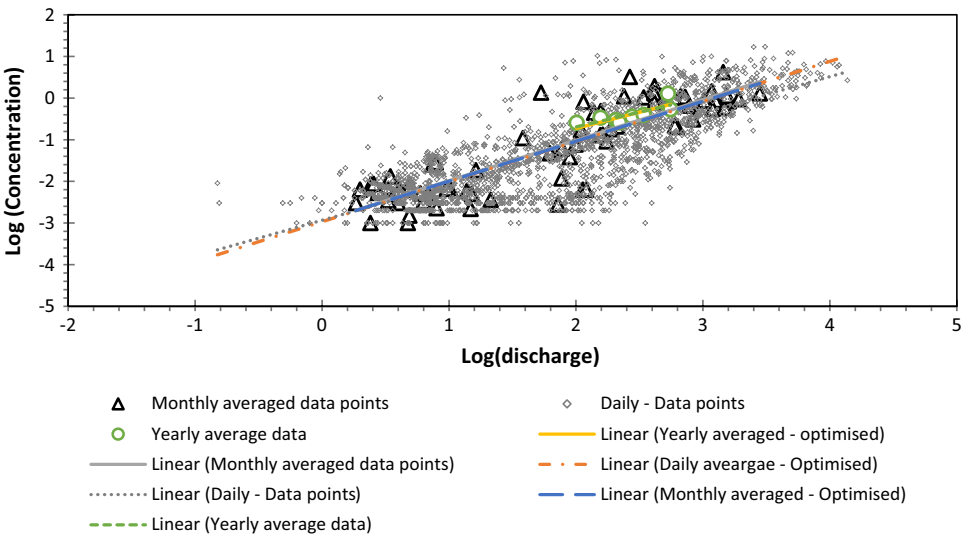


Fig. 7 SRC fitted by optimization and OLS regression

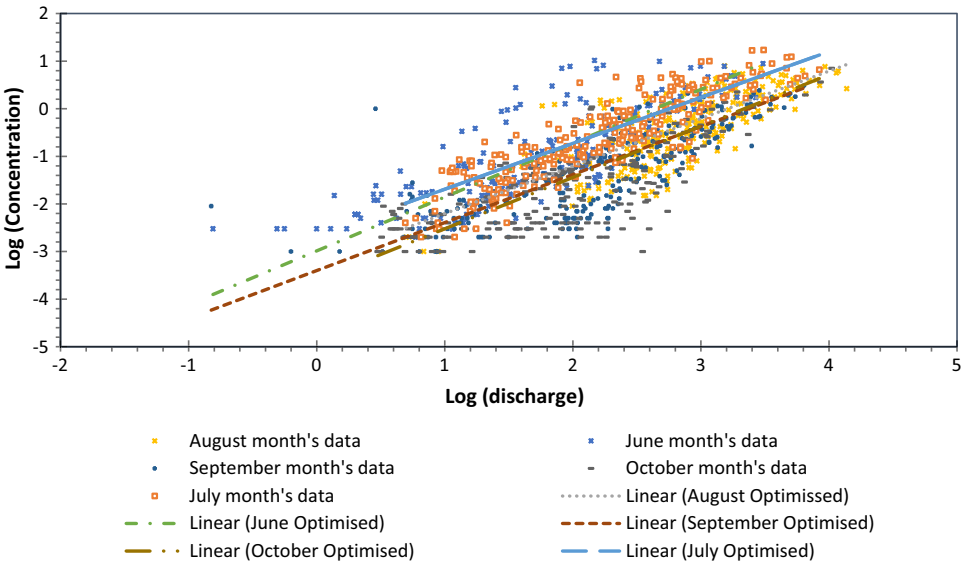


Fig. 8 Month-wise SRC fitted by optimization and OLS regression

Table 4 SRC models developed using optimization technique

SRC model		Log (a)	b	Coefficient of determination, R^2
Group of data	No. of data points			
Daily data-based SRC model	1889	−3.0131	0.9016	0.6268
Monthly averaged data-based SRC model	76	−2.9571	0.9593	0.7796
Yearly averaged data-based SRC model	9	−2.1733	0.6751	0.6048
Month-wise data-based month-wise SRC models				
June month	217	−2.9862	0.9031	0.6340
July month	260	−2.8546	1.0642	0.7751
August month	279	−3.5000	1.0980	0.5851
September month	268	−3.4000	0.9078	0.6114
October month	240	−3.5000	0.8785	0.5615

Comparison of SRC developed by OLS regression and optimization

SRC models were obtained from different groups of data, i.e., daily, monthly, yearly and month-wise though they were applied to daily time step data (discharge) to predict daily concentration and sediment load. Comparison of observed and predicted sediment load passing Sarangkheda gauging station for the period 1984–1992 disclosed that SRC fitted by optimization gave better prediction than OLS regression-fitted SRCs (Fig. 9). Table 5 shows the percentage error, NSMEF and *d* computed between observed and predicted load. The NSMEF scores of the yearly, monthly and daily models fitted by optimization are 0.4603, 0.4813 and 0.6240, respectively. The coefficient of determination obtained for the monthly SRC model fitted by optimization is 0.7796 and is the highest of all.

Despite this close agreement of the optimization-fitted SRCs with predicted sediment load for the entire 9-year period, yearly sediment load predictability was investigated. Not surprisingly, the accuracy of the yearly predictions decreased. The percentage error ranged from −79.87 to 82.65% for optimization-fitted models while for OLS regression-fitted model the error ranged from −88.24 to 98.64%. The dissimilarity of the percentage error between the entire 9-year period and yearly sediment load prediction demonstrates the competence of the SRCs fitted by optimization to round the error associated with longer periods.

The variability of predictions for yearly accumulated sediment loads with respect to data grouping was observed less for models fitted by optimization (Fig. 10), as compared to the models fitted by OLS regression (Fig. 11). This shows that the SRC models fitted by OLS regression are highly susceptible to data grouping and resolution, while those designed by the optimization technique are less influenced by the resolution of the data. This is due to the fact that the OLS regression models were further calibrated using the optimization technique by considering an objective function for minimizing the error between the observed and the predicted loads. From the results of the SRC models devised using the diversified approaches, it could be legitimated that the ones that were developed using optimization were the best.

The composition of the SRC fitting data significantly impacts the SRC produced (Horowitz 2003). Though results of the study have revealed that sediment load predictions obtained from SRCs fitted by optimization is independent of the composition of the dataset and hence, the coefficients of the obtained SRCs were analyzed. A negative correlation between the regression coefficients of the fitted SRCs is evident if the sediment flow regime is consistent (e.g., Asselman

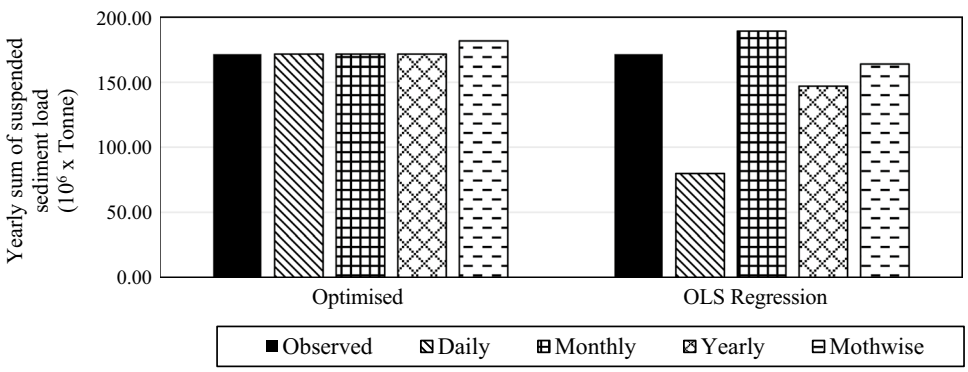


Fig. 9 Differences between actual and sediment rating curve-derived sediment load passing Sarangkheda gauging station during 1984–1992 period

Table 5 Statistical summary of developed SRC models

SRC model		Percentage Error	Nash–Sutcliffe model efficiency factor	Index of agreement, *d*
Group of data	Fitting procedure used for fitting an SRC			
Daily data-based SRC model	OLS regression	−53.42	0.5892	0.9964
	Optimization technique	−3.52E–07	0.4603	0.9953
Monthly averaged data-based SRC model	OLS regression	13.32	0.3478	0.8699
	Optimization technique	*−1.33E–07*	*0.4813*	*0.8834*
Yearly averaged data-based SRC model	OLS regression	−14.29	0.6872	0.9018
	Optimization technique	−2.01E–08	0.624	0.9031
Month-wise data-based month-wise SRC models				
June	OLS regression	−84.24	0.1988	0.3991
	Optimization technique	3.78E–07	0.6438	0.9148
July	OLS regression	17.24	0.0653	0.8324
	Optimization technique	−3.48E–05	0.4406	0.8713
August	OLS regression	−4.08	0.5757	0.9106
	Optimization technique	1.60E–11	0.579	0.9124
September	OLS regression	−49.98	0.6557	0.8387
	Optimization technique	−1.54E–07	0.9143	0.9779
October	OLS regression	−71.19	0.3879	0.5945
	Optimization technique	−7.90E–08	0.9371	0.9821

Italic values show the best performing model

The observed load is computed from the observed discharge and sediment concentration while the predicted load is computed using observed discharge and predicted sediment concentration

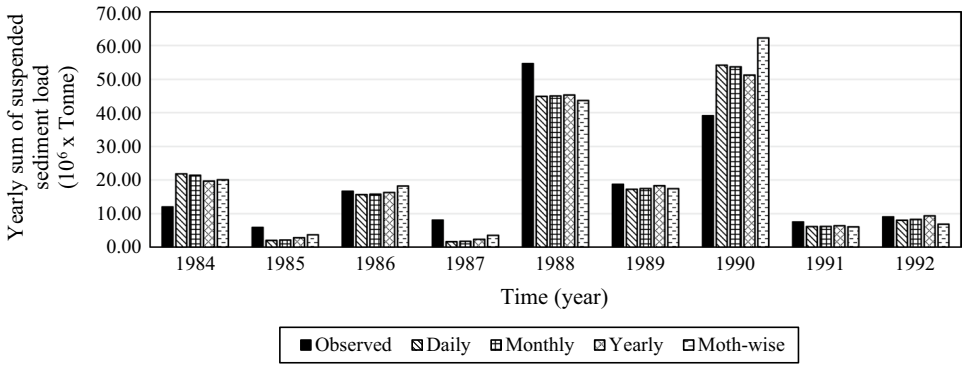


Fig. 10 Comparison of observed and predicted sediment load for the sediment rating curve fitted by optimization

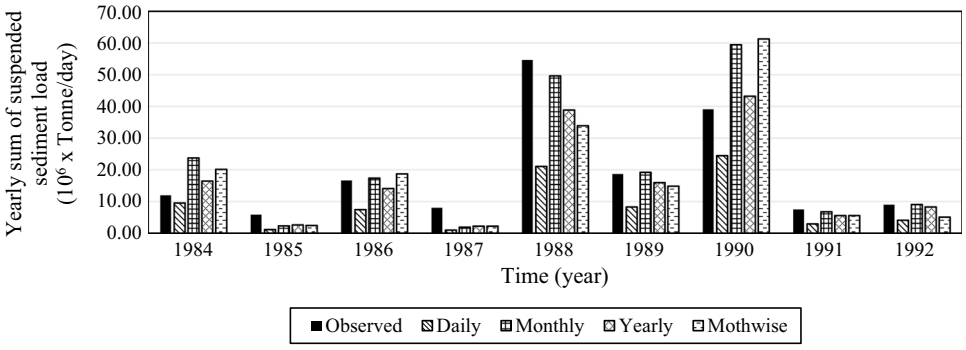


Fig. 11 Comparison of observed and predicted sediment load for the sediment rating curve fitted by OLS regression

2000; Syvitski et al. 2000). Sediment flow regime in the river reach of Sarangkheda gauging station is consistent during the 1984–1992 period, as no major activity has occurred in the upstream catchment area. Furthermore, the SRCs are obtained by grouping the observed data at a fixed location and period. Therefore, it is likely that the SRC coefficients should exhibit a good relationship. SRC coefficients obtained from daily, monthly, yearly and month-wise groups of data were plotted and are presented in Fig. 12. July and August month fitted SRCs from the month-wise models were only used to assess the relationship of the coefficients because they contribute 79% of the sediment load from a hydrological year. The coefficient of determination was found to be 0.7651 between SRC coefficients fitted by OLS regression, while for the coefficients of the SRCs fitted by optimization approach, it increased to be 0.8197. The upturn of the relationship between SRC coefficients indicates that the SRC fitted by optimization represents the flow and sediment transport regime in a better way, as compared to OLS regression-fitted SRCs.

Application of the SRC model to predict reservoir-inflowing sediment load by filling spatial and temporal data gaps

The availability of temporal and spatial data from the hydrometric observations pertaining to two consecutive hydrographic surveys is presented in Table 6. The surveys were undertaken in 1983 and 1992. Since the Sarangkheda gauging station was established only in 1984, hydrometric observations of the entire catchment area for the preceding year were missing. Moreover, the gauging station's observed data (1984–1992) do not account for the suspended sediment from region B (Fig. 2). Hence, the model was applied to bridge the spatiotemporal data gap. The SRC models developed at the Sarangkheda gauging station were transferred to the Ukai Reservoir head. Discharge inflow to the reservoir was identified from a water budget model in the form of a spreadsheet program. Elevation storage curve was utilized to obtain the storage volume. Daily inflow ($I_{t=1\,\mathrm{day}}$) was computed from the change in storage volume in accordance with the downstream release by the spillway, as well as the hydropower plant and reservoir water loss by evaporation (Eq. 10). Volume of evaporation was estimated using the observations of pan evaporimeter:

$$I_t = S_t - S_{t-1} + \left(\mathrm{VR_{ph}} + \mathrm{VR_{ULBMC}} + \mathrm{VR_{SPILLWAY}} + \mathrm{VR_{EV}}\right) \quad (10)$$

I_t = inflow at the end of the period, S_{t-1} = storage at the beginning of the period, S_t = storage at the end of the period, $\mathrm{VR_{ph}}$ = volume of release through powerhouse during period t, $\mathrm{VR_{ULBMC}}$ = volume of release through Ukai left bank main canal during period t, $\mathrm{VR_{SPILLWAY}}$ = volume of release

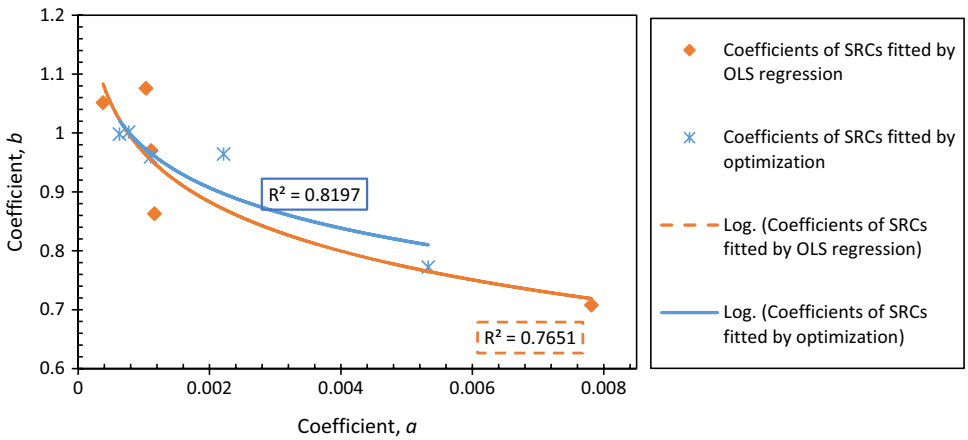


Fig. 12 Correlation between the coefficients of the SRC obtained from OLS regression and optimization fitting

Table 6 Spatial and temporal suspended sediment concentration and discharge data gaps and inconsistency

Year		1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Hydrometric observation data for regions A and B (Fig. 2)	A	Missing data	Present but inconsistent								
	B		Missing data (ungauged catchment area)								
Hydrographic survey		Survey done	–	–	–	–	–	–	–	–	Survey done

Regions A and B are as shown in Figs. 1 and 2, region B is ungauged region and data are missing for the period 1983–1992. The Sarangkheda gauging station was established in 1984, and so data of 1983 are missing

through radial gates (spillway) during period t, VR_{EV} = volume of water lost by evaporation during period t.

For the year 1983, discharge inflow was directly used as input in the SRC models to estimate the suspended sediment load contributed by the complete reservoir upstream catchment. As per the monthly SRC fitted by optimization, sediment inflowing the reservoir in the year 1983 was computed to be 23.56×10^6 tonnes. For the period 1984–1992, the daily water volume contributed by region B was discerned and it was converted into the daily discharge contribution of the region. The discharge contributed by region B was thus used as input in the SRC models to ascertain the suspended sediment load contribution of the region. Region B (ungauged catchment area, 3825 km^2) which includes the reservoir rim catchment contributes about 5% (8.36×10^6 tonnes) of sediment load of the region A (gauged catchment area) (Table 7). The suspended sediment load received by the reservoir during the study period is furnished in Table 7.

Ukai Reservoir-trapped sediments

Reservoir-trapped sediment is derived from the product of trap efficiency (TE) and total sediment load (suspended load and bed load). TE was predicted by Brune median curve, which ranged from 96.9 to 99.4% (period 1983–1992).

Reservoir-deposited sediment density and deposition volume

Three approaches of sediment bulk density were used to compute deposition volume. First, density was computed using empirical approach (Lara and Pemberton 1963; Miller 1953), second, based on the mean observed densities of Indian reservoirs having similar operational characteristics, density was computed (i.e., 1191 kg/m^3) and third, typical value of density (i.e., 1400 kg/m^3) was considered (Tebbi et al. 2012).

Estimation of the density by empirical approach was carried out in two steps. The initial sediment density was computed (Lara and Pemberton 1963), which was further processed for the effect of consolidation (Miller 1953). Sediment sampling of the submerged deposited sediments showed that the clay, silt and sand content of the sediment is 32%, 51% and 17%, respectively (Table 1). According to the reservoir operation, the sediment always remains submerged in the reservoir and is never exposed to sunlight or air. The mean initial density obtained through Eq. 8 was 963.82 kg/m^3. The average density after 9 years of compaction (final density) as inferred from Miller (1953) was 966.07 kg/m^3.

The three densities (viz. empirical approach based, mean observed density and typical density value) and sediment load trapped by the reservoir (obtained from different SRC models) were used to convert the sediment mass inflow to volumetric terms. RCL observed by two consecutive hydrographic surveys performed in the years 1983 and 1992 is 466.200×10^6 m^3. Table 8 demonstrates the difference in the observed and predicted volumes. The percentage error noticed in Table 8 urged further investigation on the certainty of the estimated density.

Empirical density estimate

The RCL predicted from the density estimated by empirical formulation resulted in an underprediction ranging from

Table 7 Total suspended load contributed by catchment

SRC model (1)		SSL_A^a (2)	SSL_{A+B}^b (3)	SSL_B^c (4)	SSL_A^d (5) (2) + (3) + (4) = (5)
Group of data used for SRC model development	Fitting procedure used for fitting an SRC				
Observed data		171.595	–	–	–
Daily data-based SRC model	OLS regression	80.788	11.546	4.349	96.682
	Optimization technique	173.379	23.021	7.771	204.171
Monthly averaged data-based SRC model	OLS regression	191.286	25.840	8.936	226.062
	Optimization technique	173.383	23.559	8.215	205.157
Yearly averaged data-based SRC model	OLS regression	148.701	22.888	9.844	181.433
	Optimization technique	173.409	25.911	10.531	209.850
Summation of month-wise SRC model	OLS regression	157.993	12.551	6.238	176.782
	Optimization technique	174.970	16.497	10.991	202.457

[a]Total suspended load contributed by region A in million tonnes for the period 1984–1992

[b]Total suspended load contributed by entire catchment in million tonnes for the year 1983

[c]Total suspended load contributed by region B in million tonnes for the period 1984–1992

[d]Total suspended load contributed by entire catchment in million tonnes for the period 1983–1992

Table 8 Percentage error between observed and computed sediment volume

SRC models		PE_a (%)	PE_b (%)	PE_c (%)
Data group	Fitting technique			
Daily data-based SRC model	OLS regression	−74.24	−79.1	−82.22
	Optimization technique	−45.6	−55.87	−62.46
Monthly averaged data-based SRC model	OLS regression	−39.77	−51.14	−58.44
	Optimization technique	−45.34	−55.66	−62.28
Yearly averaged data-based SRC model	OLS regression	−51.66	−60.79	−66.64
	Optimization technique	−44.09	−54.65	−61.42
Summation of month-wise SRC model	OLS regression	−52.9	−61.79	−67.50
	Optimization technique	−46.06	−56.24	−62.78

PE, percentage error computed between observed and predicted reservoir-deposited sediment volume. Suffix *a, b* and *c* stands for the density estimate used to derive the volume prediction, *a* implies volume computed using the empirical formula of bulk density, *b* implies volume computed using density of Indian reservoirs having similar operational characteristics and *c* implies volume computed using typical value of density (1400 kg/m^3)

39.77 to 74.24%. Such obtained underprediction cannot be considered for the estimation of daily RCL (Table 8), and uncertainty of the density estimate is assessed. The density estimated by the Lara and Pemberton (1963) and Miller (1953) approach using the sampled sand, silt and clay content depends on the particular specimens in the sample. Since these values can vary from sample to sample, particular samples can produce misleading values resulting in incorrect density estimates. The possibility of estimating a wrong density, on the basis of the collected sand silt and clay content, cannot be absolutely rejected except complete reservoir area is accurately sampled. This is, of course, physically and economically challenging as the dam rises to a maximum height of 70 m above the river bed level (Ukai Dam authority). Generally, in reservoir sediment sampling, the depth of water becomes one of the major hindrances for sample collection. Hence, due to limited samples in the spatial domain, the density representing the complete reservoir area has to be inferred from the sample-obtained mean density. In order to obtain an objective or subjective bias correction factor, which can be multiplied to the sample-estimated density, to infer reservoir representative density, the major sediment density influencing factors and the sampling program-related biases have to be evaluated. The inadequacy of such data limits us while deriving the bias correction factor. Nevertheless, in order to derive the corrected density estimate, the SRC that stood best for load prediction during SRC development stage (month-wise data-grouped SRC fitted by optimization, Table 4) is considered as the best SRC and a submerged sediment density of the reservoir is derived as a ratio of accumulated sediment load to the observed RCL (observed from hydrographic surveys conducted during 1983 and 1992). The computed representative density is found to be 528.075 kg/m^3; such low value of computed density questions the validity of the derived sediment load or the 1992 hydrographic surveyed reservoir capacity.

Statistical evaluation of the density estimated by the empirical approach (Lara and Pemberton 1963; Miller 1953) and density obtained by equating sediment load (gravimetric) with hydrographic surveyed RCL (volumetric) is performed. Assuming that the statistical population of density follows a normal distribution and considering the empirical-derived density estimate as mean of statistical sample and the density obtained by equating gravimetric sediment load with volumetric RCL as statistical-population mean, two-tailed *t* test is performed to compare the mean of statistical sample and population for the known standard deviation of the sample (for ten sampled specimens).

T test statistic (t) is a standardized value calculated from the mean of sample and population, which incorporates both the sample size and its variability. The null hypothesis is exactly sufficed if t is found to be zero; as the absolute t is increased, the significance of accepting the null hypothesis is decreased (Johnson 2017). For the present study, the null hypothesis (H_0) proposes that the statistical-sample mean is not significantly different than the statistical-population mean, and on the other hand, the alternative hypothesis (H_a) states that the statistical-sample mean is significantly different than the population mean. The critical *T* value ($t_{\alpha/2,\,\mathrm{df}}$) for significance level (α) as 95% with nine degrees of freedom (df) is found to be 2.262, and t is calculated as 6.784. As t is higher than $t_{0.975,9}$, it is determined from hypothesis testing that H_0 is rejected in favor of H_a. This does not necessarily mean that H_a is true; it only suggests that there is not sufficient evidence to accept the null hypothesis suggesting alternative hypothesis may become true. However, the statistical test clearly rejected the statistical significance of the two densities to be equal. Hence, the correctness of the reservoir capacity (observed during the hydrographic survey of 1992), the derived sediment load and the density estimate remains a question. Therefore, four scenarios of RCL between the period 1983 and 2003 are generated using the

month-wise data-grouped SRC fitted by optimization and different density estimates. It was expected that if the sediment load predictions are accurate, then the capacity loss predicted from one scenario will be in good agreement with the difference between the reservoir capacity observed during the hydrographic surveys of 1983 and 2003, resulting in one true reservoir sediment density. Densities of 528.075 kg/m^3, 966.07 kg/m^3, 1191 kg/m^3 and 1400 kg/m^3 are used to generate scenarios 1, 2, 3 and 4, respectively. Comparing the observed and predicted RCL, an overprediction of 82.10% is reported from scenario 1 while an underprediction of 0.46% is reported from scenario 2. However, scenarios 3 and 4 have produced an underprediction of 19.26% and 31.31%, respectively. The discrepancy ratio of these estimates is presented in Table 9. The density estimated from the empirical approach (966.07 kg/m^3) showed least deviation between observed and predicted capacity loss. A lower value of the error from scenario 2 validates the SRC model to produce sediment load and the density obtained by empirical approach.

In order to relate hydrometric and hydrographic observations for the prediction of reservoir capacity loss on daily timescale, correct prediction of sediment load has to be done, density estimate has to be accurate and the uncertainty from the hydrographic survey should be narrowed. The present research has highlighted major problems, which can hinder the linking of the hydrometric observations with the RCL. Uncertainties of the predicted sediment load, density estimate and hydrographic survey are interlinked. The proposed approach is beneficial as it assesses the correctness of the observed data by statistical evaluation of density estimates and generation of the capacity loss scenarios. The result of the study demonstrates the prediction of daily RCL in uncertain data conditions. It is necessary to discuss here that, in the future, more consideration should be given to ensure the correctness of the hydrographic surveys. RCL obtained from hydrographic surveys should be cross-verified with the RCL derived from the hydrometric observation, using the present approach. In addition to it, from the ongoing research, few essential points identified for quality control, minimization of the surveying error and the uncertainty are listed here: (1) Hydrographic surveys should be carried out on standardized operating procedures providing a uniform method of planning, collecting, processing and analyzing data. (2) Before the execution of the hydrographic survey, by examining the reservoir bathymetry obtained from previous surveys, critical locations should be identified, so that a comparison overlay for those particular locations can be prepared. (3) The main survey lines directions (perpendicular to the general direction of contours) and their spacing should be identified from the previous surveys experience. (4) Crossline direction, i.e., at right angles to mainline direction, should be selected as a vital quality regulation measure. So when any interpolation algorithm is used, the results can be verified with such control lines. (5) Surveying vessel speed has to be assessed for the expected range of depths in the survey area and the type of echo sounder in use to reduce the measurement and observation uncertainty. (6) A consistent datum must be used throughout all hydrographic survey projects, or the data has to be adjusted if comparison has to be made.

Table 9 Reservoir capacity loss with respect to estimated densities for the period 1983–2003

Scenario	Sediment density	RCL$_o^a$ (10^6 m^3)	RCL$_p^b$ (10^6 m^3)	DR^c (%)
Scenario 1	528.05	548.71	999.22	1.82
Scenario 2	966.07		546.20	1.00
Scenario 3	1191.00		443.05	0.81
Scenario 4	1400.00		376.91	0.69

[a]RCL$_o$, reservoir capacity loss derived from the gross reservoir capacity observed in 1983 and 2003 hydrographic surveys

[b]RCL$_p$, reservoir capacity loss computed from the sediment load trapped by the reservoir and different density estimates

[c]DR, discrepancy ratio between RCL$_o$ and RCL$_p$

Gross reservoir capacity loss prediction

The temporal lumped relationship established between the hydrometric observations of the reservoir upstream gauging station and the reservoir capacity derived from hydrographic surveys were utilized to disintegrate the observed phenomena on a daily time step.

Employing the SRC model, daily depositing sediment volume and continuous timeline of gross reservoir capacity are obtained (Fig. 13). Though the model is based on a simple approach, the validity of the developed relationship to predict the RCL was found considerable. The relationship was thus utilized to estimate the action of the extreme hydrological events to RCL (Table 10). During a flood event on August 08, 2006, about 1991×10^6 m^3 of water flowed into the reservoir in a single day, which brought about 42.16×10^6 tonnes (estimated) of sediment load causing 43.64×10^6 m^3 (estimated) of RCL (Fig. 14). This model estimate is 0.59% of 2003 surveyed reservoir capacity. The average RCL from its first impoundment in 1972 to the last bathymetry survey carried out in the year 2003 was 35.35×10^6 m^3/year. The results have shown that about 50% of the capacity loss of a year may occur during a single extreme event. The design siltation rate of the Ukai Dam is 0.149×10^3 m^3/km^2/year, whereas the observed rate in the reservoir is 0.568×10^3 m^3/km^2/year (CWC 2015), which shows that the rate of siltation was underestimated by 73.77%.

Ten extreme flood events observed in the history of the Ukai Reservoir were analyzed for the inflowing sediment load and RCL (Table 10). With respect to initial reservoir capacity, RCL of 0.512% to 0.081% was observed. Through all these

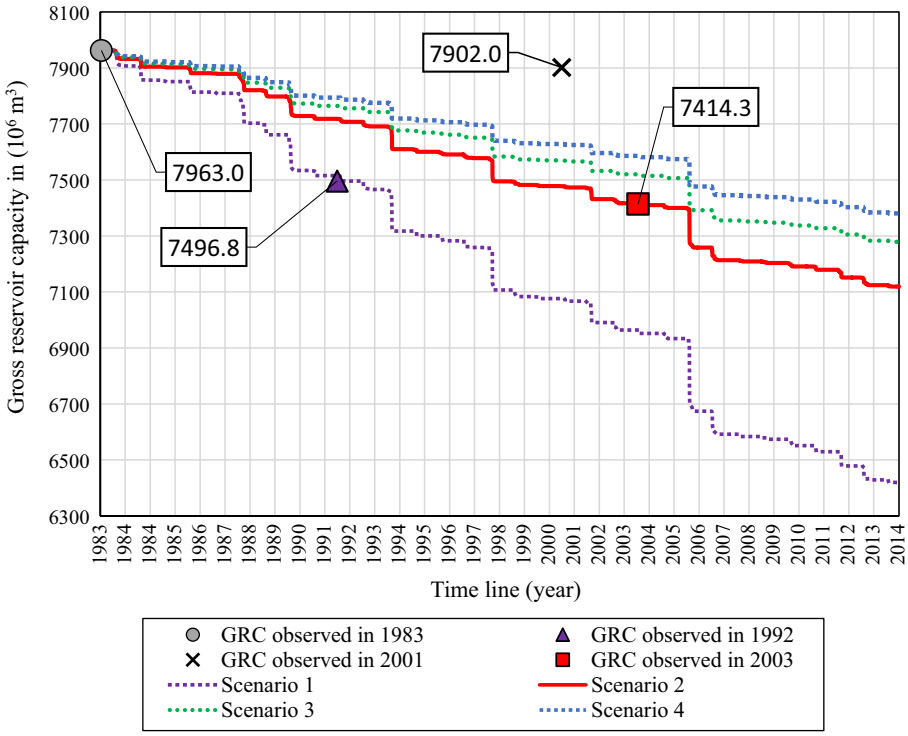


Fig. 13 Time series of predicted reservoir capacity loss

Table 10 Extreme events in reservoir catchment and its impact on capacity loss

S. no.	Date	Amount of water received by reservoir ($10^6 \times m^3$)	Estimated sediment inflow ($10^3 \times$ ton)	Estimated capacity loss ($10^6 \times m^3$)
1	08-08-2006	1991.00	42.16	43.64
2	16-09-1998	1839.63	36.11	37.38
3	07-09-1994	1350.09	19.69	20.39
4	20-08-1984	1188.00	15.33	15.87
5	17-08-1990	1022.80	11.43	11.83
6	09-07-2007	1021.88	11.41	11.81
7	04-10-1988	909.02	9.07	9.39
8	07-09-2012	905.56	9.01	9.32
9	20-08-1989	803.52	7.12	7.37
10	02-08-2013	779.93	6.72	6.96

ten events, about 173.96×10^6 m^3 of sediment volume settled in the reservoir, which reduced the capacity of the reservoir by 2.04%. It is to be noted that the estimated RCL by such hydrometric and hydrographic relationship may only provide a first-order capacity loss estimate. Yet, in data-scarce condition, it may stand as a very useful tool to understand the response of hydrological events directly on the RCL.

Conclusions

The certainty of the predicted sediment load trapped by the reservoir, estimated bulk density of the deposited sediment and the hydrographic survey observed reservoir capacity limits the accuracy of the reservoir capacity loss (RCL) prediction. Accurate prediction of sediment load using sediment rating curve (SRC) has been emphasized in the present study. A novel SRC fitting approach by means of optimization technique is proposed to predict sediment load using low-frequency (once a day) suspended sediment sampled data. Use of SRCs to produce sediment load predictions by different grouping and fitting procedures has produced the following learnings.

- Data grouping (composition of data) and curve fitting procedures adopted for the SRC model development will change the exponent and scaling coefficient of the SRC models.

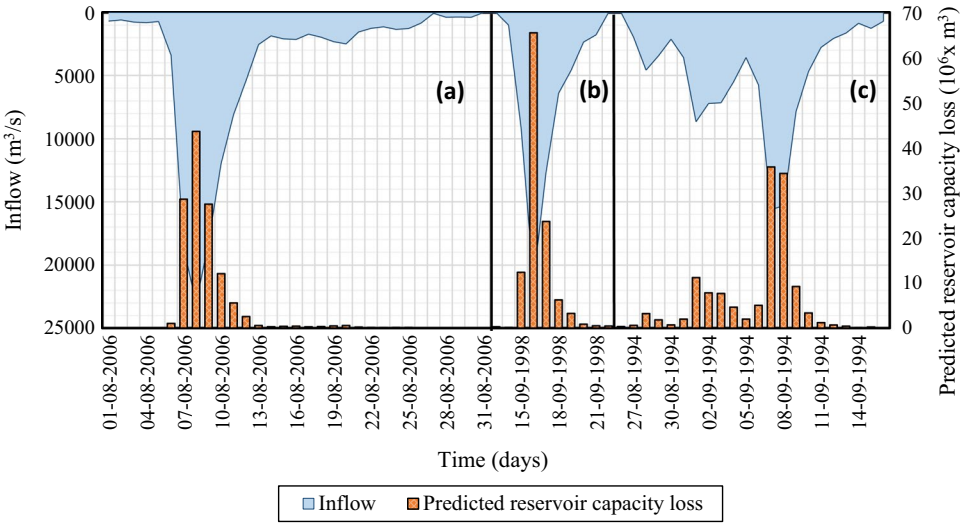


Fig. 14 Inflow and reservoir capacity loss. Three extreme events which occurred during 2006 (**a**), 1998 (**b**) and 1994 (**c**) are presented

- Degree of accuracy of the predicted load from low-frequency observations varies with the composition of the data in case of SRCs fitted by ordinary least square regression, whereas the effect of data composition to predict sediment load was observed to be low in the case of SRCs fitted by the proposed optimization approach.
- The percentage error observed between the entire period and yearly sediment load prediction demonstrated that the competence of the SRCs fitted by optimization to round the error associated with longer periods is high.
- Assessment of the relationship between the SRC coefficients has suggested that the SRCs fitted by optimization represent the flow and sediment transport regime in a better way.

Hence, if the sampling frequency is low and observations are made independent of the hydrograph (as instantaneous temporal point measurement), then the observed data can be utilized effectively by employing the fitting procedure based on optimization.

The approach developed in this paper provides a means to validate the estimate of reservoir-submerged sediment density obtained from the sampled sediment sand, silt and clay content and the empirical density predicting models. Statistical hypothesis testing (P value or T value) is required to be performed between the mean density estimated from sediment samples and the density obtained by equating sediment load prediction to the observed capacity loss. The result of the hypothesis tested should prove that both of these densities do not differ. If they differ, further investigation of the estimated density, as well as the hydrographic surveys, is required, before utilizing it to derive RCL.

Application of the developed relationship between the hydrographic and hydrometric observations to disintegrate the RCL on a daily scale will enhance the understanding of event-based capacity loss, which may stand as a useful approximation to devise a sediment management strategy.

Acknowledgements We thank Gregory L. Morris, P.E., Ph.D. for the motivation and assistance he provided in taking up this research and R. N. Sankhua, Ph.D., Director, Basin Planning, Ministry of WR, India, for comments on an earlier version of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed persons. We are grateful to the CWC, India, for collecting and archiving the hydrometric dataset. We would like to thank Ukai Dam authority for the reservoir water budget and hydrographic survey data. We are also immensely grateful to the reviewers for their insights.

Compliance with ethical standards

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

References

Annandale GW, Morris GL, Karki P (2016) Extending the life of reservoirs: sustainable sediment management for dams and run-of-river hydropower. World Bank. https://doi.org/10.1596/978-1-4648-0838-8

Arabkhedri M, Lai FS, Ibrahim N-A, Mohamad-Kasim M-R (2010) Effect of adaptive cluster sampling design on accuracy of sediment rating curve estimation. J Hydrol Eng 15:142–151. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000171

Asselman NEM (2000) Fitting and interpretation of sediment rating curves. J Hydrol 234:228–248. https://doi.org/10.1016/S0022-1694(00)00253-5

BIS-12182 (1987) Guidelines for determination of effects of sedimentation in planning and performance of reservoirs. BIS 12182

Brune G (1953) Trap efficiency of reservoirs. EOS Trans Am Geophys Union 34:407–418. https://doi.org/10.1029/TR034i003p00407

Bussi R-LX, Francés F, Benito G, Sánchez-Moya Y, Sopeña A (2013) Sediment yield model implementation based on check dam infill stratigraphy in a semiarid Mediterranean catchment. Hydrol Earth Syst Sci 17:3339–3354. https://doi.org/10.5194/hess-17-3339-2013

Bussi G, Dadson SJ, Bowes MJ, Whitehead PG (2017) Seasonal and interannual changes in sediment transport identified through sediment rating curves. J Hydrol Eng 22:06016016. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001466

Chaudhuri D (2006) Life of Maithon reservoir on ground of sedimentation: case study in India. J Hydraul Eng 132:875–880. https://doi.org/10.1061/(ASCE)0733-9429(2006)132:9(875)

Crawford CG (1991) Estimation of suspended-sediment rating curves and mean suspended-sediment loads. J Hydrol 129:331–348. https://doi.org/10.1016/0022-1694(91)90057-O

CWC (2014) Sediment data year book (including bed material), Mahi, Sabarmati, Tapi & Other West Flowing Rivers. Gandhinagar, Gujarat, India

CWC (2015) Compendium on silting of reservoirs in India. New Delhi, India

De Girolamo AM, Pappagallo G, Lo Porto A (2015) Temporal variability of suspended sediment transport and rating curves in a Mediterranean river basin: the Celone (SE Italy). Catena 128:135–143. https://doi.org/10.1016/j.catena.2014.09.020

Garg V, Jothiprakash V (2008) Trap efficiency estimation of a large reservoir. ISH J Hydraul Eng 14:88–101. https://doi.org/10.1080/09715010.2008.10514907

Harrington ST, Harrington JR (2013) An assessment of the suspended sediment rating curve approach for load estimation on the Rivers Bandon and Owenabue, Ireland. Geomorphology 185:27–38. https://doi.org/10.1016/j.geomorph.2012.12.002

Heidarnejad M, Golmaee SH, Mosaedi A, Ahmadi MZ (2006) Estimation of sediment volume in karaj dam reservoir (Iran) by hydrometry method and a comparison with hydrography method. Lake Reserv Manag 22:233–239. https://doi.org/10.1080/07438140609353900

Horowitz AJ (2003) An evaluation of sediment rating curves for estimating suspended sediment concentrations for subsequent flux calculations. Hydrol Process 17:3387–3409. https://doi.org/10.1002/hyp.1299

Johnson RA (2017) Probability and statistics for engineers, 9th edn. Richard A. Johnson published by Pearson Education, London

Kondolf GM, Gao Y, Annandale GW et al (2014) Sustainable sediment management in reservoirs and regulated rivers: experiences from five continents. Earth's Future 2:256–280. https://doi.org/10.1002/2013EF000184

Lane EW, Koelzer VA (1943) Density of sediments deposited in reservoirs. St Paul US Eng Dist Sub-Office, Hydraul Lab Univ Iowa, Iowa

Lara JM, Pemberton EL (1963) Initial unit-weight of deposited sediments. In: Federal inter-agency sedimentation conference USDA, USA, pp 818–845

Marineau MD, Wright SA (2017) Daily reservoir sedimentation model: case study from the Fena Valley Reservoir, Guam. J Hydraul Eng 143:05017003. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001344

Miller CR (1953) Determination of the unit weight of sediment for use in sediment volume computations. US Department of the Interior, Bureau of Reclamation, Project Planning Division, New York

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

Palmieri A, Shah F, Dinar A (2001) Economics of reservoir sedimentation and sustainable management of dams. J Environ Manag 61:149–163. https://doi.org/10.1006/jema.2000.0392

Salas JD, Shin H-S (1999) Uncertainty analysis of reservoir sedimentation. J Hydraul Eng 125:339–350. https://doi.org/10.1061/(ASCE)0733-9429(1999)125:4(339)

Singh VP, Jain SK, Tyagi A (2013) Risk and reliability analysis: a handbook for civil and environmental engineers. Choice Rev Online 45:45-5015. https://doi.org/10.5860/choice.45-5015

Small IF, Rowan JS, Duck RW (2003) Long-term sediment yield in Crombie Reservoir catchment, Angus; and its regional significance within the Midland Valley of Scotland. Hydrol Sci J 48:619–636. https://doi.org/10.1623/hysj.48.4.619.51408

Strand RI, Pemberton EL (1982) Reservoir sedimentation. In: Technical guideline for bureau of reclamation, technical services engineering and research center. Bureau of Reclamation's Sedimentation and River Hydraulics Group, Denver, Colorado

Sumi T, Hirose T (2009) Accumulation of sediment in reservoirs. Water Storage Transp Distrib 28:224–252

Swamee BPK, Ojha CSP (1991) Bed-load and suspended-load transport of nonuniform sediments. J Hydraul Eng 117:774–787

Syvitski JP, Morehead MD, Bahr DB, Mulder T (2000) Estimating fluvial sediment transport: the rating parameters. Water Resour Res 36:2747–2760. https://doi.org/10.1029/2000WR900133

Tebbi FZ, Dridi H, Morris GL (2012) Optimization of cumulative trapped sediment curve for an arid zone reservoir: foum El Kherza (Biskra, Algeria). Hydrol Sci J 57:1368–1377. https://doi.org/10.1080/02626667.2012.712740

Vanoni VA (1979) Mechanics of sediment transportation and alluvial stream problems. Eng Geol 14:283–284. https://doi.org/10.1016/0013-7952(79)90070-X

Verstraeten G, Poesen J (2001) Variability of dry sediment bulk density between and within retention ponds and its impact on the calculation of sediment yields. Earth Surf Process Landf 26:375–394. https://doi.org/10.1002/esp.186

Verstraeten G, Poesen J (2002) Using sediment deposits in small ponds to quantify sediment yield from small catchments: possibilities and limitations. Earth Surf Process Landf 27:1425–1439. https://doi.org/10.1002/esp.439

Waikhom SI, Yadav SM (2017) Prediction of total load transport of an Indian alluvial river to estimate unmeasured bed load through an alternative approach. Curr Sci 113:1120

Walling DE (1977a) Limitations of the rating curve technique for estimating suspended sediment loads, with particular reference to British rivers. In: Proceedings of Symposia on Erosion and solid matter transport in inland waters, pp 34–48

Walling DE (1977b) Assessing the accuracy of suspended sediment rating curves for a small basin. Water Resour Res 13:531–538. https://doi.org/10.1029/WR013i003p00531

Warrick JA (2015) Trend analyses with river sediment rating curves. Hydrol Process 29:936–949. https://doi.org/10.1002/hyp.10198

Willmott CJ (1981) On the validation of models. Phys Geogr 2:184–194. https://doi.org/10.1080/02723646.1981.10642213

Acta Geophysica (2019) 67:1471–1481
https://doi.org/10.1007/s11600-019-00330-1

RESEARCH ARTICLE - HYDROLOGY

Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting

Bibhuti Bhusan Sahoo[1] · Ramakar Jha[1] · Anshuman Singh[1] · Deepak Kumar[1]

Received: 30 August 2018 / Accepted: 13 July 2019 / Published online: 20 July 2019

Abstract
This article explores the suitability of a long short-term memory recurrent neural network (LSTM-RNN) and artificial intelligence (AI) method for low-flow time series forecasting. The long short-term memory works on the sequential framework which considers all of the predecessor data. This forecasting method used daily discharged data collected from the Basantapur gauging station located on the Mahanadi River basin, India. Different metrics [root-mean-square error (RMSE), Nash–Sutcliffe efficiency (E_{NS}), correlation coefficient (R) and mean absolute error] were selected to assess the performance of the model. Additionally, recurrent neural network (RNN) model is also used to compare the adaptability of LSTM-RNN over RNN and naïve method. The results conclude that the LSTM-RNN model ($R=0.943$, $E_{NS}=0.878$, RMSE = 0.487) outperformed RNN model ($R=0.935$, $E_{NS}=0.843$, RMSE = 0.516) and naïve method ($R=0.866$, $E_{NS}=0.704$, RMSE = 0.793). The finding of this research concludes that LSTM-RNN can be used as new reliable AI technique for low-flow forecasting.

Keywords Artificial intelligence · Long short-term memory recurrent neural network · Low flow · Hydrological time series forecasting, naïve method

Introduction

Forecasting hydrologic time series (HTS) is a vital research topic for researchers, and undoubtedly forecasting has always been prime concern in hydrological practices. The forecasting of rainfall and water level in a river is very important task for water resources engineers for planning and monitoring water resources activities like providing irrigation water, maintaining environmental flow, providing drinking water, recreational purpose and many more other activities. Hydrological processes like stream flow generation processes are not only controlled by external climatic conditions, but also by physical properties (Beven 2012). Developing appropriate models for forecasting hydrological time series is a challenging task due to the influences of many factors and complicated hydrologic processes (Gárfias-Soliz et al. 2010; Nayak et al. 2004; Sang 2013; Sang et al. 2009).

While dealing with time series forecasting in hydrology, there are two approaches generally witnessed in the literature: the first one is stochastic models and the other one is artificial intelligence (AI) techniques. The AI techniques are generally referred as black box models in the literature due to its complex and unknown underlying process. It is very difficult to say which approach is the best for hydrological time series forecasting. However, a recent study on hydrological time series forecasting by (Papacharalampous et al. 2019) and its companion studies was algorithmically proved by using large datasets (135–2537 real-world time series and 16,000–48,000 simulated time series) that at the annual and monthly time scales traditional and AI techniques can perform equally well. The gaining popularity of AI techniques for HTS forecasting is clearly witnessed in the studies. (Papacharalampous et al. 2018a) assess the one-step ahead forecasting performance of 20 univariate time series forecasting methods to a large number of geophysical and simulated time series of 91 values. "The simulation experiments reveal the most and least accurate methods for long-term forecasting applications, also suggesting that the simple methods may be competitive in specific cases." Some of those AI models used in HTS includes neural network (NN)(Atiya et al. 1999; Kişi 2007), support vector machines (SVM) (Kisi and Cimen 2011; Sahoo et al. 2018; Sivapragasam et al. 2001), extreme learning machines

✉ Bibhuti Bhusan Sahoo
bibhuti5000@gmail.com

[1] Department of Civil Engineering, National Institute of Technology, Patna, India

(ELM) (Yaseen et al. 2016), adaptive neuro-fuzzy inference system (ANFIS) (Firat and Güngör 2007), random forest (Tyralis and Papacharalampous 2017) and many more such AI techniques can be seen in the literature. Tyralis and Papacharalampous (2018) used Prophet for multi-step ahead forecasting of monthly streamflow. The Prophet algorithm is developed by Facebook for time series forecasting. These AI techniques are the viable alternative tool to study hydrological phenomena, when hydrological attributes are limited and forecasting is essential.

HTS forecasting has received the remarkable consideration by the researchers in the last few decades, and many models for HTS forecasting have showed significant performance in terms of forecasting accuracy. HTS forecasting is still one of the difficult problems and is an active research area of interest in operational hydrology. We have applied sequential LSTM-RNN model for low-flow forecasting at Mahanadi River basin, using low-flow data from the Basantapur station. Additionally, to check its adaptivity for this process, a benchmark model (naïve method) and sequential model (RNN) were used for HTS forecasting.

The main purpose of this study is to inspect the suitability of LSTM-RNN for low-flow forecasting in the selected station of Mahanadi River basin India and compare it with RNN and naïve method. The naïve forecasting method is one of the most commonly used benchmarks for time series forecasting (Hyndman and Athanasopoulos2013; Pappenberger et al. 2015). This method simply sets all forecasts equal to the last value. This naïve method is appropriate when we are interested in multi-step ahead forecasting of time series without seasonality. However, in this study, we are interested in one-step ahead forecasting of monthly values; therefore, an appropriate naïve method is the one based on all monthly values of the last year.

Reviews on modeling approaches for HTS

HTS modeling can be categorized into two groups such as parametric and nonparametric methods. The most widely used parametric time series method is autoregressive integrated moving average (ARIMA) model (Box and Jenkins 1970), that is, ARIMA (p, d, q), where p, d, q, respectively, represent the autoregressive, integrated and moving average polynomial orders. Extensive application and assessments of the various classes of such models reported for the modeling of hydrologic time series were suggested in the past (Arena et al. 2006; Chen and Rao 2002; Hipel and McLeod 1994; Komorník et al. 2006; Srikanthan and McMahon 2001; Toth et al. 2000). Traditional time series forecasting involves decomposing the data into its components such as trend component, seasonal component and noise. In contrast to parametric approaches, nonparametric methods do not have a fixed model structure and parameter. Some of the previous studies also includes the chaotic behavior, hurst phenomena, stochastic and deterministic models and their performance in HTS forecasting (Dimitriadis and Koutsoyiannis 2015; Dimitriadis et al. 2016; Koutsoyiannis and Langousis 2011; Koutsoyiannis et al. 2008).

According to Wang et al. (2009), the HTS models can be broadly divided into three groups: regression-based methods, time series models and artificial intelligence (AI)-based methods. In recent years, HTS forecasting methods have been gradually shifting from traditional statistical models to AI approaches technique. AI techniques being capable of analyzing long series, handling large-scale data, recognizing patterns hidden in historical data and then applying those patterns to predict future scenarios have become increasingly popular in HTS modeling among researchers for developing a variety of models for time series prediction (Sahoo et al. 2017). Papacharalampous et al. (2018b) used random walk (with drift), autoregressive fractionally integrated moving average (ARFIMA), exponential smoothing state-space model with Box–Cox transformation, ARMA errors, trend and seasonal components (BATS), simple exponential smoothing, Theta and Prophet methods for univariate time series forecasting along with a naïve method based on the monthly values of the last year, used for benchmarking purposes. Some of the models used by Papacharalampous et al. (2018b) are quite new or rare in HTS forecasting and beyond the scope of this study.

The artificial neural network (ANN) is one of the common AI procedures established on the conceptualization of the brain and nervous systems (Abiodun et al. 2018) and was successfully applied for HTS forecasting. A broad review of the use of ANN in the hydrological field is given by ASCE Task Committee on "Application of Artificial Neural Networks in Hydrology" (ASCE 2000a, b). A wide number of research have published to report the forecasting performance of several time series models HTS (Carlson et al. 1970; Chang et al. 2002; Chen and Rao 2002; Cheng et al. 2005; Firat and Güngör 2008; Hu et al. 2001; Jain and Kumar 2007; Keskin et al. 2006; Komorník et al. 2006; Lin et al. 2006; Nayak et al. 2004; Salas 1993; Sivapragasam et al. 2007; Zounemat-Kermani and Teshnehlab 2008). Yaseen et al. (2018) have applied Elman recurrent neural network coupled with the rolling mechanism and gray models for streamflow forecasting over various lead times. Wunsch et al. (2018) advocated the suitability of nonlinear autoregressive networks with exogenous input (NARX) model for forecasting groundwater levels in several wells in southwest Germany with a lead time of 6 months ahead. Recently, Zhang et al. (2018a) successfully implemented long short-term memory network for sewer overflow monitoring. A comprehensive review of the state-of-the-art for the application of different AI techniques in streamflow

forecasting from 2000 to 2015 was presented by Yaseen et al. (2015). The LSTM-RNN has been found very useful in continuous time series processing such as word recognition and speech recognition. Some studies have investigated the feasibility of LSTM in soil moisture modeling and agricultural applications (Fang et al. 2017), water table depth predicting (Zhang et al. 2018b) drought forecast (Xu et al. 2018). Here, the performance of LSTM-RNN in low-flow forecasts was examined.

Case study and catchment description

In this paper, the monthly low-flow HTS at gauging station namely Basantapur (82°78′E, 21°72′N) of Mahanadi River basin, India, was analyzed. The Mahanadi River basin was situated between 80°28′E–86°43′E and 19°8′N–23°32′N (Fig. 1). It travels a distance of 851 km from the source of origin before falling into the Bay of Bengal. The major part of the basin is covered with agricultural land accounting for 54.27% of the total area, and 4.45% of the basin is covered by water bodies. The Mahanadi River basin has an average annual rainfall of 1572 mm, over 70% is precipitated during the southwest monsoon between June and October. The daily discharge data from June 1971 to May 2010 of the stations Basantapur were obtained from central water commission (CWC) Bhubaneswar and were used to calculate the monthly low-flow at Basantapur station.

Defining low flow and its importance

An appropriate definition of low flow differs from person to person according to the need of the study (Pyrce 2004). Low-flow situation is determined by a certain percentile of discharge (Ahn and Palmer 2016) or a truncation level in a stream. A significant amount of past studies has reported a number of low-flow indices such as Q_{95} (Laaha and Blöschl 2005), Q_{85} (Giuntoli et al. 2013), Q_{75} (Demirel et al. 2013; Jha and Smakhtin 2008; Pyrce 2004), used for low-flow study, where Q_{95}, Q_{85},Q_{75} are the discharge equaled or surpassed for the duration of 95%, 85%, and 75% of the observation period, respectively. In the study, 'low-flow' threshold is taken as the Q_{75} discharge, i.e., the flow is equal to or surpassed for the duration of 75% of the observation period which was obtained from the daily discharge data. It is to be noted that the hydrological, topographical and climatic conditions of rivers in India, in general, are quite different, and the approach suggested by Q_{75} (Jha et al. 2008; Pyrce 2004) may be applicable for Mahanadi River based on actual field conditions.

The reliable prediction of future low flow has many important applications in water resources planning and management. It is also important for the environmental/ecological discharge (Tegos et al. 2018). The importance of low flow encouraged researchers to apply different types of forecasting approaches to evaluate and forecast low flows in rivers. Low flows can be critical in determining how much

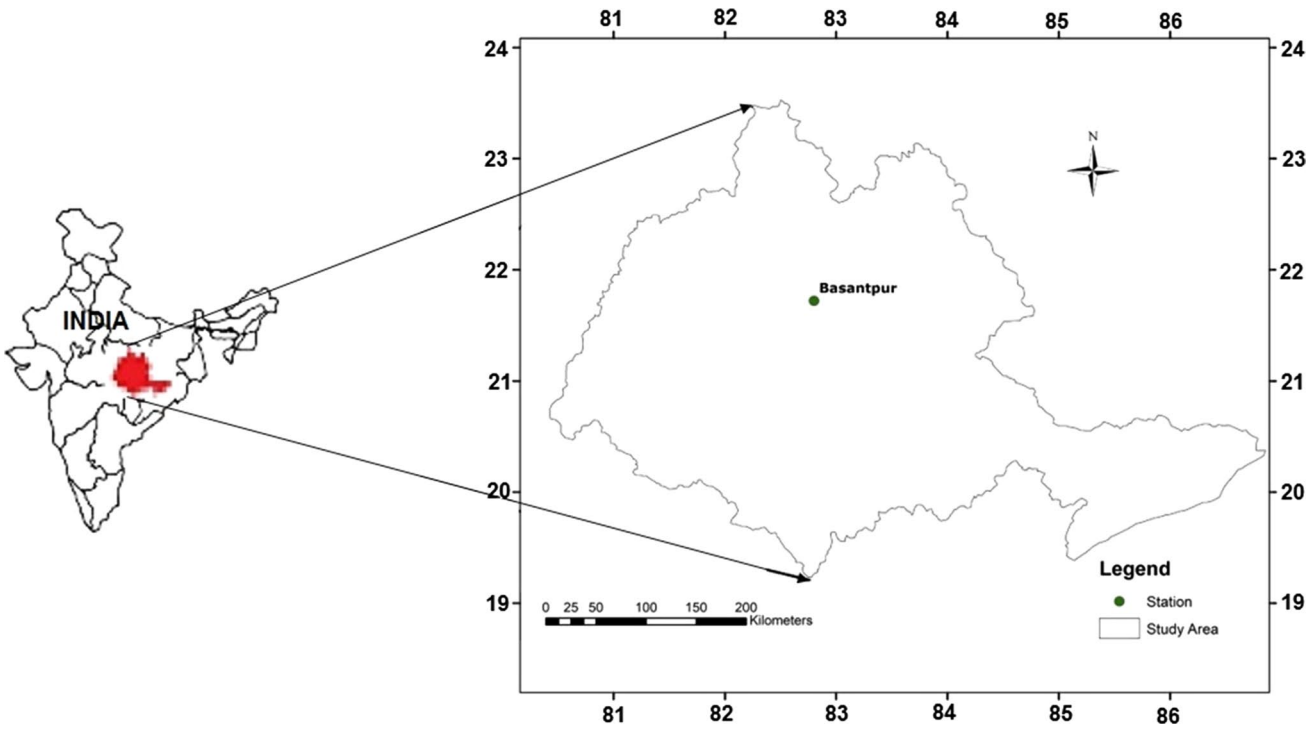


Fig. 1 Study area and selected station

water must by pass a run-of-river hydro-plant to maintain downstream river ecology and how much is available for power generation in the dry season. "Thermal power stations are dependent on cooling water and information on low flows when the availability of water for abstraction and dilution of cooling water is at a minimum is essential for design purposes" (WMO 2008). For all these applications, there may be a need to forecast flows in order to implement restriction on water use to minimize the risk of very severe restriction in the future. In some instances, licenses to extract water in excess of the available supplies have been issued and thus low-flow forecasts are essential forecast tool. The ultimate goal of the understanding of low-flow processes is to facilitate the development of early warning systems for low-flow adaptation and mitigation which is very crucial in managing water resources in the study river basin. Thus, low-flow assessment plays a crucial role in low-flow managing (Dracup et al. 1980) along with many environmental purposes related to the better management and sustainable development of water resources. Further, detailed information about the various aspect of the low flow can be found in manual on low-flow estimation—prediction by Gustard and Demuth (2009), and a review on low-flow hydrology by Smakhtin (2001).

Theoretical overview

Recurrent neural networks (RNNs)

In recent times, due to successful application of deep learning especially in the field of sequential prediction like statistical language modeling, chaotic time series, ecological modeling for dynamic systems control and finance and marketing motivated researchers to use deep learning for time series forecasting for hydrology events (Assaad et al. 2008; Cinar et al. 2017; Mikolov et al. 2010). The concept behind RNNs is to make use of arbitrarily input data over long sequences, such that it repeats the same task to every element in the sequence and output dependence on the previous computation. In more technicality, it consists of memory cell which captures information till sequence of training data completed. The architecture of RNN's varies according to its application, many-to-one model (useful when we want to predict at the current time step given all the previous inputs), many-to-many model (useful when we want to predict multiple future time steps at once given all the previous inputs) and several other variations. The final structure selection depends upon the problem statement which depends on phenomena. In our study, many-to-one for one-step ahead forecasting model is used, i.e., to predict the current month's low-flow value given all the previous month's low-flow values as input to the model.

RNN is a connectionist model described by interconnections and suitable for modeling temporal dependencies of unspecified duration inputs and the output using internal memory. The feature of RNN is that there is no instantaneous flow of information taking place among the neurons rather than the loops. Thus, it makes possible to keep the influence of the information for a variable at a particular period till the sequential time series complete. The memory of RNN is coded by the recurrent connections, and the output comes from each neuron itself (Assaad et al. 2008). Figure 2 shows a typical RNN structure in unrolled (network of complete sequence) form of full connected network.

where x_t is the input at a time t. The black square in Fig. 2 gets inputs from other neurons at a previous time step x_{t-1}, $s_t$is the hidden state at time step t. It is the "memory" of the network. s_t is calculated based on the previous hidden state and the input at the current step. s_t captures information about what happened in all the previous time steps and is given by Eq. (1)

$$s_t = f(Ax_t + Ws_{t-1}) \tag{1}$$

y_t is the output at step t. For example, if we wanted to predict the next sequence in a time series, it would be a vector of probabilities across the time series. The function f usually is a nonlinear activation function as tanh. RNN shares the same parameters (A, B, W) Fig. 2 across all steps performing the same task at each step, just with different inputs.

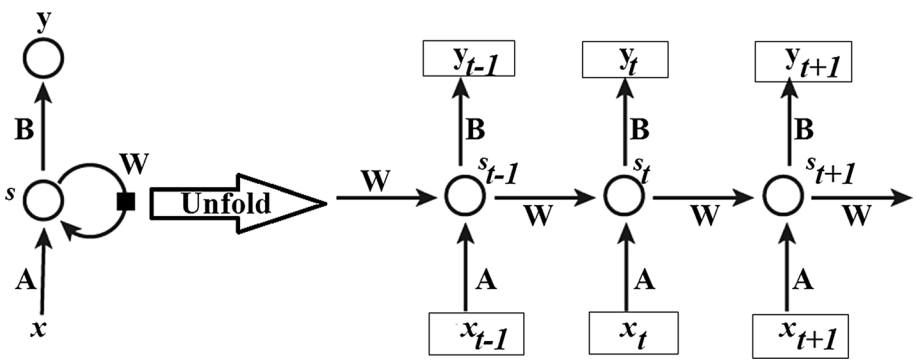


Fig. 2 Unfolding of Recurrent Neural Network (RNN) (LeCun et al. 2015)

Long short-term memory recurrent neural network (LSTM-RNN)

Although the RNNs have the higher competency to deal with nonlinear time series in an effective manner, but there are still some gradient issues to train long time lags, which specially needed for time series forecasting, more specifically for hydrology time series. Also, it has issue with predetermined time lags to learn temporal sequence processing and finding optimal time window size automatically (Gers 2001; Gers et al. 1999). Therefore, to overcome such limitations over RNN, an LSTM-RNN model is adopted in this paper to forecast low flow.

This state-of-the-art approach LSTM-RNN is proposed by Hochreiter and Schmidhuber (1997). The objective of this work is to develop a robust many-to-one LSTM model for hydrological time series. Similar to RNN, LSTM also consists a memory cell (Abidogun 2005) consisting four basic elements: an input gate, a neuron with a self-recurrent connection (a connection to itself), a forget gate and an output gate. The three nonlinear gates present in the block are the summation unit, which controls the inside-outside movement of information via activations cell through multiplications. This multiplication takes place at each input and output cell by their respective gates, while forget gate multiplies previous state (memory cell's self-recurrent connection) and allowing the cell to forget or remember its previous state using sigmoid activation function. In general, gate activation function ('f_t') is taken as logistic sigmoid, so that gate activation is between 0 (gate close) and 1 (gate open), whereas tanh or logistic sigmoid is for output activation function ('O_t') to overcome the vanishing gradient problem, whose second derivative can sustain for a long range before going to zero. Further, augmentation is possible which depends upon the different problem statement. The weights ('peephole' connection) join the cell to the gates, which is presented in Fig. 3, and the rest of the connection is unweighted (or equivalently, a fixed weight). The memory block output connects the rest of the network through output gate multiplication.

The model input is denoted as $x=(x_1, \ldots, x_j, \ldots, x_t)$, and the output sequence is denoted as $y = \left(x_{t+1}, \ldots, x_{t+i}, \ldots, x_{t+t'}\right)$ where t is prediction period and t' is the next time step prediction. In the case of low-flow prediction, x can be considered as historical input data, and y is the single lag period series. The goal of LSTM-RNN is to predict low-flow discharge in the next time step based on previous data and is calculated by the following equation:

$$i_t = \sigma\left(W_{ix} \cdot x_t + W_{ih} \cdot h_{t-1} + W_{ic} c_{t-1} + b_i\right) \tag{2}$$

$$f_t = \sigma\left(W_{fx} \cdot x_t + W_{fh} \cdot h_{t-1} + W_{fc} \cdot c_{t-1} + b_f\right) \tag{3}$$

$$c_t = f_j \cdot c_{t-1} + i_t \cdot g\left(W_{cx} \cdot x_t + W_{ch} \cdot h_{t-1} + b_c\right) \tag{4}$$

$$o_t = \sigma\left(W_{ox} \cdot x_t + W_{oh} \cdot h_{t-1} + W_{oc} \cdot c_t + b_o\right) \tag{5}$$

$$h_t = o_t \cdot h\left(c_t\right) \tag{6}$$

$$y_t = W_{yh} \cdot h_t + b_y \tag{7}$$

where σ denotes the sigmoid function

The memory block is outlined in a box and consists of an input gate, an output gate and a forget gate, where the outputs of three gates are, respectively, represented as follows: i_t, o_t, f_t. The activation vectors for each cell and memory block are, respectively, denoted as c_t and h_t. The weight

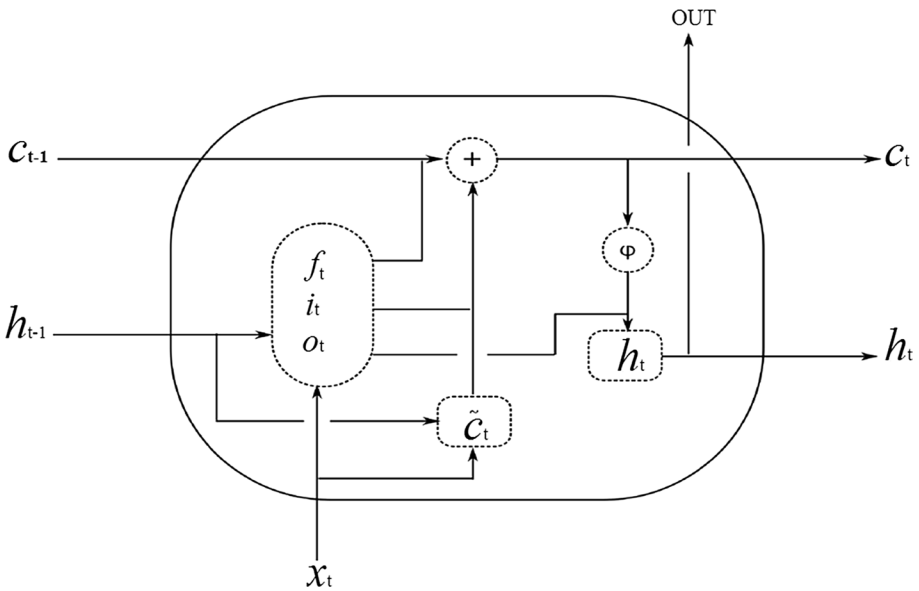


Fig. 3 LSTM memory block with one cell with three gated layers: forget gate f_t, input gate i_t and output gate o_t, controlling the activation of cells c_{t-1} and c_t

matrices W and bias vectors b are utilized to build connections between the input layer, output layer and the memory block.

Naïve method

Naïve forecast is the simple and very effective forecasting model and considered as benchmark against most of sophisticated models (Hyndman and Athanasopoulos 2018). Using the naïve approach, forecasts are produced that are equal to the last observed value. Naïve method is also useful for highly seasonal data. The low flows are highly seasonal as in this case, we set each forecast to be equal to the last observed value from the same season of the year (e.g., the same month of the previous year). Suppose the historical data be denoted by, $y_1, \ldots y_T$, then forecasting can be denoted by Eq. (8),

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)} \tag{8}$$

where $\hat{y}_{T+h|T}$ is a short-hand for the estimate of $\hat{y}_{T+h}$ based on the data $y_1, \ldots y_T$ and h is the forecast horizon, m is the seasonal period and k is the integer part of $(h-1)/m$ (i.e., the number of complete years in the forecast period prior to time $T+h$). This looks more complicated than it really is. For example, with monthly data, the forecast for all future February values is equal to the last observed February value. The interested reader can find more detail about the method in (Hyndman and Athanasopoulos 2018).

Model development and performance evaluation

The objective of study is to implement many-to-one LSTM-RNN model. The assembly of time delay model is developed using "Keras: The Python Deep learning library" (Chollet 2016). The low-flow dataset is divided into training and testing to build the models. The partitioning of the dataset (training and testing) generally varies with the problem of interest. Hence, there is no data division and depend upon problem. We used 70% data for training model and remaining 15% for validation and 15% for testing. Considering the view point of simple model, the RNN has a single-layered topology, i.e., one input layer, one hidden layer and one output layer. LSTM-RNN is constructed using one input layer, one LSTM layer with memory blocks and one output layer. Both the models were tested with 1, 2,3,4 and 5 neurons in the hidden layer with a lag of 1, 3, 6, 9 and 12, and the best configuration was selected based on the RMSE.

A simple data pre-processing step was tested to check model sensitivity, by adopting time series transformation to a logarithmic scale prior training the model (Bandara et al. 2017). Finally, in the post-processing stage, the forecasted value was back-transformed into their actual scale, by taking the exponent of each generated output value. This process is adopted to stabilize the variance of a time series. Firstly, we calculated the Q_{75} value from the original discharge data. Then, we transformed the Q_{75}, i.e., the low-flow time series in this study using natural logarithm. All the Q_{75} low-flow time series discharge data are greater than one so while we do log transform of the Q_{75} time series, there is no chance of getting any value negative. The selection of model architecture includes a selection of model input. Therefore, various time steps were tested in between 1 and 12, and finally 12 time steps were fixed for building the model (Ouyang and Lu 2018). Hence, as input, vector 3D array (number of sample = 468, number of time steps = 12, output = 1) was used to train (70% of the sample data) the model in both cases; in addition, the loss function mean square error and optimizer Adam were used to compile the model as the final step in building the model with 2000 epochs.

Some techniques are recommended for HTS forecasting model performance evaluation according to the published literature related to calibration, validation and application of hydrological models (Schoups et al. 2008). Four performance evaluation criteria used in this study are computed in the following section.

The coefficient of correlation (R)

$$R = \frac{\sum_{i=1}^{N}\left(Q_{75,\text{obs}} - \bar{Q}_{75,\text{obs}}\right)\left(Q_{75,\text{for}} - \bar{Q}_{75,\text{for}}\right)}{\sqrt{\sum_{i=1}^{N}\left(Q_{75,\text{obs}} - \bar{Q}_{75,\text{obs}}\right)^2}\sqrt{\sum_{i=1}^{N}\left(Q_{75,\text{for}} - \bar{Q}_{75,\text{for}}\right)^2}} \tag{9}$$

Root-mean-squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(Q_{75,\text{obs}} - Q_{75,\text{for}}\right)^2}{N}} \tag{10}$$

Nash–Sutcliffe efficiency coefficient (E_{NS})

$$E_{\text{NS}} = 1 - \left[\frac{\sum_{i=1}^{N}\left(Q_{75,\text{obs}} - Q_{75,\text{for}}\right)^2}{\sqrt{\sum_{i=1}^{N}\left(Q_{75,\text{obs}} - \bar{Q}_{75,\text{obs}}\right)^2}}\right], \quad -\infty \le E_{\text{NS}} \le 1 \tag{11}$$

Mean absolute error (MAE)

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|Q_{75,\text{obs}} - Q_{75,\text{for}}\right| \tag{12}$$

where $Q_{75,\text{obs}}$ = observed Q_{75}; $Q_{75,\text{for}}$ = forecasted Q_{75}; $\bar{Q}_{75,\text{obs}}$ = average observed; $\bar{Q}_{75,\text{for}}$ = average forecasted Q_{75}; N = number of data points.

Table 1 Forecasting models are implemented using R, RMSE, E_{NS} and MAE values during validation period

	LSTM- RNN	RNN	Naïve
RMSE	0.487	0.561	0.793
E_{NS}	0.878	0.843	0.704
R	0.943	0.935	0.866
MAE	0.361	0.391	0.617

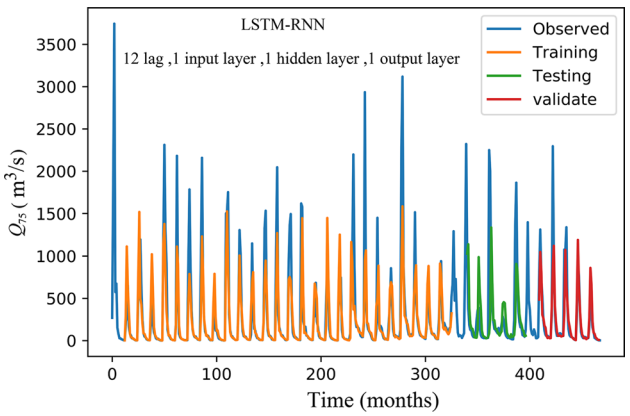


Fig. 4 Training (from June 1971 to September 1998), testing (from October 1998 to July 2004) and validation (from August 2004 to May 2010) for LSTM-RNN

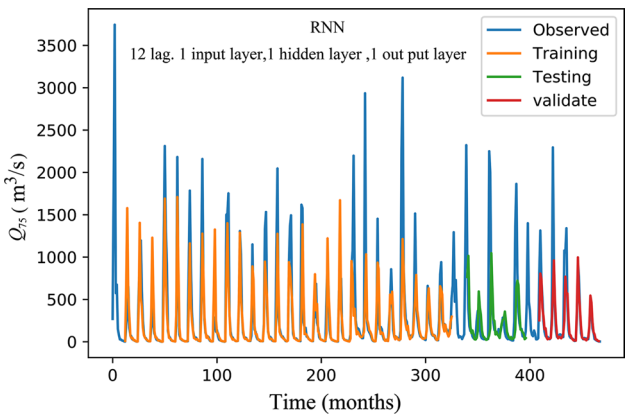


Fig. 5 Training (from June 1971 to September 1998), testing (from October 1998 to July 2004) and validation (from August 2004 to May 2010) for RNN

We used 70% data for training model and remaining 15% for validation and 15% for testing, i.e., from June 1971 to September 1998 for training, from October 1998 to July 2004 for testing and from August 2004 to May 2010 for validation in case of LSTM-RNN and RNN, while for naïve method, we used June 1971 to July 2004 for training and August 2004 to May 2010 for validation.

Result and discussion

As earlier stated, the main purpose of this study is to inspect the suitability of LSTM-RNN for low-flow HTS forecasting and compare with RNN and naïve method.

The outcomes accomplished in this study suggest that the LSTM-RNN method is an effective technique to model the monthly low-flow discharge time series and can give significant prediction performance than the traditional RNN and benchmark naïve method for time series approaches. The results indicate that the best performance can be obtained by LSTM-RNN with a lag of 12 considering RMSE evaluation criteria during the validation phases (Table 1). For RNN, the best performance is also obtained at lag 12.

Overall, the RNN and LSTM-RNN models can give good prediction performance and could be successfully applied to establish the forecasting models that could provide accurate and reliable monthly low-flow prediction. But long-range dependence (also known as the Hurst phenomenon) cannot be modeled by AI learning regression models. This modeling can be made within stochastic frameworks and is mostly important for probabilistic forecasting (long-range dependence leads to wider prediction intervals)(Tyralis and Koutsoyiannis 2014). The results suggest that the LSTM-RNN model was superior to the RNN for monthly low-flow time series forecasting at Basantapur station in the Mahanadi River basin. From Table 1, it is observed that the LSTM–RNN has outperformed in forecasting low flow for the representative station in the Mahanadi River basin compared to RNN and naïve method. The Nash–Sutcliffe model efficiency coefficient (E_{NS}) is 0.889 for LSTM-RNN, 0.825 for RNN and 0.704 for naïve method indicating LSTM-RNN has better predictive power than other two methods. When comparing forecast methods applied to a single time series or to several time series with the same units, the MAE is popular as it is easy to both understand and compute. A forecast method that minimizes the MAE will lead to forecasts

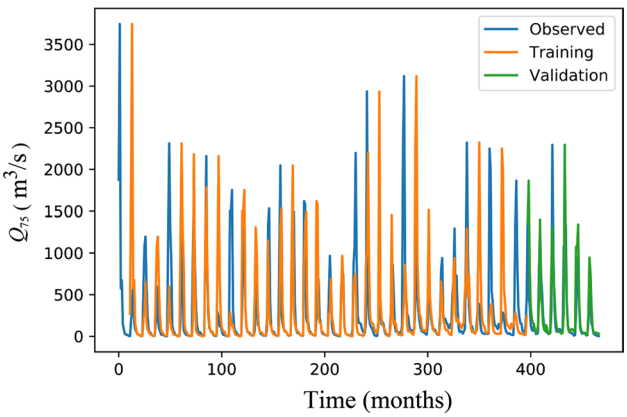


Fig. 6 Training (from June 1971 to July 2004) and validation (from August 2004 to May 2010) for naïve method

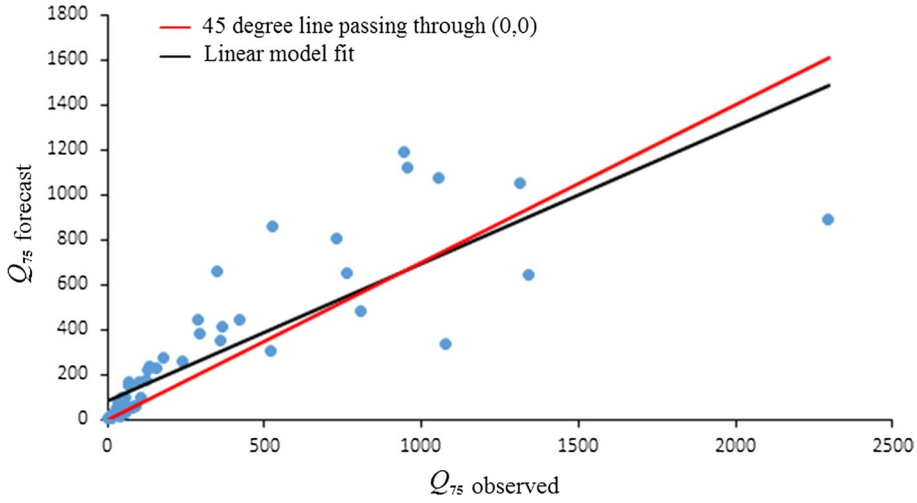


Fig. 7 Scatterplot of the forecasted versus their corresponding observed Q_{75} values for LSTM-RNN in the validation period

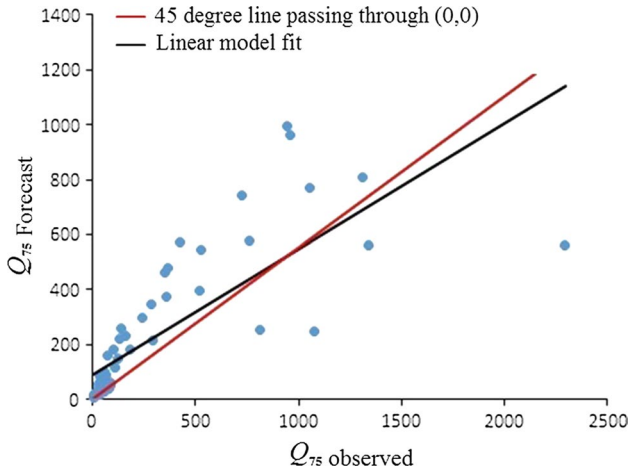


Fig. 8 Scatterplot of the forecasted versus their corresponding observed Q_{75} values for RNN in the validation period

of the median, while minimizing the RMSE will lead to forecasts of the mean. It is clearly seen from Table 1 that LSTM–RNN has MAE 0.361 which is better than RNN's MAE 0.391. For this case study of forecasting low flow for the station Basantapur in the Mahanadi River basin, the LSTM-RNN LSTM-RNN's performance is increased by 7.67% than the performance of RNN. However, a large comparison study can be used to provide generalized results about the forecasting performance of black box models (e.g., LSTM-RNN, RNN, ANN) (Papacharalampous et al. 2018a, c, 2019; Tyralis and Papacharalampous 2018). Figures 4 and 5 show the LSTM-RNN plots of data during training, testing and validation period, and Fig. 6 showing the training and validation results for naïve method. The scatterplot of the forecasted versus observed Q_{75} during the validation period is shown in Figs. 7, 8 and 9, respectively, for LSTM-RNN, RNN and naïve method.

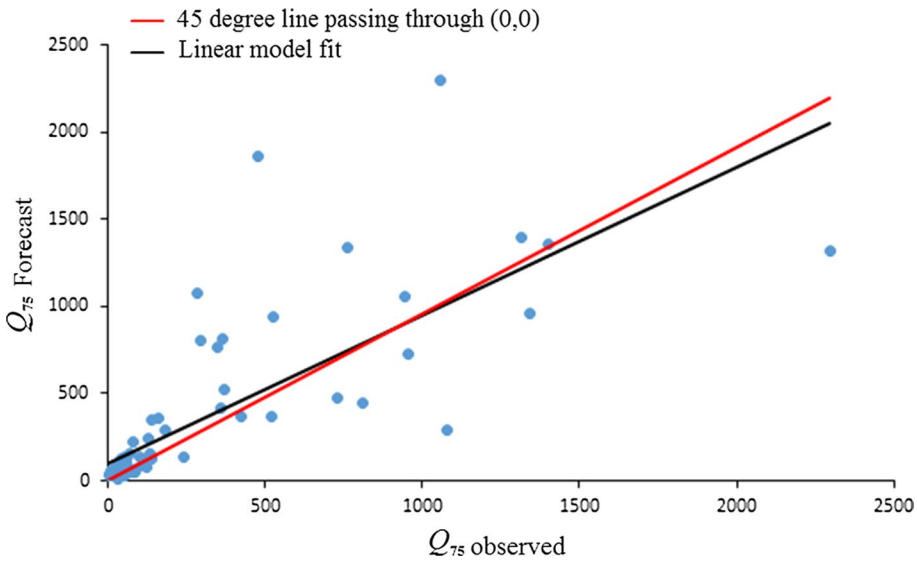


Fig. 9 Scatterplot of the forecasted versus their corresponding observed Q_{75} values for naïve method in the validation period

Conclusions

An effort was made in this paper to explore the suitability of LSTM-RNN over RNN for hydrological time series. As a case study, monthly low-flow discharge data are used to implement the forecasting models. The standard statistical performance evaluation measures are adopted to evaluate the performances of various models applied.

LSTM-RNN's ability to forget, remember and update the information pushes it one-step ahead of RNN. The results obtained in this study indicate that the LSTM-RNN method can be used to model low-flow HTS at Basantapur station in the Mahanadi River basin, India, and can give satisfactory performance over RNN and naïve method. LSTM-RNN is well-suited to learn from experience to classify, process and predict time series given time lags of unknown size and bound between important events. Time series prediction involves processing of patterns that evolve the appropriate response at a particular point in time and depends not only on the current value of the observable but also in the past.

Therefore, the results of the study are encouraging, and the authors advocate that LSTM-RNN approaches can be used in modeling the low-flow hydrological time series for the selected station, and this may provide some ideas for researchers and engineers who apply data-driven AI approaches for modeling low-flow hydrological time series forecasting. This work recommends the performance of such model can be improved for low-flow hydrological time series forecasting by adopting several stacked layers (multiple hidden LSTM layers) and a GRU (Gated Recurrent Unit) LSTM that would be scope of this work.

Compliance with ethical standards

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

References

Abidogun OA (2005) Data mining, fraud detection and mobile telecommunications: call pattern analysis with unsupervised neural networks. University of the Western Cape, Cape Town

Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H (2018) State-of-the-art in artificial neural network applications: a survey. Heliyon 4:e00938

Ahn KH, Palmer RN (2016) Use of a nonstationary copula to predict future bivariate low flow frequency in the Connecticut river basin. Hydrol Process 30:3518–3532

Arena C, Cannarozzo M, Mazzola MR (2006) Multi-year drought frequency analysis at multiple sites by operational hydrology–a comparison of methods. Phys Chem Earth Parts A/B/C 31:1146–1163

ASCE (2000a) Artificial neural networks in hydrology. I: Preliminary concepts. J Hydrol Eng 5:115–123

ASCE (2000b) Artificial neural networks in hydrology. II: Hydrologic applications. J Hydrol Eng 5:124–137

Assaad M, Boné R, Cardot H (2008) A new boosting algorithm for improved time-series forecasting with recurrent neural networks. Inf Fusion 9:41–55

Atiya AF, El-Shoura SM, Shaheen SI, El-Sherif MS (1999) A comparison between neural-network forecasting techniques-case study: river flow forecasting. IEEE Trans Neural Netw 10:402–409

Bandara K, Bergmeir C, Smyl S (2017) Forecasting across time series databases using long short-term memory networks on groups of similar series. arXiv preprint arXiv:171003222

Beven KJ (2012) Rainfall-runoff modelling: the primer. Wiley, New York

Box G, Jenkins G (1970) Time series analysis; forecasting and control. Holden-Day, San Francisco

Carlson RF, MacCormick A, Watts DG (1970) Application of linear random models to four annual streamflow series. Water Resour Res 6:1070–1078

Chang F, Chang LC, Huang HL (2002) Real-time recurrent learning neural network for stream-flow forecasting. Hydrol Process 16:2577–2588

Chen H-L, Rao AR (2002) Testing hydrologic time series for stationarity. J Hydrol Eng 7:129–136

Cheng C, Chau K, Sun Y, Lin J (2005) Long-term prediction of discharges in Manwan Reservoir using artificial neural network models. Adv Neural Netw 2005:975

Chollet F (2016) Keras. https://github.com/fchollet/keras/tree/master/keras

Cinar YG, Mirisaee H, Goswami P, Gaussier E, Aït-Bachir A, Strijov V (2017) Position-based content attention for time series forecasting with sequence-to-sequence RNNs. In: International conference on neural information processing. Springer, pp 533–544

Demirel MC, Booij MJ, Hoekstra AY (2013) Identification of appropriate lags and temporal resolutions for low flow indicators in the River Rhine to forecast low flows with different lead times. Hydrol Process 27:2742–2758

Dimitriadis P, Koutsoyiannis D (2015) Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst-Kolmogorov processes. Stoch Environ Res Risk Assess 29:1649–1669

Dimitriadis P, Koutsoyiannis D, Tzouka K (2016) Predictability in dice motion: how does it differ from hydro-meteorological processes? Hydrol Sci J 61:1611–1622

Dracup JA, Lee KS, Paulson EG (1980) On the definition of droughts. Water Resour Res 16:297–302

Fang K, Shen C, Kifer D, Yang X (2017) Prolongation of SMAP to spatiotemporally seamless coverage of continental US using a deep learning neural network. Geophys Res Lett 44:11–15

Firat M, Güngör M (2007) River flow estimation using adaptive neuro fuzzy inference system. Math Comput Simul 75:87–96

Firat M, Güngör M (2008) Hydrological time-series modelling using an adaptive neuro-fuzzy inference system. Hydrol Process 22:2122–2132

Gárfias-Soliz J, Llanos-Acebo H, Martel R (2010) Time series and stochastic analyses to study the hydrodynamic characteristics of karstic aquifers. Hydrol Process 24:300–316

Gers F (2001) Long short-term memory in recurrent neural networks Unpublished PhD dissertation. Ecole Polytechnique Fédérale de Lausanne, Lausanne

Gers FA, Schmidhuber J, Cummins F (1999) Learning to forget: continual prediction with LSTM. 850–855

Giuntoli I, Renard B, Vidal J-P, Bard A (2013) Low flows in France and their relationship to large-scale climate indices. J Hydrol 482:105–118

Gustard A, Demuth S (2009) Manual on low-flow estimation and prediction. Opera

Hipel KW, McLeod AI (1994) Time series modelling of water resources and environmental systems, vol 45. Elsevier, Amsterdam

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780
Hu T, Lam K, Ng S (2001) River flow time series prediction with a range-dependent neural network. Hydrol Sci J 46:729–745
Hyndman RJ, Athanasopoulos G (2018) Forecasting: principles and practice. OTexts, Melbourne
Jain A, Kumar AM (2007) Hybrid neural network models for hydrologic time series forecasting. Appl Soft Comput 7:585–592
Jha R, Smakhtin V (2008) A review of methods of hydrological estimation at ungauged sites in India, vol 130. IWMI, Colombo
Jha R, Sharma K, Singh V (2008) Critical appraisal of methods for the assessment of environmental flows and their application in two river systems of India. KSCE J Civ Eng 12:213–219
Keskin ME, Taylan D, Terzi O (2006) Adaptive neural-based fuzzy inference system (ANFIS) approach for modelling hydrological time series. Hydrol Sci J 51:588–598
Kişi Ö (2007) Streamflow forecasting using different artificial neural network algorithms. J Hydrol Eng 12:532–539
Kisi O, Cimen M (2011) A wavelet-support vector machine conjunction model for monthly streamflow forecasting. J Hydrol 399:132–140
Komorník J, Komorníková M, Mesiar R, Szökeová D, Szolgay J (2006) Comparison of forecasting performance of nonlinear models of hydrological time series. Phys Chemi Earth Parts A/B/C 31:1127–1145
Koutsoyiannis D, Langousis A (2011) Precipitation, Treatise on Water Science, edited by P. Wilderer and S. Uhlenbrook, 2, 27–78. Academic Press, Oxford
Koutsoyiannis D, Yao H, Georgakakos A (2008) Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods/Prévision du débit du Nil à moyen terme: une comparaison de méthodes stochastiques et déterministes. Hydrol Sci J 53:142–164
Laaha G, Blöschl G (2005) Low flow estimates from short stream flow records—a comparison of methods. J Hydrol 306:264–286
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436
Lin J-Y, Cheng C-T, Chau K-W (2006) Using support vector machines for long-term discharge prediction. Hydrol Sci J 51:599–612
Mikolov T, Karafiát M, Burget L, Cernocký J, Khudanpur S (2010) Recurrent neural network based language model. In: Interspeech, p 3
Nayak PC, Sudheer K, Rangan D, Ramasastri K (2004) A neuro-fuzzy computing technique for modeling hydrological time series. J Hydrol 291:52–66
Ouyang Q, Lu W (2018) Monthly rainfall forecasting using echo state networks coupled with data preprocessing methods. Water Resour Manag 32:659–674
Papacharalampous G, Tyralis H, Koutsoyiannis D (2018a) One-step ahead forecasting of geophysical processes within a purely statistical framework. Geosci Lett 5:12
Papacharalampous G, Tyralis H, Koutsoyiannis D (2018b) Predictability of monthly temperature and precipitation using automatic time series forecasting methods. Acta Geophys 66(4):807–831
Papacharalampous G, Tyralis H, Koutsoyiannis D (2018c) Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: a multiple-case study from Greece. Water Resour Manag 32:5207–5239
Papacharalampous GA, Tyralis H, Koutsoyiannis D (2019) Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. Stoch Environ Res Risk Assess. https://doi.org/10.1007/s00477-018-1638-6
Pyrce R (2004) Hydrological low flow indices and their uses Watershed Science Centre, (WSC) Report
Sahoo S, Russo T, Elliott J, Foster I (2017) Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US. Water Resour Res 53(5):3878–3895
Sahoo BB, Jha R, Singh A, Kumar D (2018) Application of support vector regression for modeling low flow time series. KSCE J Civ Eng 23(2):923–934
Salas JD (1993) Analysis and modeling of hydrologic time series. In: Maidment DR (ed) Handbook of hydrology, vol 19. McGraw Hill, New York, pp 19.1–19.72
Sang Y-F (2013) A review on the applications of wavelet transform in hydrology time series analysis. Atmos Res 122:8–15
Sang Y-F, Wang D, Wu J-C, Zhu Q-P, Wang L (2009) The relation between periods' identification and noises in hydrologic series data. J Hydrol 368:165–177
Schoups G, Van de Giesen N, Savenije H (2008) Model complexity control for hydrologic prediction. Water Resources Res 44:W00B03
Sivapragasam C, Liong S-Y, Pasha M (2001) Rainfall and runoff forecasting with SSA–SVM approach. J Hydroinform 3:141–152
Sivapragasam C, Vincent P, Vasudevan G (2007) Genetic programming model for forecast of short and noisy data. Hydrol Process 21:266–272
Smakhtin V (2001) Low flow hydrology: a review. J Hydrol 240:147–186
Srikanthan R, McMahon T (2001) Stochastic generation of annual, monthly and daily climate data: a review. Hydrol Earth Syst Sci Discuss 5:653–670
Tegos A, Schlüter W, Gibbons N, Katselis Y, Efstratiadis A (2018) Assessment of environmental flows from complexity to parsimony—lessons from Lesotho. Water 10:1293
Toth E, Brath A, Montanari A (2000) Comparison of short-term rainfall prediction models for real-time flood forecasting. J Hydrol 239:132–147
Tyralis H, Koutsoyiannis D (2014) A Bayesian statistical model for deriving the predictive distribution of hydroclimatic variables. Clim Dyn 42:2867–2883
Tyralis H, Papacharalampous G (2017) Variable selection in time series forecasting using random forests. Algorithms 10:114
Tyralis H, Papacharalampous GA (2018) Large-scale assessment of Prophet for multi-step ahead forecasting of monthly streamflow. Adv Geosci 45:147–153
Wang W-C, Chau K-W, Cheng C-T, Qiu L (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J Hydrol 374:294–306
WMO (2008) Manual on low flow estimation and prediction. WMO, Geneva
Wunsch A, Liesch T, Broda S (2018) Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX). J Hydrol 567:743–758
Xu L, Chen N, Zhang X, Chen Z (2018) An evaluation of statistical, NMME and hybrid models for drought prediction in China. J Hydrol 566:235–249
Yaseen ZM, El-Shafie A, Jaafar O, Afan HA, Sayl KN (2015) Artificial intelligence based models for stream-flow forecasting: 2000–2015. J Hydrol 530:829–844
Yaseen ZM, Jaafar O, Deo RC, Kisi O, Adamowski J, Quilty J, El-Shafie A (2016) Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. J Hydrol 542:603–614
Yaseen ZM, Fu M, Wang C, Mohtar WHMW, Deo RC, El-Shafie A (2018) Application of the hybrid artificial neural network coupled with rolling mechanism and grey model algorithms for streamflow forecasting over multiple time horizons. Water Resour Manag 32(5):1883–1899
Zhang D, Lindholm G, Ratnaweera H (2018a) Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. J Hydrol 556:409–418

Zhang J, Zhu Y, Zhang X, Ye M, Yang J (2018b) Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. J Hydrol 561:918–929

Zounemat-Kermani M, Teshnehlab M (2008) Using adaptive neuro-fuzzy inference system for hydrological time series prediction. Appl Soft Comput 8:928–936