



# Automatic lithology modelling of coal beds using the joint interpretation of principal component analysis (PCA) and continuous wavelet transform (CWT)

THINESH KUMAR<sup>1</sup>, SEELAM NARESH KUMAR<sup>1</sup> and G SRINIVASA RAO<sup>2,\*</sup> 

<sup>1</sup>*CMPDIL, Coal India Limited, Ranchi 834 008, India.*

<sup>2</sup>*Department of Applied Geophysics, IIT (ISM) Dhanbad, Dhanbad 826 004, India.*

\*Corresponding author. e-mail: [gsrao@iitism.ac.in](mailto:gsrao@iitism.ac.in)

MS received 24 December 2021; revised 10 August 2022; accepted 24 August 2022

Identification of thin interbedded non-coal bands and coal seams with varying carbon contents within a coal seam is of paramount interest in coal exploration due to its banded nature. The manual interpretation and conventional modelling based on Fourier/Walsh transform techniques fail to derive such information accurately from geophysical well log data due to its non-stationary nature. The present study proposes a combined principal component analysis (PCA) and continuous wavelet transform (CWT) algorithm for automatic lithological modelling of geophysical well log data. In the first step of well log lithology modelling, a median filter is applied on well log data to preserve the thinner beds and other valuable geological signatures of coal seams. In the second step, the filtered log data is subjected to PCA, and the variance level of PC scores is determined to study the physical relationship of input parameters. The third step is to apply CWT on the selected PC scores and determine lithological discontinuities from the modulus maxima lines drawn on the wavelet scalogram. For filling the lithology skeleton with the proper interpretation, a database is also created by correlating the selected PC score values with input parameters. We have applied the proposed algorithm to gamma ray, density, and resistivity logs of two boreholes located in the Bistrampur and Jharia coalfields of eastern India. The results of the proposed PCA-CWT based modelling match well with core data and manual interpretations of the boreholes. At a few depth ranges, the proposed algorithm also reveals some additional lithological discontinuities that were not mapped in the core data. The study further conveys that PCA-CWT-based lithological modelling of geophysical logs is helpful to pace up the exploration work in coal blocks with poor core recovery.

**Keywords.** Principal component analysis (PCA); continuous wavelet transform (CWT); lithology modelling; well log; coal exploration.

---

## 1. Introduction

In coal exploration, the detection of coal seams of varying thickness (thick/thin) as well as fine to prominent variations within thick coal seams is important to obtain ash, moisture, and volatile

material percentage present in coal beds down the earth. More significantly, due to the banded occurrence of Indian coal seams, detailing the fossil fuel is crucial in terms of non-coal bands and coal seams with varying carbon contents. Coring boreholes can provide these details to an extent but in

the case of structurally disturbed areas, obtaining core samples with proper recovery is difficult. On the other hand, geophysical well log data possesses immense information about these formation properties and geological interfaces (Crain 1986; Dewan 1983). However, correctly identifying the depth of different lithological units from the well log signal is an extremely tedious job due to its nonlinear characteristics of varying wavelengths and frequencies.

Traditionally, cross-plotting of different log parameters such as gamma ray, density, sonic, neutron, and resistivity are used for differentiating carbonaceous beds (coal, shaly coal, and carbonaceous shale) and non-coal litho-units (sandstone, shaly sand, sandy shale, and shale) (Chatterjee and Paul 2012; Anyiam *et al.* 2018). The main demerits of this technique are the precise identification of different carbonaceous and non-coal beds in an overlapping data cloud and the necessity for more log parameters. As an extension of this technique, modules are incorporated into the commercial software for obtaining litholog as the final output. However, the outcomes are subjected to the boundaries drawn by the user based on the cut-off values of carbonaceous beds in different depth ranges without showing details on the varying carbon contents in those depth ranges.

Over the past few decades, principal component analysis (PCA) (Elek 1988; Lim *et al.* 1998), wavelet transform and its combination with Fourier transform (FT) and short time Fourier transform (STFT) (Alvarez *et al.* 2003; Pan *et al.* 2008; Coconi-Morales *et al.* 2010; Ouadfeul and Aliouane 2011; Chandrasekhar and Rao 2012; Javid and Tokmechi 2012) were widely used in the hydrocarbon exploration to map the lithological discontinuities from geophysical logs (density, self-potential, gamma ray, and resistivity). Besides lithology identification, wavelet transform was applied on resistivity logs to identify reservoir fluid types (Yue *et al.* 2006). Wavelet decomposed density log curves along with interpreted resistivity image logs is used to determine fracture zone interval (Zhang *et al.* 2011). Wavelet transform is also applied on 1D and 2D permeability data to identify boundaries, faults, and fractures (Panda *et al.* 2000). Comparatively, studies based on PCA and CWT techniques are very limited in coal exploration. Ren *et al.* (2018) applied the principal component analysis on well logs for coal texture identification. Recently, Chen *et al.* (2021) used wavelet transform and linear discrimination

analysis to interpret the thin-layer coal texture from well log data. Lie *et al.* (2020) have used multi-scale wavelet analysis along with PCA on density, resistivity, gamma ray, spontaneous potential, and caliper logs for the reconstruction of crucial well logs such as neutron and sonic for coal bed methane (CBM) exploration in Qinshui basin, China.

The present study attempted to model the lithological discontinuities from geophysical logs based on the PCA and continuous wavelet transform (CWT). For this purpose, the proposed technique was applied to gamma ray, density, and resistivity log datasets recorded in Jharia and Bistrampur coalfields located in eastern India. Initially, well log datasets are filtered and zero-centered to avoid spurious spikes not related to the sub-surface geology and single- or double-parameter dominance. Subsequently, these log datasets are subjected to PCA and CWT to prepare the lithology skeleton. Finally, the obtained results from PCA-CWT are validated against core data and manual interpretations of the boreholes.

## 2. Methodology

### 2.1 Principal component analysis (PCA)

The principal component analysis (PCA) is a dimensionality reduction technique. It detects the linear relationship between different variables and replaces a group of correlated data values through uncorrelated data values called principal scores/components. Here, the first principal component is derived along the direction of maximum variation in the data cloud. While the second component is derived along the next direction that contains maximum variation in the data cloud. These components are orthogonal to each other, confirming less or no redundancy in the data derived. The above concept of PCA can be visualised through the illustration using only two parameters, as shown in figure 1. For a more extensive set of log parameters, mathematics remains unchanged, but it is challenging to visualise the geometry (Lim 2003).

The method of PCA involves three steps. The first step is to standardise the data sets by subtracting each of the values from their respective means and dividing by the standard deviation of the time series to reduce the dominance of a single parameter (Lim 2003) and to identify the

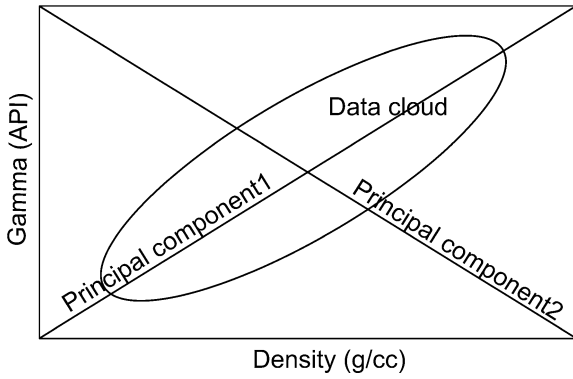


Figure 1. Principal components of two-dimensional systems (after Kassenaar 1991).

correlation between them. The following equations can represent this.

$$\text{Var}(x) = \frac{\sum_{i=1}^n (x_i - X)(x_i - X)}{n}, \quad (1)$$

$$\text{Cov}(x, y) = \frac{\sum_{i=1}^n (x_i - X)(y_i - Y)}{n}. \quad (2)$$

$X$  and  $Y$  are the respective mean values of the time series. A useful way to identify covariance values between all the variables is to put them in matrix form.

$$C^{n \times n} = \left( C_{ij}, C_{ij} = \text{cov}(Dim_i, Dim_j) \right), \quad (3)$$

where  $C^{n \times n}$  is a matrix of  $n$  rows and  $n$  columns, and  $Dim_x$  is the  $x$ th dimension. The matrix form for calculating the covariance value of three different variables is as follows:

$$C = \begin{pmatrix} \text{cov}(x, x) & \text{cov}(x, y) & \text{cov}(x, z) \\ \text{cov}(y, x) & \text{cov}(y, y) & \text{cov}(y, z) \\ \text{cov}(z, x) & \text{cov}(z, y) & \text{cov}(z, z) \end{pmatrix}. \quad (4)$$

The main diagonal states the covariance of the same variable, whereas the upper and lower triangle of the matrix are the covariance between different values (also  $\text{cov}(x, y) = \text{cov}(y, x)$ ).

In the second step, the eigenvalues and eigenvectors are calculated from the covariance matrix, which throws light on patterns present in the data. Once eigenvectors are found from the covariance matrix, the next step is to arrange those values/scores in the highest to lowest order, i.e., in the order of significance. Typically, the first two components contain maximum variances of the data of interest needed for evaluation. In comparison, the other components are less significant (i.e., components generated from lesser eigenvalues), and these

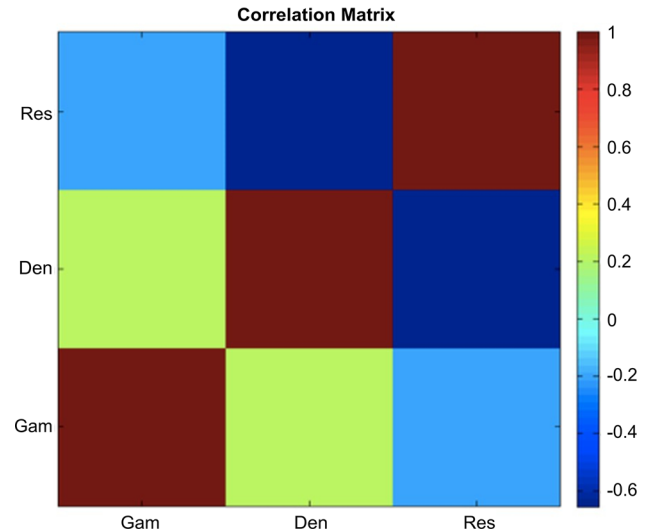


Figure 2. Covariance matrix between gamma, density, and resistivity parameters.

can be avoided without losing much information. As the PCA depends on the covariance matrix of the datasets, it facilitates in identifying the common physical property orientation (figure 2). Therefore, the cross-plots of PC scores against original data sets help us to interpret the underlying physical properties and trends in the dimensionless logs containing the maximum variance of the input logs.

## 2.2 Continuous wavelet transform (CWT)

Continuous wavelet transform is a time-frequency localisation tool used in signal processing for analysing non-stationary signals. It can be defined as the convolution product of the signal with the mother wavelet (Goupillaud *et al.* 1984)

$$C_s(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} S(z) \psi^* \left( \frac{z-b}{a} \right) dz, \quad (5)$$

where  $S(z)$  is the signal considered for analysis, and  $\psi^*(z)$  is the complex conjugate of the mother wavelet. For most practical purposes, the mother wavelet is usually expected to have zero mean, as defined in the following equation (6) (Daubechies 1992).

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0. \quad (6)$$

Here the purpose of normalising constant  $\frac{1}{\sqrt{a}}$  is to keep the same energy level of wavelet coefficients in all scales (Daubechies 1992).

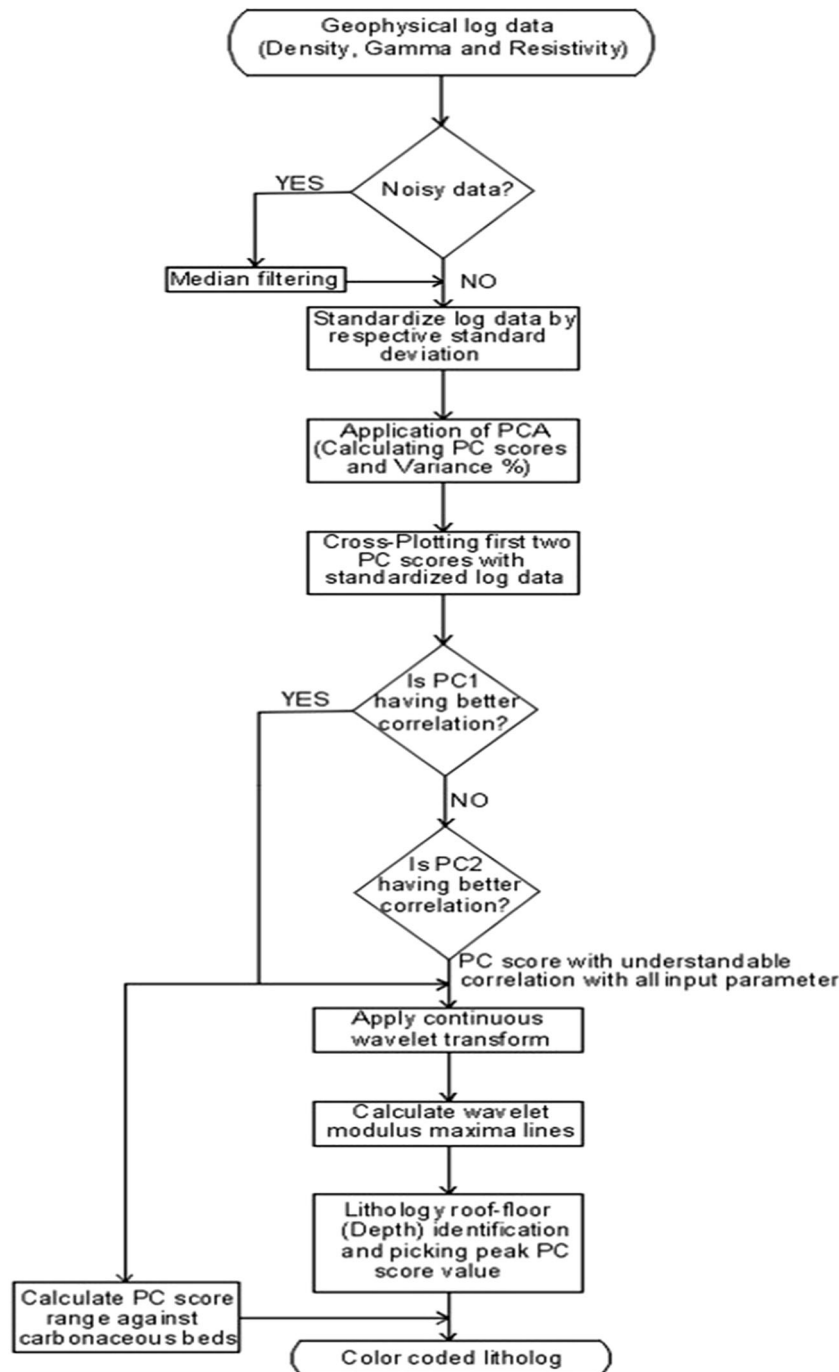


Figure 3. Flowchart for automatic lithology segmentation and generating colour-coded litholog.

CWT possesses high sampling on the lower scale and coarse sampling on the larger scale. These aspects are explained in terms of  $a \in R$  &  $b \in R$ , which are called scale (inversely proportional to frequency) and translation (directly proportional to time). When the value of the scale increases, the wavelet takes into account only the long-time behaviour of the signal  $S(z)$ , and when the value of the scale decreases, the wavelet focuses on small-

scale features of the signal (Farge 1992; Kumar and Foufoula-Georgiou 1997). Further, the understanding and enhancing time-frequency localisation accuracy of CWT entirely depends on the choice of the mother wavelet (Kumar and Foufoula-Georgiou 1997; Polikar 1999). This depends on the kind of information one wants to extract from the signal (Farge 1992). The most desirable analysing wavelet for any signal whose frequency



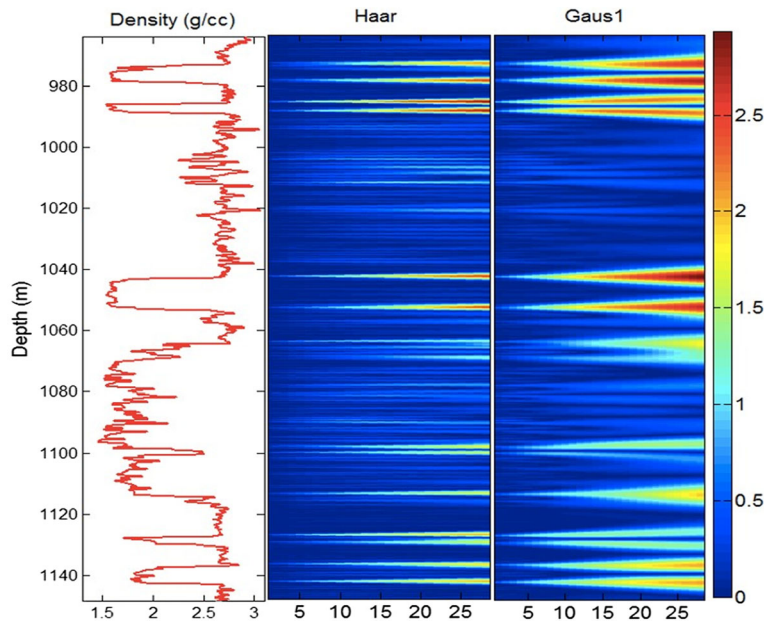


Figure 4. Illustrates how the Haar wavelet scalogram meticulously demarcated all thin beds without much smearing than the Gaus1 mother wavelet.

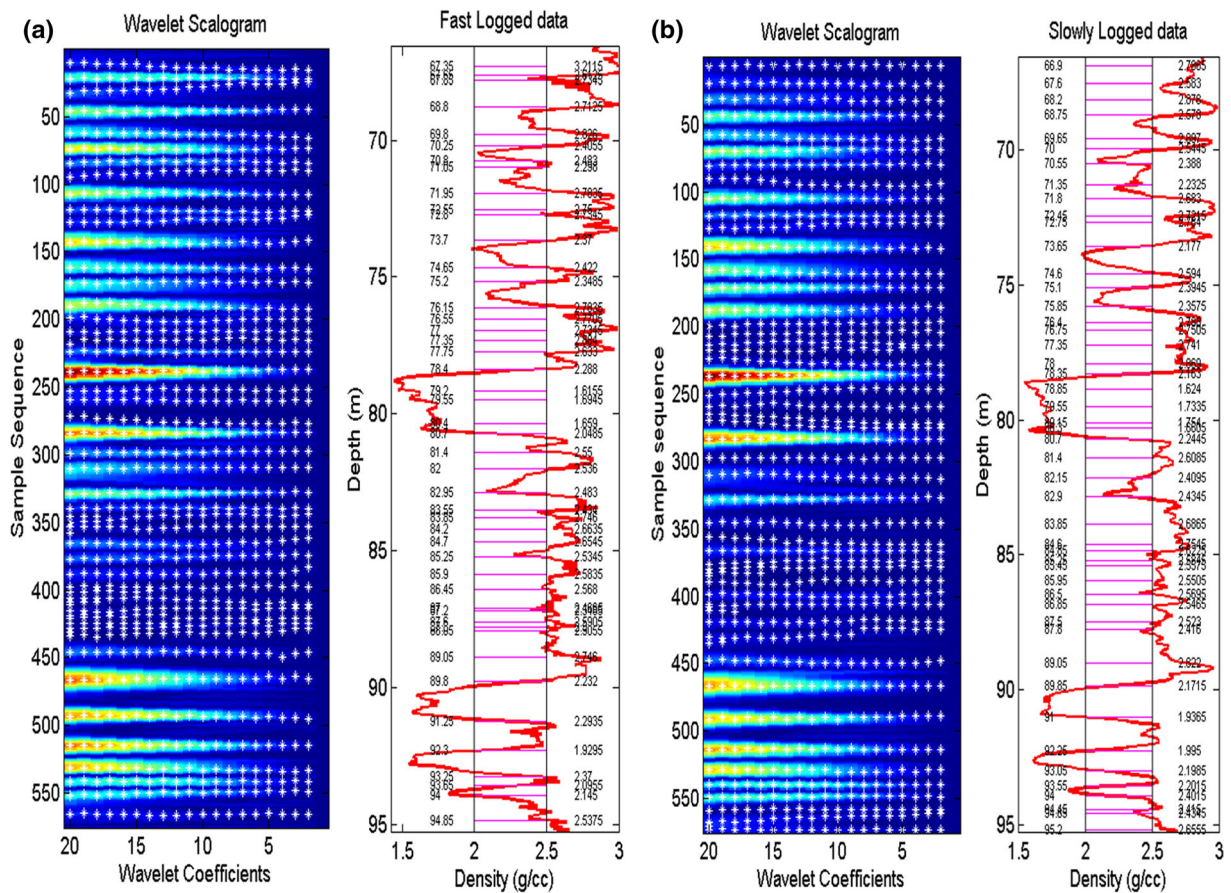


Figure 5. Haar wavelet scalogram of (a) slowly logged density log and (b) fast logged density log.

varies over time (non-stationary) should be orthogonal, local, and universal (Kumar and Foufoula-Georgiou 1997).

### 2.3 Combined PCA and CWT algorithm and procedure for automatic lithological modelling

The steps followed for the application of combined PCA and CWT algorithm on well log data for automatic lithological modelling are explained below (figure 3):

- (1) Standardise the data sets by dividing them by their standard deviation values.
- (2) Datasets are median filtered with filter length from 3 to 6 and compared with original data for further refinement if required.
- (3) Filtered data is subjected to PCA and generates a Pareto chart to check the variance level of PC scores.
- (4) Cross-plotting PC scores with log datasets for identifying physical properties and trends possessed by individual PC scores. PC score with better correlation with original log datasets and variance level is subjected to wavelet transform.
- (5) Wavelet modulus maxima values are traced automatically. Thereby its corresponding roof-floor depth and peak PC values for roof-floor depth are picked.
- (6) Database preparation by picking PC score ranges for different carbonaceous beds against conventional density, gamma, and resistivity ranges.
- (7) Based on the PC scores obtained in step-6, colour-coded litholog was generated by assigning coal = 1, shalycoal = 2, carbshale = 3, and non-coal = 4.

#### 2.3.1 Selection of mother wavelet

The main purpose of using CWT in this study is automatic lithology demarcation of well log signals. There are two critical factors that should be considered to arrive at an effective solution. One factor is the shape of well log signal against lithology and the second factor is the logging speed. In order to select the optimum mother wavelet, these two factors were assessed in the present study. Most of the earlier researchers suggested that the Gaus1 mother wavelet is suitable for finding lithological

discontinuity in well log signals (Singh *et al.* 2017; Chandrasekhar and Rao 2012). We have also identified major discontinuities with considerable accuracy by using the Gaus1 mother wavelet (figure 4). However, the maxima against inter-bedding inside the coal packets, which is essential in the Indian coal scenario, ended up in smearing/mixing with other maxima. Thus, we believe that the Gaus1 mother wavelet may not be suitable for well log analysis of Indian coal (figure 4).

On the other hand, Haar, being symmetric, mother wavelet appears to be highly suitable for picking blocky discontinuity and possesses good time localisation (Rongxi 2015). This feature of the Haar mother wavelet gives an advantage over other popular mother wavelets such as Gauss, Morlet, Symlet, and Daubechies, etc., in delineating major discontinuity and inter-bedding within thick coal seams without smearing/mixing with the next maxima (figure 4). Typically, in coal exploration, geophysical logs are recorded at 4–5 m/min speed (slowly logged data) considering the shoulder effects. However, there are occasions where the logistics and sub-surface geological conditions (Motur clay formation, faulted boreholes, etc.) largely limit the execution of the logging operation with the optimum speed. In such cases, the logging speed has to be increased to 10–15 m/min (fast logged data). Figure 5 illustrates the efficacy of continuous wavelet transform with Haar as the basis function in delineating the beds both in the case of slowly logged and fast logged data.

## 3. Application to well-log data of coal basins

### 3.1 Median filtering of well log datasets

Raw geophysical well log datasets are always noisy in nature. A purpose-based filter should be used to enhance the signal of interest without affecting trends and signatures about actual geology down the borehole. In oil and gas sectors, the low-frequency signal associated with thicker lithofacies is considered to be important, and less/no care is shown against the inter-bedding of fine thickness inside them. While in coal exploration, thin beds are also given exorbitant weightage along with thicker beds as the carbonaceous contents occur in outcrops as a minimum and up to the maximum depth of 0.50–1.16 km (MECL 1987, 2019). In order to record the response from finer bands present inside the coal seam, the log datasets in this

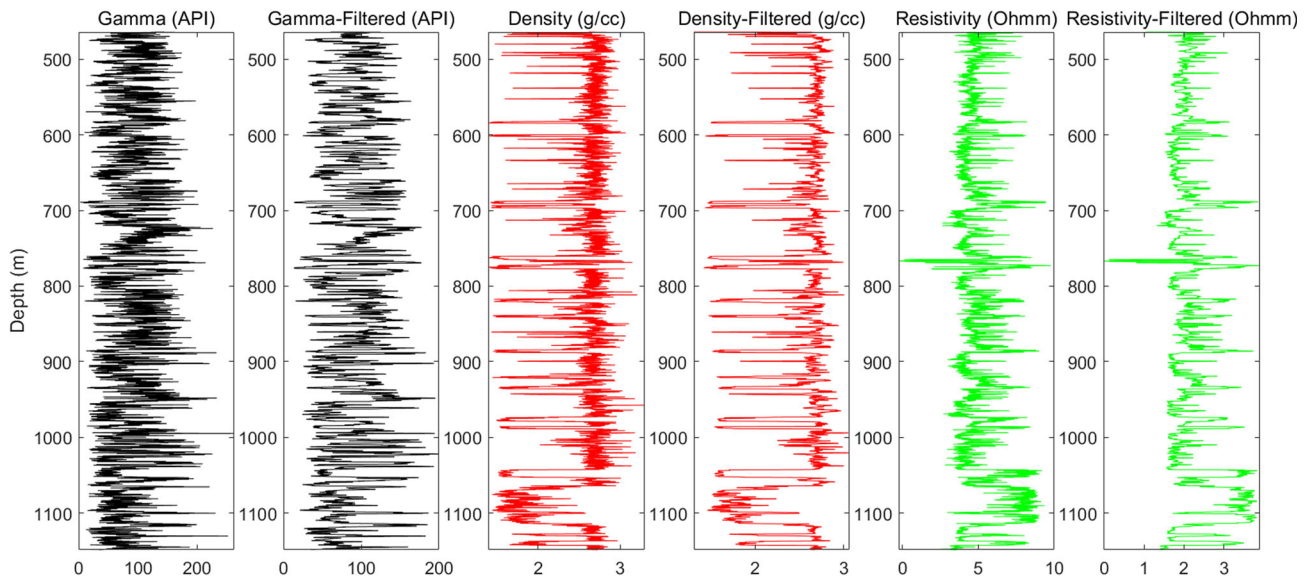


Figure 6. Showing raw geophysical data and their corresponding median filtered data.

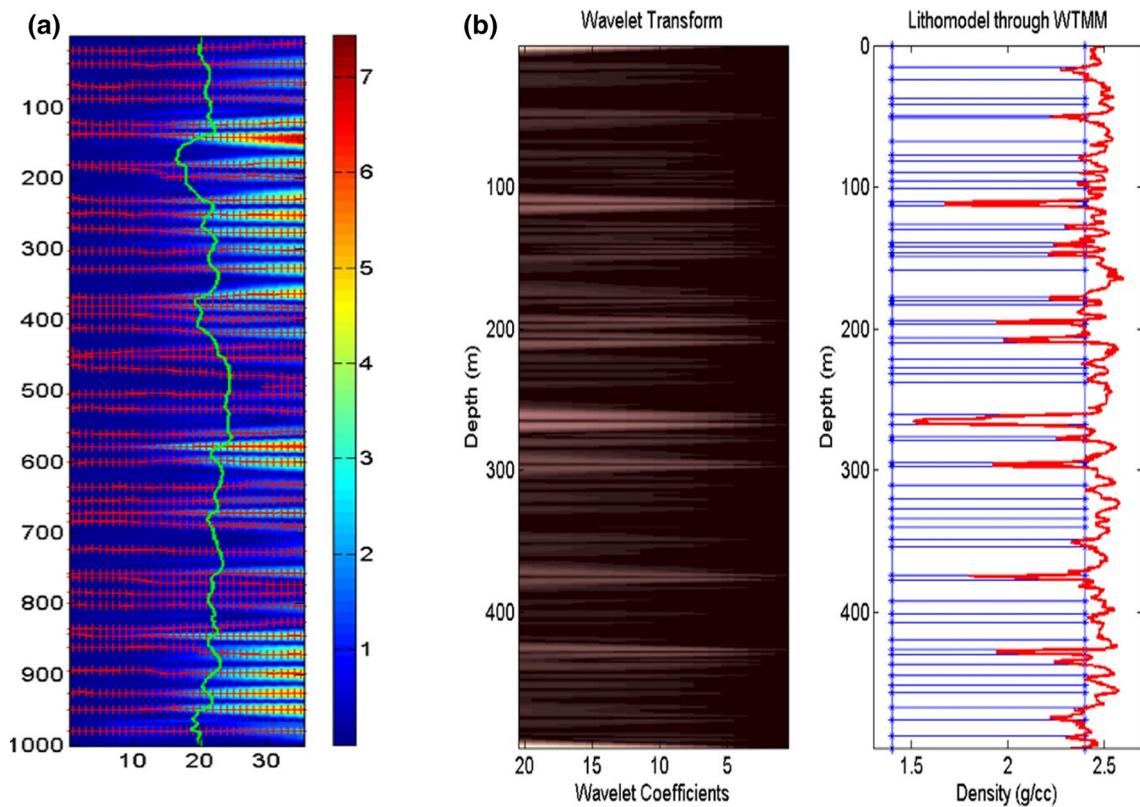


Figure 7. (a) Showing the efficiency of wavelet-based lithology identification on synthetic data and (b) lithology demarcation by tracing modulus maxima lines in wavelet scalogram on a density log.

study are acquired using slim-hole geophysical logging tools with sampling intervals from 0.01 to 0.05 m. Such sampling rate is sufficient enough to detect finer thickness bands present inside the coal seam. Once these mild signatures occurring inside thick coal seams are given proper attention, they

can provide a rough idea about the pureness of a coal seam.

In conventional interpretation, a moving average filter is used to visualise the datasets for interpretation that collapses/reshapes such mild signatures. To avoid this, median filtering developed by



Tukey (1977) is used in this study. If a signal with plenty of non-stationary spikes over a wide bandwidth, the median filter holds good, but it fails while the same non-stationary spikes convey certain information. To avoid such data loss, a proper geological understanding of the study area is mandatory before enforcing median filtering on the signal. Therefore, a window length of 3 can be tried initially on the coal-based well log signals. If there is no drastic change, then the window length can be extended up to 6 for further refining the datasets (Pratt 1978).

In general, gamma ray logs are a reliable indicator of lithological discontinuities, but the noisy clustered spikes occur at lithological variation in the signal, vague out the sedimentary succession shown by the signal down the borehole, especially the thin beds in the data. The density log, which is the paramount parameter in coal exploration, also suffers from a similar kind of turmoil. However, it differentiates the carbonaceous portion from non-coal. Whereas trends present in the non-coal region and inter-bedding inside coal seams are not distinct, both in the density and resistivity logs. With a good understanding of the area geology and using a median filter with appropriate window length, non-stationary spikes which are not related to sub-surface geology are

eradicated, and thereby, trends in the datasets become distinct and preserved (figure 6).

### 3.2 PCA analysis

In the density log, only carbonaceous deposits show a notable change in the signature. Whereas other rocks, such as sandstone, shale, shaly sand, and sandy shale, share some overlapping density values and do not show much difference on the density log. On the other hand, these lithologies can be differentiated both on gamma ray and resistivity logs due to their varied shale content and water content, respectively. Therefore, the thickness of any lithology noted in one parameter may differ from another parameter. Moreover, manually picking this variation is a difficult and time-consuming task. PCA helps in the automatic picking of these variations based on PC scores. Such PC scores can be identified precisely by cross-plotting PC scores with a good percentage of variance level with input log parameters.

### 3.3 Stratigraphy identification using CWT

Identification of very fine dirt bands within the thicker coal seams is of paramount interest in coal

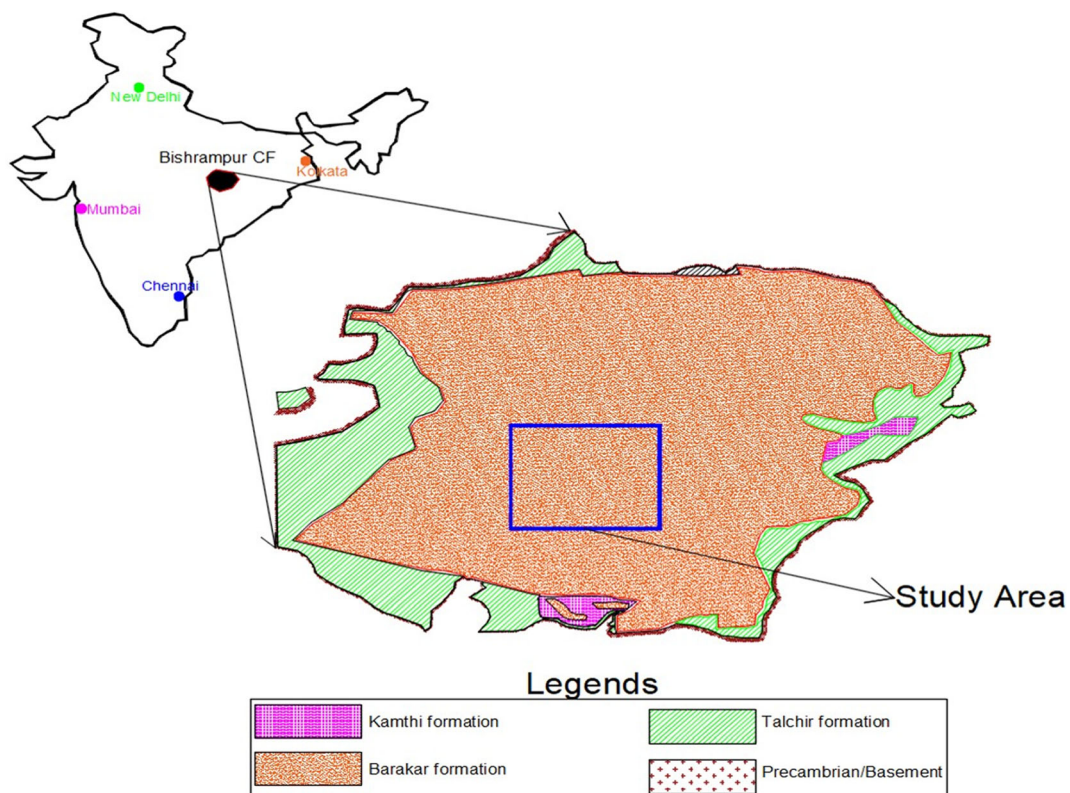


Figure 8. Showing the location of Bisrampur Coalfield in India and the location of the study area in Bisrampur coalfield.



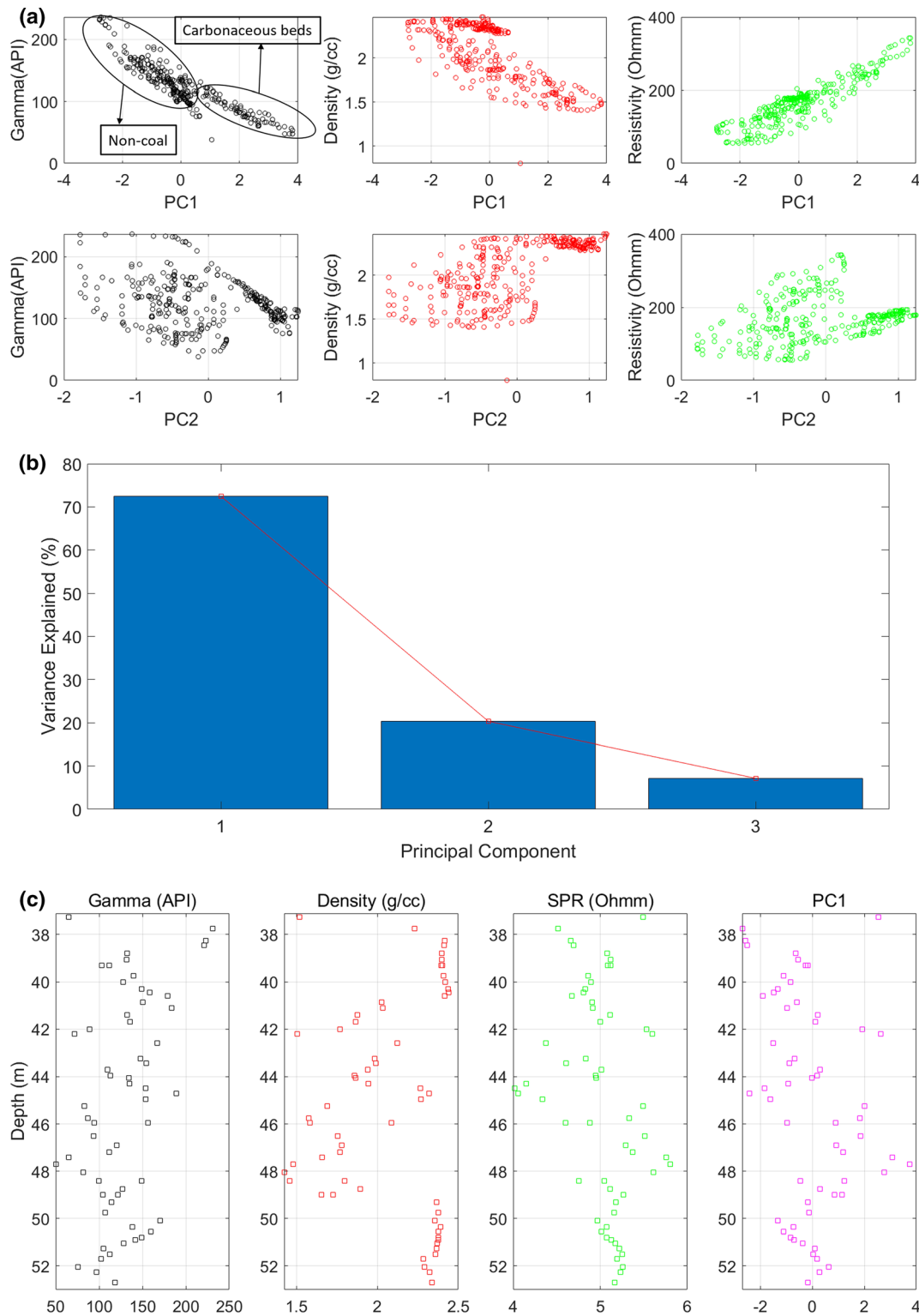


Figure 9. (a) Cross-plot between the first two PC scores and well log datasets (density, gamma, and resistivity), (b) Pareto-chart shows the variance percentage-hold by three principal components, and (c) shows the database categorising different carbonaceous beds.

exploration due to its banded nature. Geophysical well logs can identify such fine bands present inside coal seams due to their vertical resolution. However, picking such thin bands following the

conventional interpretation techniques is very difficult due to their varying thickness from one parameter to another. The application of CWT with appropriate time varying basis function on

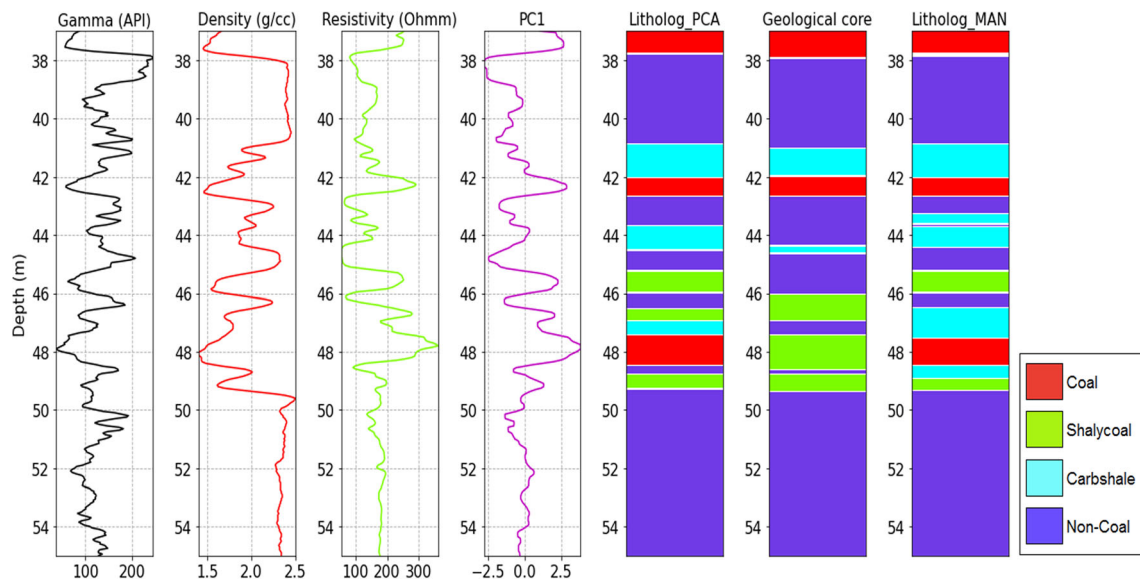


Figure 10. Showing geophysical parameters, PC1 score followed by PCA-based litholog, core, and manual interpretation.

Table 1. The various parameter ranges for carbonaceous beds for the borehole located in Bishrampur Coalfield.

Lithology	Density (g/cc)	Gamma (API)	Resistivity ( $\Omega$ m) (expressed in logarithmic scale)	PC1
Coal	1.280–1.560	0.00–50.00	$\geq 200$	$\geq 2.00$
Shalycoal	1.561–1.850	50.10–100.00	199–150	1.99–0.60
Carbshale	1.851–2.050	100.10–140.00	149–100	–0.15–0.59

well log can deduce different information present in the signal in terms of different wavelet coefficients. Therefore, CWT is applied in this study on selected PC score values for demarcating such fine beds from well log data.

The modulus maxima lines obtained using the Haar wavelet show a pointing nature when the signal of interest is approached with the proper scale with respect to the frequency of interest (figure 5a and b). In contrast, minute lithology variations inside thick coal seams or isolated thin bands are lost when approached with higher scale values while tracing modulus maxima values. Modulus maxima points between 3 and 4 scales are used for lithology picking from fine to coarser thickness, as shown in figure 7(a and b). Thus, wavelet transform is applied on selected PC score with appropriate scale, and the generated modulus maxima are projected on the signal. Maxima occurrence along the sample sequence is looked at directly from the depth and PC score along with original datasets to get roof-floor information. Further, PC score peak values for a respective lithology (roof-floor) are picked and compared with

the database decided in comparison with the original datasets. Values falling in the category of coal, shalycoal, carbshale, and non-coal are given codes as 1, 2, 3, and 4, respectively. These codes are colour coded for generating litholog for visualisation.

### 3.4 Database generation for interpretation

The main purpose of database generation is to derive a correlation between the generated PC score values and the input log parameter values. For developing a database, the density log being the important parameter for coal exploration is considered as the base along with the other log parameters. Here, the database is categorised as coal, shalycoal, and carbshale based on values exhibited by PC1 against all three carbonaceous lithologies. In general, the most widely used standard value ranges in the coalfields are considered for deciding density cut-off values. While in the case of the virgin area, these values are chosen either based on the available regional datasets or scout boreholes drilled in the study area.

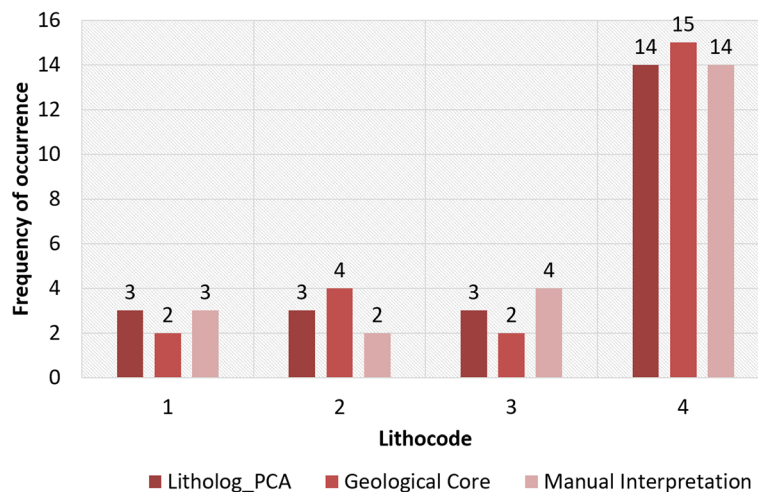


Figure 11. Showing the frequency of occurrence of demarcated lithologies.

In the present study, the validity of the proposed algorithm is also tested on two boreholes located in Bisrampur and Jharia coalfields of eastern India.

#### 4. Study on Bisrampur Coalfield

The Bisrampur Coalfield belongs to Lower Gondwana and covers an area of  $\sim 1036 \text{ km}^2$  in eastern India (figure 8). It comprises Talchir, Kaharbari, Barakar, and Kamthi formations and shows a gentle dip of  $2^\circ\text{--}3^\circ$ . Coal-bearing Barakar rocks are developed to a thickness of about 150 m. Several coal seams have been reported from different localities of Bisrampur coalfield. Detailed prospecting conducted by the Indian Bureau of Mines in the southwestern part of the field has indicated more than one coal horizon. Pasang seam is the thickest seam on the horizon, and the coal in this seam is of non-coking type.

PC1 is found to negatively correlate with density log and a positive correlation with resistivity log (figure 9a). Interestingly, PC1 of the gamma-ray log shows clear discrimination between carbonaceous and non-coal lithologies (figure 9a). It is observed that all the three geophysical logs of the borehole show noticeable sharp variations against the carbonaceous beds (low in case of density and gamma; high in case of resistivity) (figure 10). Whereas for the non-coal lithology, the gamma-ray log variations are only well pronounced compared to the other parameters (figure 10). PC1, which is the combination of all three parameters and maximum variance holder, indicates a positive PC score against carbonaceous beds and PC score  $\leq 0$

against non-coal lithologies. Therefore the cross plot of the gamma ray log and PC1 clearly segregates non-coal portions (higher API) from carbonaceous beds (lower API) (figure 9a). In contrast, cross-plotting these variables with PC2 does not show any understandable relations (figure 9a). Therefore, PC1 having a prominent variance level possesses merits of all the three variables, and suitable for interpretation (figure 9b). Carbonaceous bed values of original log inputs and their corresponding PC1 values are shown in figure 9(c). It indicates that the PC1 values vary from  $-0.15$  to  $0.59$  for carbshale,  $1.99\text{--}0.60$  for shalycoal, and  $2.00$  for coal (table 1).

Figure 10 shows the comparison of litholog derived from the core, manual interpretation, and PCA-CWT modelling. In the depth range from 38 to 42 m (figure 10), the gamma-ray log has drastically changed from its maximum of 235.955 API at 38 m depth to 100–200 API values up to a depth of 42 m. Gamma-ray log values  $>200$  API indicate non-coal (pure shale in the present case), and  $<200$  API may be the presence of carbonaceous beds (coal, shaly coal, and carbonaceous shale)/non-coal beds in the instant case, which could be verified using density and resistivity logs. Similarly, in the above depth range from 38 to 42 m and 50–55 m, the variations observed in the gamma-ray log are duly checked with density and resistivity parameters for the presence of carbonaceous and non-coal beds. PCA-CWT-based interpretation matches well with core data and manual interpretations up to a depth range from 37.5 to 43.0 m (figure 10).

The frequency of occurrence of carbonaceous beds in litholog derived from core data, manual

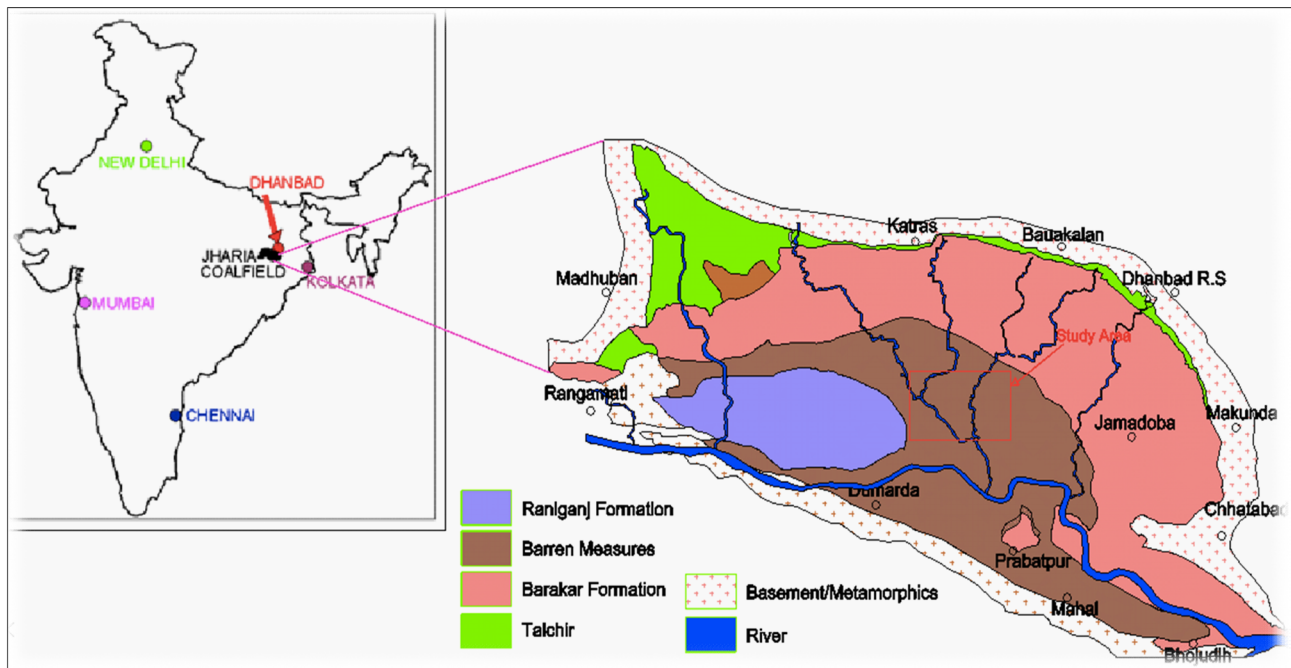


Figure 12. Showing the location of Jharia Coalfield in India and the location of the study area in Jharia Coalfield.

interpretation, and the PCA-CWT proposed technique is given in figure 11. The occurrence of coal, shalycoal, carbshale, and non-coal beds match up to 43.5 m in both interpretations. Beyond 43.5 m depth, the interpretations derived from all three methods differ either in terms of lithology or thickness (figure 11). This could be due to less parting between the carbonaceous beds, lack of resolution in geological information in the core, and cut-off fixing in the carbonaceous packets for manual interpretation (figure 11). Around 44 m, it is evident from the geophysical log that a carbonaceous shale bed of thickness 0.8 m is present, and similar results were also found in the proposed PCA-CWT algorithm. However, the carbonaceous bed was not identified in core lithology at ~44 m due to poor core recovery. In the depth range from 46 to 49.5 m, the core data indicate only a series of shalycoal beds. In contrast to this, the proposed PCA-CWT algorithm and manual interpretation register different carbonaceous beds such as coal, shalycoal, and carbshale. In the taken depth range, due to less parting between the carbonaceous beds and lack of resolution in geological information in core data, the discrete coal beds are marked as one in core litholog. These discrete coal beds with less parting in between them are explained well in PC1 (maximum variance holder) with merits of all geophysical parameters and good resolution of thin bands inside the coal packets. Further, around

44 m depth manually interpreted litholog shows a slightly higher thickness of carbonaceous beds. This could be due to the difference in density cut-off values chosen by the interpreter for identifying coal to shalycoal, shalycoal to carbshale sequences during manual interpretation.

## 5. Study on Jharia Coalfield

Jharia Coalfield is situated about 260 km north-west of Calcutta, mainly in the heart of Damodar valley. The coalfield is roughly sickle in shape and covers an area of  $\sim 456 \text{ km}^2$  with an extension of  $\sim 18 \text{ km}$  in the north–south direction and a maximum of  $\sim 38 \text{ km}$  in an east–west direction (figure 12). The general stratigraphic succession of the area is that the basement metamorphic rocks are overlain by the Talchir, Barakar, and Raniganj formations. Barakar Formation contains a major coal-bearing horizon of Jharia Coalfield, and coals of this formation can be divided into: (i) low volatile coals containing up to 26% volatiles, (ii) medium volatile coals containing 26–28% volatiles, and (iii) high volatile coals containing over 28% volatiles. The coal sequences of the overlying Raniganj Formation have slightly higher moisture content than the Barakar Formation.

PC1 against density and resistivity parameters show a good correlation for coal-bearing sequences



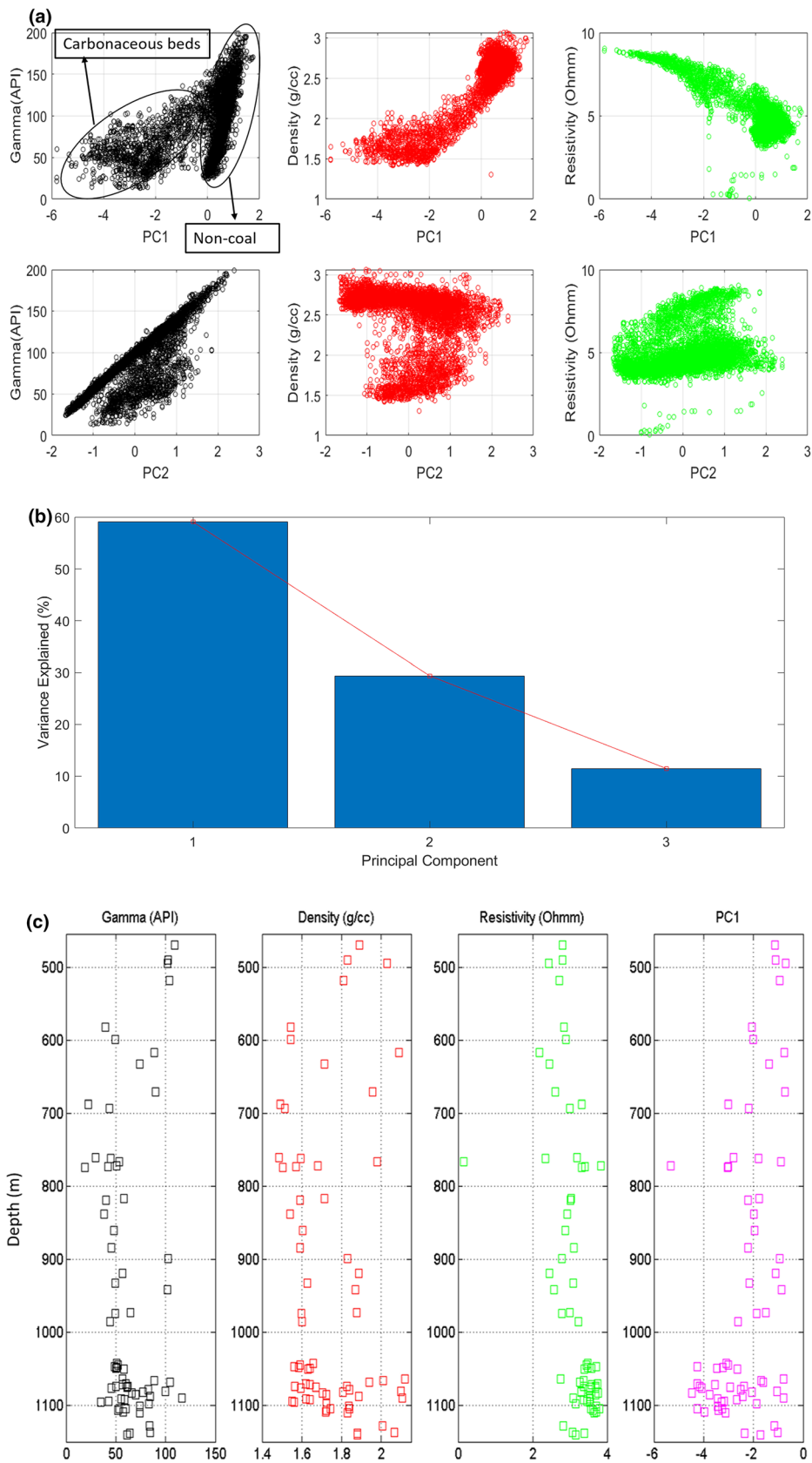


Figure 13. (a) Cross-plot between the first two PC scores and well log datasets (density, gamma, and resistivity), (b) Pareto-chart shows the variance percentage-held by three principal components, and (c) shows the database categorising different carbonaceous beds.

Table 2. The various parameter ranges for carbonaceous beds for the borehole located in Jharia Coalfield.

Lithology	Density (g/cc)	Gamma (API)	Resistivity ( $\Omega$ m) (expressed in logarithmic scale)	PC1
Coal	1.280–1.560	0.00–40.00	>3.75	$\leq -4.00$
Shalycoal	1.561–1.85	40.10–100.00	3.74–3.50	–3.99 to –2.00
Carbshale	1.851–1.99	100.10–140.00	3.49–2.20	–1.99 to –0.50

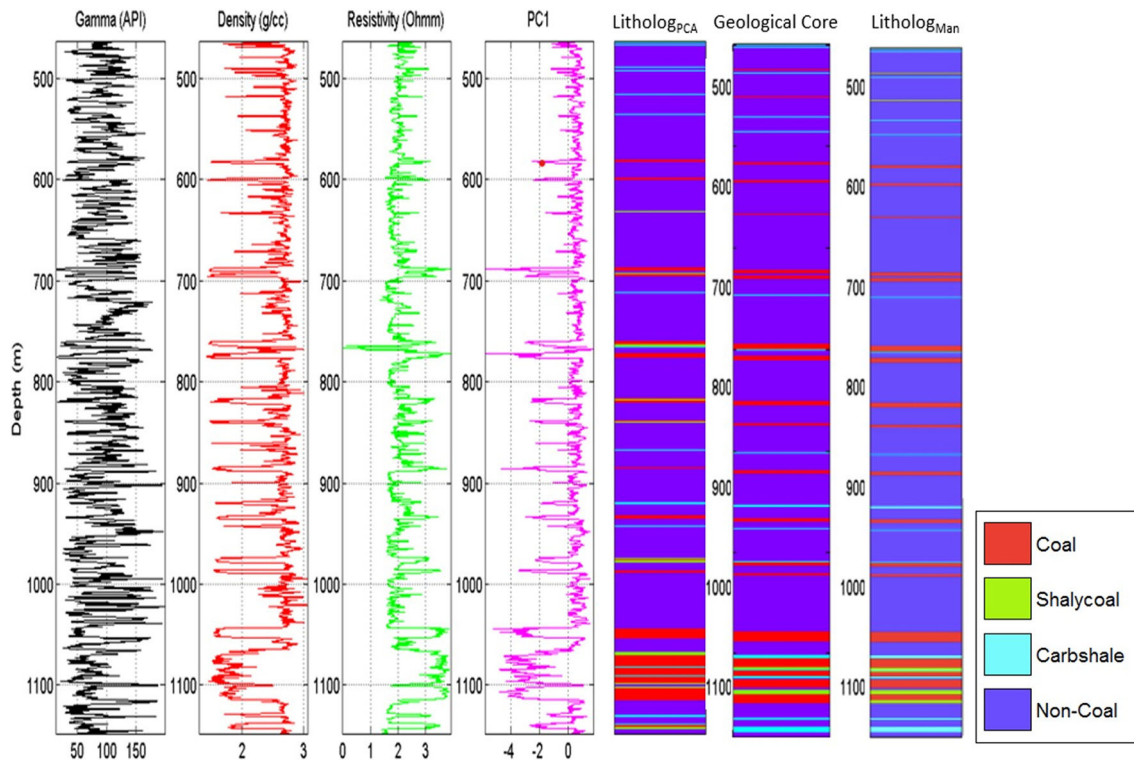


Figure 14. Showing geophysical parameters, PC1 score followed by PCA-based litholog, core, and manual interpretation.

(figure 13a). The carbonaceous and non-coal beds are resolved in the cross plot of PC1 and gamma ray due to the strong response of all three input parameters against the carbonaceous beds and silence in signatures of density and resistivity parameters except gamma ray log. All the coal seams in the borehole from 500 m belong to the Barakar Formation, which is the main reason that density and resistivity values share some overlapping ranges, whereas the gamma-ray log fluctuates strictly as per the shale content in the formation. It is also noticed that cross-plotting of PC2 with original datasets also shows a strong positive correlation with gamma-ray log, and its correlation with density and resistivity logs does not show much change (figure 13a). Therefore, it is suggested that PC1 can alone be used to interpret different carbonaceous beds instead of using all three PC scores (figure 13b). Carbonaceous bed

values of original log inputs and their corresponding PC1 values are shown in figure 13(c). It indicates that PC1 values vary from  $-3.99$  to  $-2.00$  for shalycoal,  $-1.99$  to  $-0.50$  for carbshale, and  $\leq -4.00$  for coal (table 2).

At depth ranges from 460 to 500 m, one coal and a few thin carbshale beds are noted in the core data (figure 14). Whereas PCA indicates all the thin beds as carbshale (figure 14). In the core log, depth ranges such as 758–766, 816–822, 837–841, 973–979, and 1064–1116 m are marked as coal in most places, but in PCA-based interpretation, few shalycoal/carbshale sequences along with coal were identified at these depths (figure 14). These discrepancies between the core log and PCA-CWT interpretation indicate the influence of characteristic gamma, density, and resistivity logs on PC score. Thus PCA-CWT-based lithology modelling more accurately deduced the roof-floor of minute

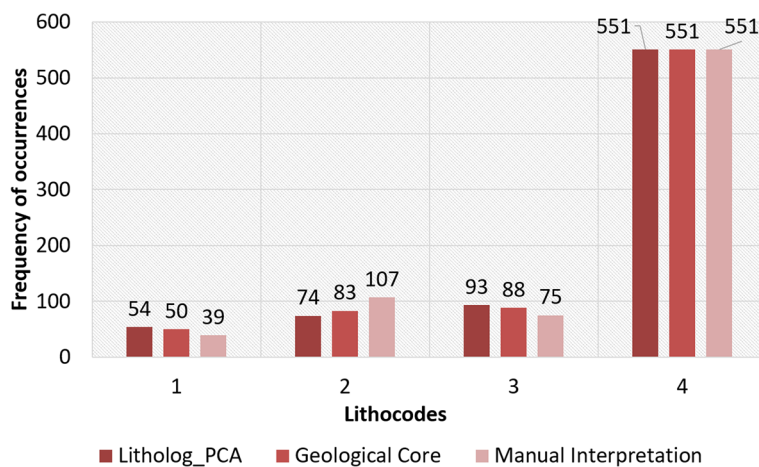


Figure 15. Showing the frequency of occurrence of demarcated lithologies.

variations in the well logs rather than core log. It is also observed that core lithology matches quite well with the manual interpretation of geophysical logs except in 758–766, 816–822, 837–841, 973–979, and 1064–1116 m depth ranges. These variations are attributed to using an averaging filter that collapses thin bed variations, boundary conditions between different carbonaceous beds, and interpreter value ranges. Further, the frequency of occurrence of carbonaceous beds derived from core data, manual interpretation, and the proposed PCA-CWT technique is shown in figure 15. All three interpretations show varying counts between coal and shalycoal compared with the rest as indicated in depth ranges 758–766, 816–822, 837–841, 973–979, and 1064–1116 m.

As discussed in the earlier sections, the applicability of the PCA-CWT technique was tested only on a few wells from Jharia and Bistrampur coal fields, which come under coking and non-coking categories, respectively. Although the study showed convincing results in detecting thin carbonaceous beds of thickness  $\geq 0.4$  m, the applicability of the technique needs to be tested on more wells from other coalfields holding banded power grade coal seams, e.g., Talcher and IB valley coalfields, Odisha.

## 6. Conclusions

In this study, a combined principal component analysis (PCA) and continuous wavelet transform (CWT) algorithm is developed for lithological modelling of well log signals in coal exploration. This proposed algorithm was successfully implemented on gamma ray, density, and resistivity logs of two boreholes located in Bistrampur and Jharia

coalfields of eastern India. The major findings of this study are summarised below:

- CWT based on Haar mother wavelet is found to be more useful in delineating major to minor stratigraphic changes irrespective of logging speed.
- PCA reveals that PC1 accentuates major lithological discontinuities within the coal seams compared to PC2. The cross-plot of PC1 with original log (gamma ray, density, and resistivity) datasets differentiates the non-coal portions from carbonaceous beds (shalycoal, carbshale, and coal).
- The study demonstrates that the PCA-CWT is suitable for detecting thin carbonaceous beds (shalycoal, carbshale, and coal) of thickness  $\geq 0.4$  m.
- The predicted lithology from PCA-CWT modelling match well with core data and manual interpretations of the boreholes. At few depth ranges, the proposed algorithm is able to reveal additional lithological discontinuities that were not identified in the core data.
- In coal blocks having limited coring boreholes or structurally disturbed areas especially faulted boreholes, PCA-CWT modelling of geophysical logs can be an alternative for deriving litholog.

## Acknowledgements

The authors sincerely express their gratitude to Sri. Manoj Kumar, Chair-Cum-Managing Director, CMPDIL, and Sri. S Gomasta, Director (T/CRD), CMPDIL, for providing permission to use the field datasets and publishing the results. Help rendered by Sri. Arjun Hembrom, General Manager (Exploration), CMPDIL and Sri. Rajiv Kumar Singh, Chief

Manager (Geology), CMPDIL, are also acknowledged. This work was supported by a financial grant (Ministry of Science and Technology, DST-FIST/197/2018-19/580) from the Department of Science and Technology, Government of India.

### Author statement

Thinesh Kumar: Visualisation, methodology, writing – initial draft. Seelam Naresh Kumar: Critical review of the manuscript. G Srinivasa Rao: Conceptualisation, overall supervision, and finalisation of the manuscript.

### References

- Alvarez G, Sanso B, Michelena R J and Jimenez J R 2003 Lithologic characterisation of a reservoir using continuous wavelet transform; *IEEE Trans. Geosci. Remote Sens.* **41(1)** 59–65.
- Anyiam O A, Mode A W and Okara E S 2018 The use of cross-plots in lithology delineation and petrophysical evaluation of some wells in the western Coastal Swamp, Niger Delta; *J. Pet. Explor. Prod. Technol.* **8(1)** 61–71.
- Chandrasekhar E and Rao V E 2012 Wavelet analysis of geophysical well-log data of Bombay offshore basin, India; *Math. Geosci.* **44(8)** 901–928.
- Chatterjee R and Paul S 2012 Application of cross-plotting techniques for delineation of coal and non-coal litho-units from well logs in Jharia Coalfield, India; *Geomaterials* **2** 94–104.
- Chen S, Liu P, Tang D, Tao S and Zhang T 2021 Identification of thin-layer coal texture using geophysical logging data: Investigation by wavelet transform and linear discrimination analysis; *Int. J. Coal Geol.* **239** 103727.
- Coconi-Morales E, Ronquillo-Jarillo G and Campos-Enrriquez J O 2010 Multi-scale analysis of well-logging data in petrophysical and stratigraphic correlation; *Geofis. Int.* **49(2)** 55–67.
- Crain E R 1986 *The log analysis handbook: Quantitative log analysis methods*; volume 1, Penn Well, Tulsa, pp. 91–95.
- Daubechies I 1992 *Ten lectures on wavelets*; Society for Industrial and Applied Mathematics, Philadelphia.
- Dewan J T 1983 *Essentials of modern open-hole log interpretation*; PennWell Books.
- Elek I 1988 Some applications of principal component analysis: Well to well correlation, zonation; *Geobyte* **3** 46–53.
- Farge M 1992 Wavelet transforms and their applications to turbulence; *Ann. Rev. Fluid Mech.* **24(1)** 395–458.
- Goupillaud P, Grossmann A and Morlet J 1984 Cycle-octave and related transforms in seismic signal analysis; *Geoexploration* **23(1)** 85–102.
- Javid M and Tokhmechi B 2012 Formation interface detection using gamma ray log: A novel approach; *J. Min. Environ.* **3(1)** 41–50.
- Kassenaar J D C 1991 An application of principal components analysis to borehole geophysical data; In: *Proceedings of the 4th International MGLS/KEGS Symposium on Borehole Geophysics for Minerals, Geotechnical and Groundwater Applications*, pp. 211–218.
- Kumar P and Foufoula-Georgiou E 1997 Wavelet analysis for geophysical applications; *Rev. Geophys.* **35(4)** 385–412.
- Lim J S, Kang J M and Kim J 1998 Artificial intelligence approach for well-to-well log correlation; In: *SPE India Oil and Gas Conference and Exhibition*, Society of Petroleum Engineers, OnePetro.
- Lim J S 2003 Multivariate statistical techniques including PCA and rule-based systems for well log correlation; *Dev. Petrol. Sci.* **51** 673–688.
- Li W, Chen T, Song X, Gong T and Liu M 2020 Reconstruction of critical coalbed methane logs with principal component regression model: A case study; *Energy Explor. Exploit.* **38(4)** 1178–1193.
- MECL 1987 Geological report on exploration for coal: Kapuria block, Jharia Coalfield, Jharkhand, Unpublished report, vol. I, pp. 127.
- MECL 2019 Geological report on detailed exploration for coal: Kapuria block, Jharia Coalfield, Jharkhand, Unpublished report, vol. I, pp. 298.
- Ouadfeul S A and Aliouane L 2011 Multifractal analysis revisited by the continuous wavelet transform applied in lithofacies segmentation from well-logs data; *Int. J. Appl. Phys. Math.* **1(1)** 10.
- Panda M N, Mosher C and Chopra A K 2000 Application of wavelet transforms to reservoir data analysis and scaling; In: *Proc. SPE 71st Annual Technical Conference and Exhibition*, Denver, CO, USA, pp. 251–266.
- Pan S Y, Hsieh B Z, Lu M T and Lin Z S 2008 Identification of stratigraphic formation interfaces using wavelet and Fourier transforms; *Comput. Geosci.* **34(1)** 77–92.
- Polikar R 1999 The story of wavelets; In: *Physics and modern topics in mechanical and electrical engineering*, pp. 192–197.
- Pratt W K 1978 *Digital image processing*; John Wiley & Sons.
- Ren P, Xu H, Tang D, Li Y, Sun C, Tao S, Li S, Xin F and Cao L 2018 The identification of coal texture in different rank coal reservoirs by using geophysical logging data in north-west Guizhou, China: Investigation by principal component analysis; *Fuel* **230** 258–265.
- Rongxi Z 2015 Wavelet analysis of well-logging data in petrophysical and stratigraphic correlation; In: *2015 IEEE International Conference on Computational Intelligence & Communication Technology IEEE*, pp. 18–21.
- Singh A, Maiti S and Tiwari R K 2017 Selection of optimum wavelet in CWT analysis of geophysical downhole data; *J. Indian Geophys. Union* **21(2)** 153–166.
- Tukey J W 1977 *Exploratory data analysis*; vol. **2**, pp. 131–160.
- Yue W, Tao G and Liu Z 2006 Identifying reservoir fluids by wavelet transform of well logs; *SPE Reservoir Evaluation & Engineering* **9(05)** 574–581.
- Zhang X F, Pan B Z, Wang F and Han X 2011 A study of wavelet transforms applied for fracture identification and fracture density evaluation; *Appl. Geophys.* **8(2)** 164–169.