Front. Earth Sci. 2022, 16(3): 557–567
https://doi.org/10.1007/s11707-021-0926-5

RESEARCH ARTICLE

Probabilistic Fisher discriminant analysis based on Gaussian mixture model for estimating shale oil sweet spots

Kun LUO, Zhaoyun ZONG (✉)

School of Geosciences, China University of Petroleum, Qingdao 266555, China

Abstract The delineation of shale oil sweet spots is a crucial step in the exploration of shale oil reservoirs. A single attribute such as total organic carbon (TOC) is conventionally used to evaluate the sweet spots of shale oil. This study proposes a probabilistic Fisher discriminant approach for estimating shale oil sweet spots, in which the probabilistic method and Gaussian mixture model are incorporated. Statistical features of shale oil facies are obtained based on the well log interpretation of the samples. Several key parameters of shale oil are projected to data sets with low dimensions in each shale oil facies. Furthermore, the posterior distribution of different shale oil facies is built based on the classification of each shale oil facies. Various key physical parameters of shale oil facies are inversed by the Bayesian method, and important elastic properties are extracted from the elastic impedance inversion (EVA-DSVD method). The method proposed in this paper has been successfully used to delineate the sweet spots of shale oil reservoirs with multiple attributes from the real pre-stack seismic data sets and is validated by the well log data.

Keywords probabilistic Fisher discriminant analysis, sweet spots, shale-oil facies, Bayesian inversion

1 Introduction

Shale oil is typically composed of both free oil and adsorbed oil, and is mainly retained or stored in the organic-rich source rocks (Raji et al., 2015; Wang et al., 2015). The probabilistic Fisher discriminant approach aims to delineate the shale oil sweet spots from the measured seismic data sets and well-log data. Total organic carbon (TOC), maturity, high Young's modulus, and low Poisson's ratio are the most important indicators of sweet spots in the shale reservoir (Zong et al., 2012b; Ouadfeul and Aliouane, 2016; Verma et al., 2016; Chen et al., 2018). The selection of proper reservoir inversion parameters is crucial in the accurate characterization of the shale reservoir. The four most important reservoir physical properties used to characterize unconventional reservoirs are the total organic carbon (TOC), porosity, pyrolysis S1, and clay content (Jarvie, 2012; Gorynski et al., 2019; Chen et al., 2020; Luo and Zong, 2020). This information together with fracturing-related features can be used to find and characterize the sweet spots of shale oil reservoirs.

The content of oil and gas is a key indicator of shale oil sweet spots. Particularly, the S1 value in pyrolysis of organic matter can be used to estimate the shale oil content in the source rock approximately and delineate the sweet spots of the shale oil reservoir (Song et al., 2013). In the pyrolysis of organic matter, gas is continuously released from the samples. Espitalie et al. (1977a) defined the first peak (S1) in the pyrolysis as the hydrocarbon content in the rock samples. Moreover, the S1 to TOC ratio, referred to as oil saturation index (OSI), can be used to determine the potential producibility of oil (Jarvie, 2012). The mobility of shale oil is affected by the properties of the retained oil and the light hydrocarbon component (Bao et al., 2016).

The pyrolysis S1 is related to the mobility of shale oil. Gorynski et al. (2019) used pyrolysis S1 data to assess the mobility of tight oil. The free or adsorbed hydrocarbon in the source rock can be characterized quantitatively by pyrolysis S1 from 3D pre-stack seismic data sets (Luo and Zong, 2019). Chen et al. (2020) used a large amount of Rock-Eval pyrolysis data in the Junggar Basin of China to evaluate tight oil resources. Pyrolysis S1 and fracture pressure gradient (FG) are key properties of any unconventional shale reservoir and the abundance of total organic carbon is the critical parameter to delineate geological sweet spots (Robert et al., 2015).

The pre-stack seismic inversion has been widely used to

Received February 13, 2021; accepted June 17, 2021

E-mail: zongzhaoyun@upc.edu.cn

estimate TOC, clay content, and porosity. Ogiesoba and Hammes (2014) identified TOC-rich zones from 3D post-stack seismic attributes in the Eagle Ford Shale quantitatively. Yin et al. (2014, 2015) proposed a pragmatic approach, which combines the rock physics model and Bayesian inversion to directly inverse physical parameters of reservoirs from the elastic impedance of three different angles. Wang (2017) introduced a method for the pre-stack seismic inversion of shale gas sweet spots based on Yin's works, and quantitatively inversed TOC, clay content and porosity. Ouadfeul and Aliouane (2016) inverted TOC using the multilayer perceptron neural network from three-dimensional seismic data in the Barnett Shale Formation. Yu et al. (2020) employed a rock-physics-based Bayesian method to inverse the total organic carbon in the Dongying Basin, China.

The Young's modulus and Poisson's ratio are directly related to the brittleness of rocks (Rickman et al., 2008). Yin et al. (2004) inversed Poisson's ratio, P-wave impedance, and S-wave impedance from pre-stack seismic data sets based on elastic impedance. The shale oil sweet spots have relatively high Young's modulus, but low Poisson's ratio (Sena et al., 2011). Zong et al. (2012a) derived the YPD-Zoeppritz function, which uses the correlation among Young's modulus, Poisson's ratio, P-wave and S-wave velocity, and density to estimate Young's modulus and Poisson's ratio. The elastic impedance (EI) has been widely used in the inversion of reservoir attributes (Connolly, 1999). In the YPD-EI function (Zong et al., 2013), Young's modulus, Poisson's ratio, and density can be directly inverted from elastic impedance variation with incident angle. Huang (1984) proposed a model for the prediction of fracture pressure and considered several attributes such as overburden stress, rock strength, and Poisson's ratio. Ma et al. (2020) utilized an AVAZ inversion method based on azimuth seismic data and derived the approximate correlation of the reflection coefficient, which is characterized by Young's modulus, Poisson's ratio, density, and fracture weaknesses parameters. Ma et al. (2020) calculated fracture pressure by an improved shale anisotropy fracture pressure model, which is based on Huang's model (1984). Hu et al. (2015) assumed isotropic pressure coupling and estimated fracture pressure gradient (FG) quantitatively using the Poisson's ratio, effective stress, and pore-pressure gradient.

Fisher discriminant analysis (Fisher, 1936), also known as Fisher's LDA, was initially used to distinguish the two-species problem. It has been used in many fields such as text classification (AbuZeina and Al-Anzi, 2018), face recognition (Wang et al., 2016; Wu et al., 2017), and other data classification applications (Jiang and Chen, 2020; Guo et al., 2021). The FDA needs to find an optimal projection by maximizing the between-class distance and minimizing the within-class distance simultaneously. It is a supervised technique that requires class labels of the training data. Several effective Fisher discriminant functions have been proposed. However, the most commonly used discriminant functions are the ratio trace criterion (Rao, 1948), trace ratio criterion (Wang et al., 2007), and vectors ratio criterion (Foley and Sammon, 1975). Multiple inversed features of the shale oil reservoir can be used with Fisher's discriminant analysis for the classifications and estimation of the shale oil sweet spots.

This study proposes a PFDA technique in which the probabilistic method is incorporated into the Fisher discriminant analysis to delineate the sweet spots of shale oil. This method could use multiple attributes from shale oil reservoirs to distinguish the sweet spots. The discriminate criterion was converted to a Gaussian mixture probability model, which was in turn expanded in the Fisher discriminant analysis method. Additionally, Young's modulus, Poisson's ratio, and density from real pre-stack seismic data sets were inverted based on the method proposed by Zong et al. (2012a, 2013). Inversion was performed for TOC, S1, clay content, and porosity using rock physics principles in the Bayesian method. With these inversion attributes of shale oil reservoirs, the sweet spots could be delineated by our PFDA method and the introduced method was applied to real seismic data and validated with the well logs.

2 Methods

2.1 Fisher discriminant analysis with probabilistic approach for sweet spots of shale oil

The main idea in the Fisher discrimination analysis is to project all types of data in a certain direction to obtain the highest separation distance among all projections. Fisher discrimination analysis attempts to find an optimal low-dimensional space to store the most discriminative information. FDA and its developments have been successfully applied in face recognition, speech recognition, and many other fields. However, Fisher discriminant analysis (FDA) focuses on a linear problem and due to the nonlinear separability of the shale oil sweet spots of different types, the FDA alone may not be able to achieve the delineation of the sweet spot of shale oil. Bayesian methods are often adopted to solve these linear restrictions. In this study, the conditional probability methods and the Fisher discriminant analysis (PFDA) are integrated and combined.

The information obtained from the lithofacies, petrophysical curves, and sedimentologic information was employed to categorize shale oil facies. Some reservoir attributes such as total content of organic matter, the thickness of source rock, maturity of kerogen, mobility of shale oil, and pyrolysis S1 may be used to distinguish the sweet spots in the delineation of shale oil. Given that there are m classes and n samples in different facies G_k, $1 \leqslant k \leqslant m$:

$$\begin{cases} G_1 : \boldsymbol{x}_1^{(1)}, \boldsymbol{x}_2^{(1)}, ..., \boldsymbol{x}_n^{(1)} \\ G_2 : \boldsymbol{x}_1^{(2)}, \boldsymbol{x}_2^{(2)}, ..., \boldsymbol{x}_n^{(2)} \\ \cdots\cdots\cdots\cdots \\ G_m : \boldsymbol{x}_1^{(m)}, \boldsymbol{x}_2^{(m)}, ..., \boldsymbol{x}_n^{(m)} \end{cases}, \tag{1}$$

where $\boldsymbol{x}_j^{(k)}, 1 \leqslant j \leqslant n$, denote output features of a sample from the supervised facies. $N = 3n$, denotes the total samples of all classes.

Shale oil resources are organic-rich mudstones that have generated oil that is stored in the organic-rich mudstone intervals or continuous organic-lean intervals. The formation of shale oil resources can be categorized into three different types by their dominant organic matter and lithological composition. The three types are (G_1) organic-rich mudstones with high maturity, (G_2) organic-lean intervals of sand, and (G_3) organic-rich mudstones with low maturity. Several reservoir parameters influence the quality of the shale oil resources. Shale oil resource systems vary considerably from one system to another (Jarvie, 2012), but the best resources share some commonalities such as marine shales with excellent TOC values, high thermal maturity, high clay content, and relatively high mobility of shale oil. Generally, the high mobility of shale oil is associated with high free hydrocarbon content in the shale, which corresponds with a high pyrolysis S1 in the shale. In particularly, each type of shale oil facies in our study area has six main characteristics which enable the delineation and indication of the sweet spots of shale oil resources.

Based on the classification of G_1, G_2, G_3, let $\boldsymbol{x}_j^{(k)} = [J_1, J_2, ..., J_6]$ denote the characteristic matrix of a sample. $J_1, J_2, ..., J_6$ represent TOC, pyrolysis S1, clay content, Young's modulus, porosity, and Poisson's ratio, respectively.

The target of fisher discriminant analysis is to find the projection matrix $\boldsymbol{W}$, by maximizing the between-class scatter distance, and meanwhile minimizing the within-class scatter distance (Wu et al., 2017), which can be formulated as

$$\arg\max_{\boldsymbol{W}} \frac{\boldsymbol{W}^{\mathrm{T}} \mathbf{S}_b \boldsymbol{W}}{\boldsymbol{W}^{\mathrm{T}} \mathbf{S}_\omega \boldsymbol{W}}, \tag{2}$$

where $\boldsymbol{S}_w$ and $\boldsymbol{S}_b$ denote as the within scatter matrix and the between scatter matrix, respectively, which can be defined as

$$\boldsymbol{S}_w = \sum_{k=1}^{m} \sum_{j=1}^{n} (\boldsymbol{x}_j^{(k)} - \boldsymbol{\mu}_k)(\boldsymbol{x}_j^{(k)} - \boldsymbol{\mu}_k)^{\mathrm{T}}, \tag{3}$$

$$\boldsymbol{S}_b = \sum_{k=1}^{m} (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^{\mathrm{T}}, \tag{4}$$

where $\boldsymbol{\mu}_k$ is the mean vectors of atoms in the kth class G_k, and $\boldsymbol{\mu} = \frac{1}{N} \sum_{k=1}^{m} \sum_{j=1}^{n} \boldsymbol{x}_j^{(k)}$ denotes the mean vector of all samples from all classes.

Specifically, the optimal solution (Wang et al., 2016) of Eq. (2) can be written as follows:

$$\boldsymbol{W}^* = \operatorname{argmin} - \boldsymbol{W}^{\mathrm{T}} \mathbf{S}_b \boldsymbol{W}$$

$$s.t. \boldsymbol{W}^{\mathrm{T}} \mathbf{S}_\omega \boldsymbol{W} = 1. \tag{5}$$

We can use the Lagrange multiplier method to get Eq. (6):

$$\boldsymbol{S}_b \boldsymbol{W} = \lambda \boldsymbol{S}_\omega \boldsymbol{W}. \tag{6}$$

If the $\boldsymbol{S}_\omega$ is nonsingular, we can obtain the matrix form:

$$\boldsymbol{S}_\omega^{-1} \boldsymbol{S}_b \boldsymbol{W} = \boldsymbol{W} \Lambda, \tag{7}$$

where $\boldsymbol{W} = [w_1, ..., w_q]$ the eigenvectors; $\Lambda = \operatorname{diag}[\lambda_1, ..., \lambda_q]$, $\lambda_1 \geqslant \lambda_2, ..., \geqslant \lambda_q \geqslant \lambda_{q+1} = 0$, is a diagonal matrix of eigenvalues; q is the rank of the matrix of $\boldsymbol{S}_\omega^{-1} \boldsymbol{S}_b$.

when we get the optimal projection direction, the raw data can then be projected onto a new hyperplane:

$$y_j^{(k)} = \boldsymbol{W}^{\mathrm{T}} x_j^{(k)}, \tag{8}$$

where $y_j^{(k)}$ are the samples after the projection of the raw data; $\boldsymbol{W}$ is the optimal projection direction.

Assume that variables y in each type conform to the Gaussian mixture model with finite components N_d. Especially the variables y in each type G_j can be generally sampled by Monte Carlo simulation and their distributions could be expressed as

$$p(y|G_j) = \sum_{d=1}^{N_d} \pi_d N(y; \boldsymbol{\mu}_j^d, \boldsymbol{\Sigma}_j^d), \tag{9}$$

where N is the Gaussian distribution, $\boldsymbol{\mu}_j^d$ represents the vector of means in each type G_j; $\boldsymbol{\Sigma}_j^d$ and π_d are the covariance matrix and the weights of each Gaussian distribution in the type G_j, respectively. Then the prior distribution (Grana and Rossa, 2010) can be written as

$$p(y) = \sum_{n=1}^{N_d} \pi_n N(y; \boldsymbol{\mu}^n, \boldsymbol{\Sigma}^n), \tag{10}$$

where π_n represents the weights of each Gaussian distribution of all samples; $\boldsymbol{\Sigma}^n$ and $\boldsymbol{\mu}^n$ are the covariance matrix and the vector of means of all samples. Then, the probability of each type G_j can be obtained

$$p(G_j) = \frac{p(y)}{p(y|G_j)} = \frac{\sum_{n=1}^{N_d} \pi_n N(y; \boldsymbol{\mu}^n, \boldsymbol{\Sigma}^n)}{\sum_{d=1}^{N_d} \pi_d N(y; \boldsymbol{\mu}_j^d, \boldsymbol{\Sigma}_j^d)}. \tag{11}$$

The type G_j of given sample y by comparing the probability of each category:

$$G_j \longrightarrow \arg\max_j P(G_j). \tag{12}$$

The algorithm used for solving PFDA is as follows:

Algorithm 1 PFDA algorithm for estimating shale oil sweet spots.

Input: Data matrix $\boldsymbol{x}_j^{(k)}$ of each class G_j

1: Calculate the within scatter matrix $\boldsymbol{S}_w$ in (3) and the between scatter matrix $\boldsymbol{S}_b$ in Eq. (4).

2: Calculate the discriminative projection matrix $\boldsymbol{W}$ by maximizing the function given in Eq. (2).

3: Obtain the samples $y_j^{(k)}$ after projection of the raw data set by calculating the function given in Eq. (8).

4: Calculate the means vectors and covariance matrix, $\boldsymbol{\mu}_j^d$, $\boldsymbol{\Sigma}_j^d$, $\boldsymbol{\Sigma}^n$, $\boldsymbol{\mu}^n$.

5: Estimate the shale oil sweet spots by the discriminant probability function given in Eqs. (11) and (12).

End

2.2 Pyrolysis S1, TOC, clay content and porosity estimation from elastic impedance using Bayesian inversion

This study adopts Bayesian inversion to predict the pyrolysis S1, total organic carbon, clay content, and porosity from the elastic impedance data sets. In the Bayesian method, the posterior distribution of model attributes $p(\boldsymbol{m}|\boldsymbol{d})$ can be acquired through calculation of the prior distribution $p(\boldsymbol{m})$ and the likelihood function $p(\boldsymbol{d}|\boldsymbol{m})$ (Grana et al., 2017):

$$p(\boldsymbol{m}|\boldsymbol{d}) = \frac{p(\boldsymbol{d}|\boldsymbol{m})p(\boldsymbol{m})}{p(\boldsymbol{d})}, \tag{13}$$

where $\boldsymbol{d} = [EI_1, EI_2, EI_3]$ and $\boldsymbol{m} = [S1, \mathrm{TOC}, \mathrm{clay}, \mathrm{por}]$ are the elastic impedance and the model parameters, respectively. The statistical rock-physics model can be written as

$$[EI_1, EI_2, EI_3] = \mathbf{f}_{\mathrm{RPM}}(S1, \mathrm{TOC}, \mathrm{clay}, \mathrm{por}) + \varepsilon, \tag{14}$$

where ε is the error that represents the difference between model predictions and real data; and $\mathbf{f}_{\mathrm{RPM}}$ is the rock-physics model.

The posterior distribution is affected by observed data $\boldsymbol{d}$ only through $p(\boldsymbol{d}|\boldsymbol{m})$, which is regarded as a function of $\boldsymbol{m}$, referred to as likelihood function (Li et al., 2020):

$$p(\boldsymbol{d}|\boldsymbol{m}) = \prod_{j=1}^{j} p(\boldsymbol{d}_j|\boldsymbol{m}). \tag{15}$$

We assume that the prior distribution is a Gaussian mixture model (GMM) with finite components N_c, which can be expressed as:

$$p(\boldsymbol{m}) = \sum_{n=1}^{N_c} \pi_n N(\boldsymbol{m}; \boldsymbol{\mu}_{\boldsymbol{m}}^n, \boldsymbol{\Sigma}_{\boldsymbol{m}}^n), \tag{16}$$

where N is the Gaussian distribution; $\boldsymbol{\mu}_{\boldsymbol{m}}^n$, $\boldsymbol{\Sigma}_{\boldsymbol{m}}^n$ and π_n represent the vector of prior means, the covariance matrix and the weights of each Gaussian distribution, respectively.

Finally, the posterior distribution can be represented by a mixture Gaussian distribution:

$$p(\boldsymbol{m}|\boldsymbol{d}) = \sum_{k=1}^{N_c} \lambda_{\boldsymbol{m}|\boldsymbol{d}}^k N(\boldsymbol{m}; \boldsymbol{\mu}_{\boldsymbol{m}|\boldsymbol{d}}^k, \boldsymbol{\Sigma}_{\boldsymbol{m}|\boldsymbol{d}}^k), \tag{17}$$

where $\lambda_{\boldsymbol{m}|\boldsymbol{d}}^k$ are mixture coefficients of the posterior, $\boldsymbol{\Sigma}_{\boldsymbol{m}|\boldsymbol{d}}^k$ are the covariance matrices of each Gaussian component, and $\boldsymbol{\mu}_{\boldsymbol{m}|\boldsymbol{d}}^k$ are the conditional means. The petrophysical parameters can be attained by maximizing the posterior distribution:

$$m^* \longrightarrow \underset{m}{\mathrm{argmax}}\, p(\boldsymbol{m}|\boldsymbol{d}). \tag{18}$$

2.3 Elastic impedance inversion for Young's modulus and Poisson's ratio

YPD function was used to estimate Young's modulus and Poisson's ratio and the properties were extracted from the elastic impedance inversion using the EVA-DSVD method proposed by Zong et al. (2013):

$$EI(\theta) = A_0 \left(\frac{E}{E_0}\right)^{a(\theta)} \left(\frac{\sigma}{\sigma_0}\right)^{b(\theta)} \left(\frac{\rho}{\rho_0}\right)^{c(\theta)}. \tag{19}$$

If we take the logarithm on both sides of Eq. (19), then the Eq. (19) can be rewritten as

$$\ln\frac{EI(\theta)}{A_0} = a(\theta)\ln\left(\frac{E}{E_0}\right) + b(\theta)\ln\left(\frac{\sigma}{\sigma_0}\right) + c(\theta)\ln\left(\frac{\rho}{\rho_0}\right), \tag{20}$$

where A_0, σ_0, ρ_0, and E_0 are the statistic average of elastic impedance, Poisson's ratio, density and Young's modulus, respectively; $a(\theta)$, $b(\theta)$, and $c(\theta)$ are the coefficients associated with incident angles:

$$a(\theta) = \frac{1}{4}\sec^2\theta - 2k\sin^2\theta, \tag{21}$$

$$b(\theta) = \frac{1}{4}\sec^2\theta\frac{(2k-3)(2k-1)^2}{k(4k-3)} + 2k\sin^2\theta\frac{(2k-1)}{(4k-3)}, \tag{22}$$

$$a(\theta) = \frac{1}{2} - \frac{1}{4}\sec^2\theta, \tag{23}$$

Poisson's ratio, density, and Young's modulus can be determined from elastic impedances using EVA-DSVD method and Eq. (20).

Figure 1 illustrates the summarized workflow described thus far.

3 Results

The method proposed is applied to a shale oil area in southern China. The well logs in Fig. 2 show organic content, pyrolysis S1, clay content, porosity, Poisson's ratio, and Young's modulus. Based on the lithofacies, petrophysical curves and available sedimentologic information, the distribution of shale oil facies can be divided into three different shale oil facies (G_1,G_2,G_3), as shown in Fig. 3. The classification is carried out by three steps: 1) Sand and clay facies were distinguished based on the lithofacies; 2) The distribution of G_3 was obtained by differentiating the facies containing a TOC smaller than 1.8 and the clay content below 0.36; 3) Then, the combination of porosity and pyrolysis was used as the means to classify the remaining part, where S1 > 7 mg/g and porosity > 0.065 are indicators of the facies, and the rest is of G_2 type.

The mean and standard deviation of various parameters are listed in Fig. 3, where variation in the statistical parameters is evident for different shale oil facies. For instance, G_1 has a relative high value of TOC, pyrolysis S1 and Poisson's ratio compared with G_2 and G_3. Besides, it has a lower value of Young's modulus. The porosity of G_1 and G_2 is relatively lower than G_3, but G_1 is slightly higher than G_2.

Figure 4 displays the cross plot of the key evaluation parameters of the shale oil facies. Intersection analysis was performed on the pairwise plots of the features to primarily examine the influence of each parameter on the classification of the oil shale facies. The results show that each pair of the parameters can simply distinguish the facies of G_1, G_2, and G_3. However, some overlap areas are detected in the cross-plots, which may cause errors in the characterization of shale oil sweet spots. The rock physics analysis results demonstrated in Figs. 4(b)–4(f), and Fig. 4(m) suggest that high-quality shale oil face G_1 has a relatively high value of TOC, S1, clay content, and Poisson's ratio. Figure 4(a) and Fig. 4(i) indicate that high Poisson's ratio and low Young's modulus are the significant features of G_1. In theory, high porosity is beneficial for a shale oil face, but there is no obvious difference in porosity in each facies according to Fig. 4(e), Fig. 4(j), Fig. 4(n), and Fig. 4(h).

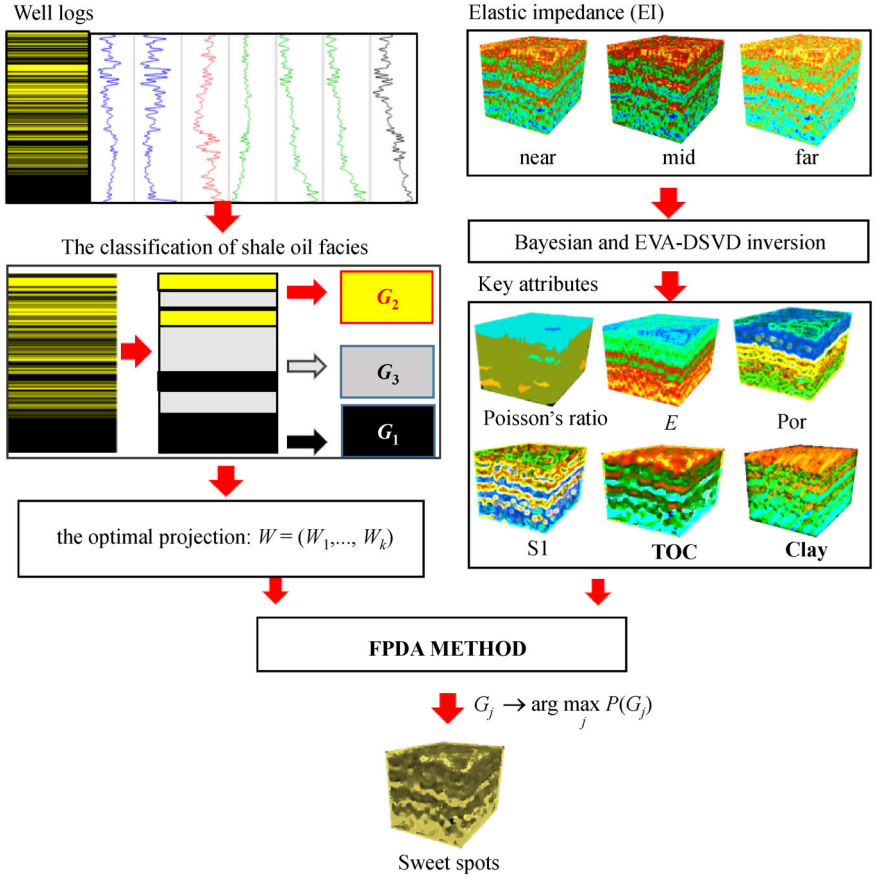


Fig. 1 The workflows of the probabilistic Fisher discriminant approach for estimating shale oil sweet spots.

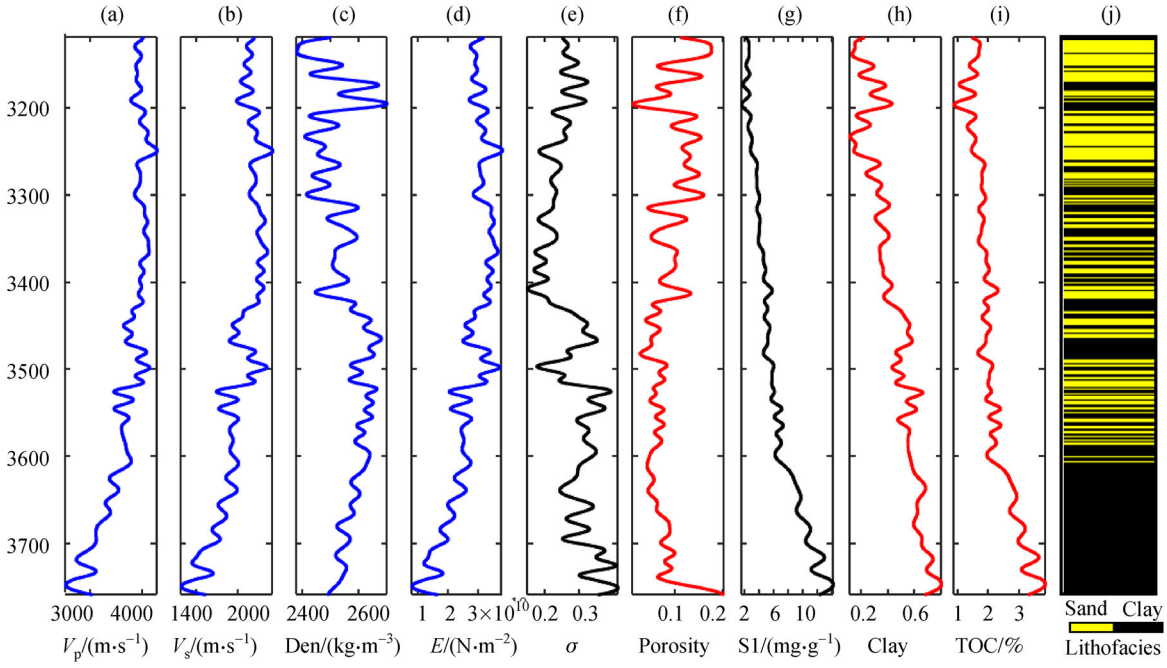
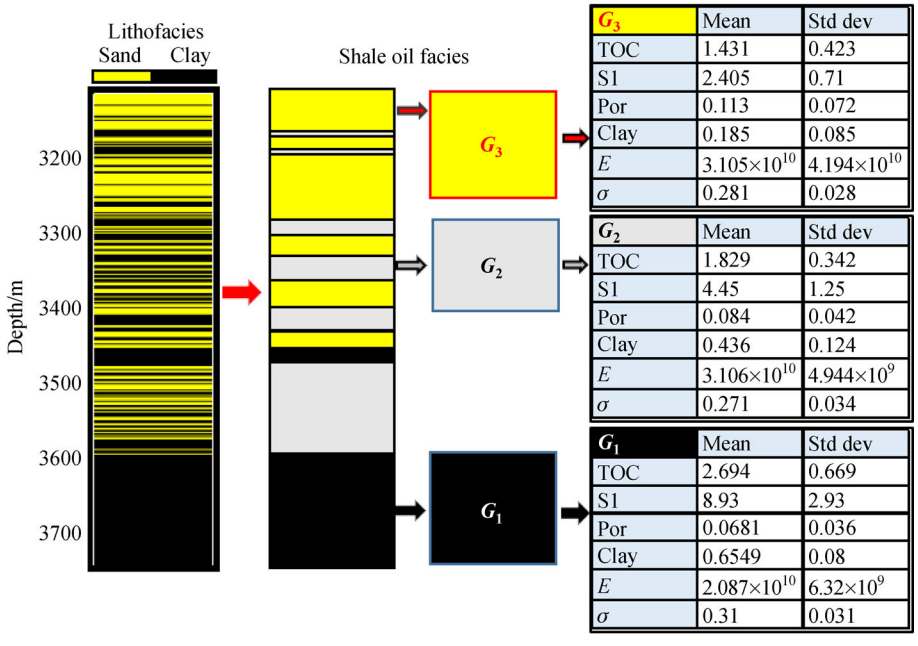


Fig. 2 From left to right: (a) P-wave velocity, (b) S-wave velocity, (c) density, (d) Young's modulus, (e) Poisson's ratio, (f) porosity, (g) pyrolysis S1, (h) clay, (i) total organic carbon, and (j) lithofacies distribution.



G_3	Mean	Std dev
TOC	1.431	0.423
S1	2.405	0.71
Por	0.113	0.072
Clay	0.185	0.085
E	3.105×10^{10}	4.194×10^{10}
σ	0.281	0.028

G_2	Mean	Std dev
TOC	1.829	0.342
S1	4.45	1.25
Por	0.084	0.042
Clay	0.436	0.124
E	3.106×10^{10}	4.944×10^{9}
σ	0.271	0.034

G_1	Mean	Std dev
TOC	2.694	0.669
S1	8.93	2.93
Por	0.0681	0.036
Clay	0.6549	0.08
E	2.087×10^{10}	6.32×10^{9}
σ	0.31	0.031

Fig. 3 The classification of shale oil facies and the statistical characteristics of key evaluation parameters of shale oil; the three types are (G_1) organic-rich mudstones with high maturity, (G_2) organic-lean intervals of sand, and (G_3) organic-rich mudstones with low maturity.

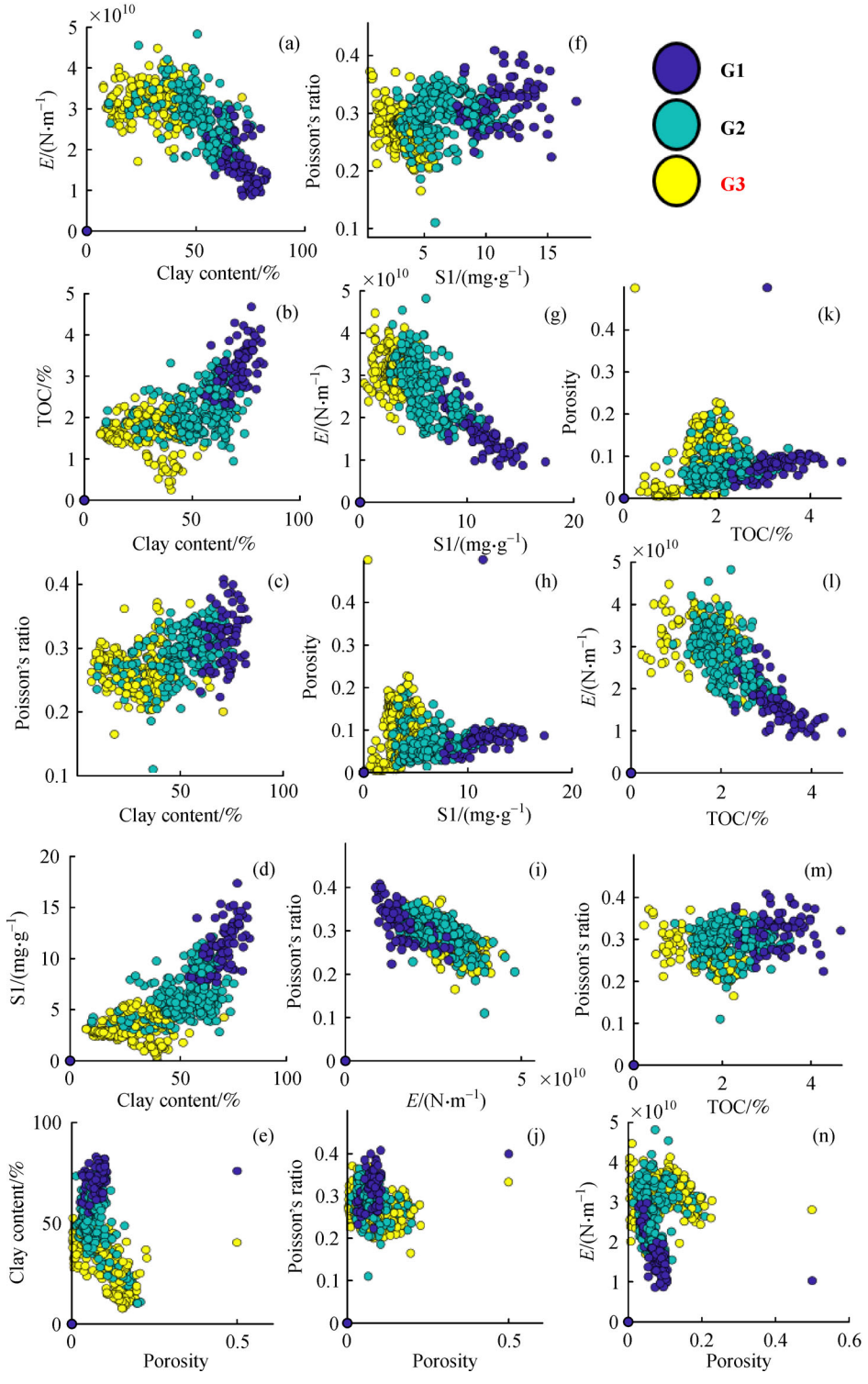


Fig. 4 Cross plot of key parameters of shale oil facies: (a) Young's modulus and clay content, (b) total organic carbon (TOC) and clay content, (c) Poisson's ratio and clay content, (d) S1 and clay content, (e) Porosity and clay content, (f) Poisson's ratio and clay content, (g) Young's modulus and S1, (h) Porosity and S1, (i) Poisson's ratio and Young's modulus, (j) Poisson's ratio and porosity, (k) Porosity and TOC, (l) Young's modulus and TOC, (m) Poisson's ratio and TOC, (n) Porosity and Young's modulus. The yellow circles, black circles and dark blue circles represent the facies of G_1, G_2, and G_3, respectively.

After using Fisher dimension reduction, the original shale oil key parameters, which have six dimensions (TOC, Clay, S1, Por, E, σ), are turned to a new group of two-dimensional data sets as shown in Fig. 5, where G_1, G_2, and G_3 are depicted by circles of different colors. The coordinate "dimension 1" and "dimension 2" in Fig. 5 can be obtained by function 9. The optimal projection calculated by function 3 is

$$\boldsymbol{W} = \begin{bmatrix} 0.1716 & 0.0442 & -0.2896 & 0.0216 & -0.00096 & 0.9403 \\ 0.1073 & 0.0668 & 0.3668 & 0.0246 & -0.0163 & 0.9212 \end{bmatrix}^{\mathrm{T}}. \quad (24)$$

Figure 5 demonstrates that the Fisher method can effectively realize the reduction of high-dimensional data, and furtherly separate different types of facies. Figure 5 shows a more integrated result, where all the previous factors are accounted for in generating the cross plot. After dimension reduction, the six factors of shale oil are converted to two-dimensional data sets in Fig. 5. There is some overlap between G_1 and G_2 due to the initial classification errors, but their core distributions are different. Gaussian mixture model in Eq. (9) can be used to rebuild the distribution of each type based on the new two-dimensional data.

The optimal projection $\boldsymbol{W}$ obtained in Eq. (24) was applied to the inversion results of shale oil key attributes depicted in Fig. 6 and two-dimensional data sets were generated. Figures 6(a)–6(f) are inversion results of clay fraction, total organic carbon content, Young's modulus, pyrolysis S1, Poisson's ratio, and porosity, respectively. Figure 7(a) is the delineation of sweet spots from TOC results shown in Fig. 6(b), where TOC $>$ 2.1 is the predetermined criterion for the good shale oil sweet spots. Fig. 7(b) displays the sweet spots of shale oil reservoirs based on the classification of shale oil facies, determined by the PFDA method. The results depicted in Fig. 7(b) are validated by the well logs, where the final result obtained by the FPDA method is in good agreement with the well log. Although, G_1 exhibits a smaller distribution range in Fig. 7(b), the best sweet spots shown in Fig. 7(a) are consistent with the best shale oil face G_1 in Fig. 7(b).

4 Conclusions

A PFDA method was proposed to differentiate the shale oil sweet spots in this study, in which the probabilistic method and Gaussian mixture model are combined with Fisher's method. This method was successfully used to delineate the sweet spots of shale oil reservoirs from the real pre-stack seismic data sets and is validated by the well log data. Additionally, a new workflow for the evaluation of shale oil sweet spots was introduced. Six key attributes of the shale oil sweet spots were selected based on the rock physics analysis, and these features were inversed by the Bayesian and EVA-DSVD methods. The PFDA was finally used to utilize more comprehensive and low-dimensional data to effectively delineate the sweet spots.

The accuracy of the PFDA method is relatively sensitive to the prior classifications. While significant errors in the inversion process can adversely affect the accuracy of the results, minimal error would not have a significant impact on the accuracy of the results generated by the PFDA method.

G_1	Mean	Std dev
Dimension 1	137.13	11.69
Dimension 2	38.5274	4.79

G_2	Mean	Std dev
Dimension 1	108.4	12.5371
Dimension 2	42.942	4.2

G_3	Mean	Std dev
Dimension 1	80.63	5.4
Dimension 2	36.25	2.76

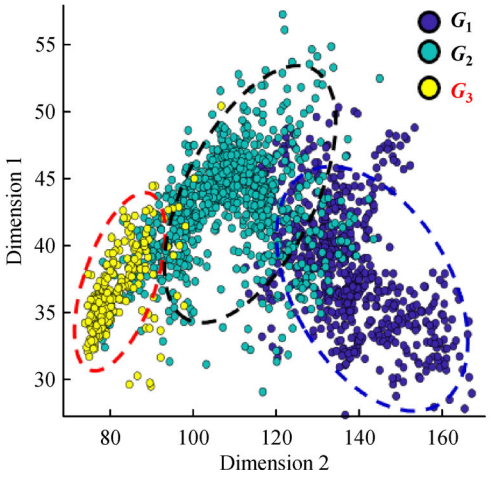


Fig. 5 The mean and standard deviation of new two-dimensional data sets of each type and the cross plot of shale oil facies after Fisher dimension reduction. The coordinate "dimension 1" and "dimension 2" are obtained by the dimension reduction of model parameters.

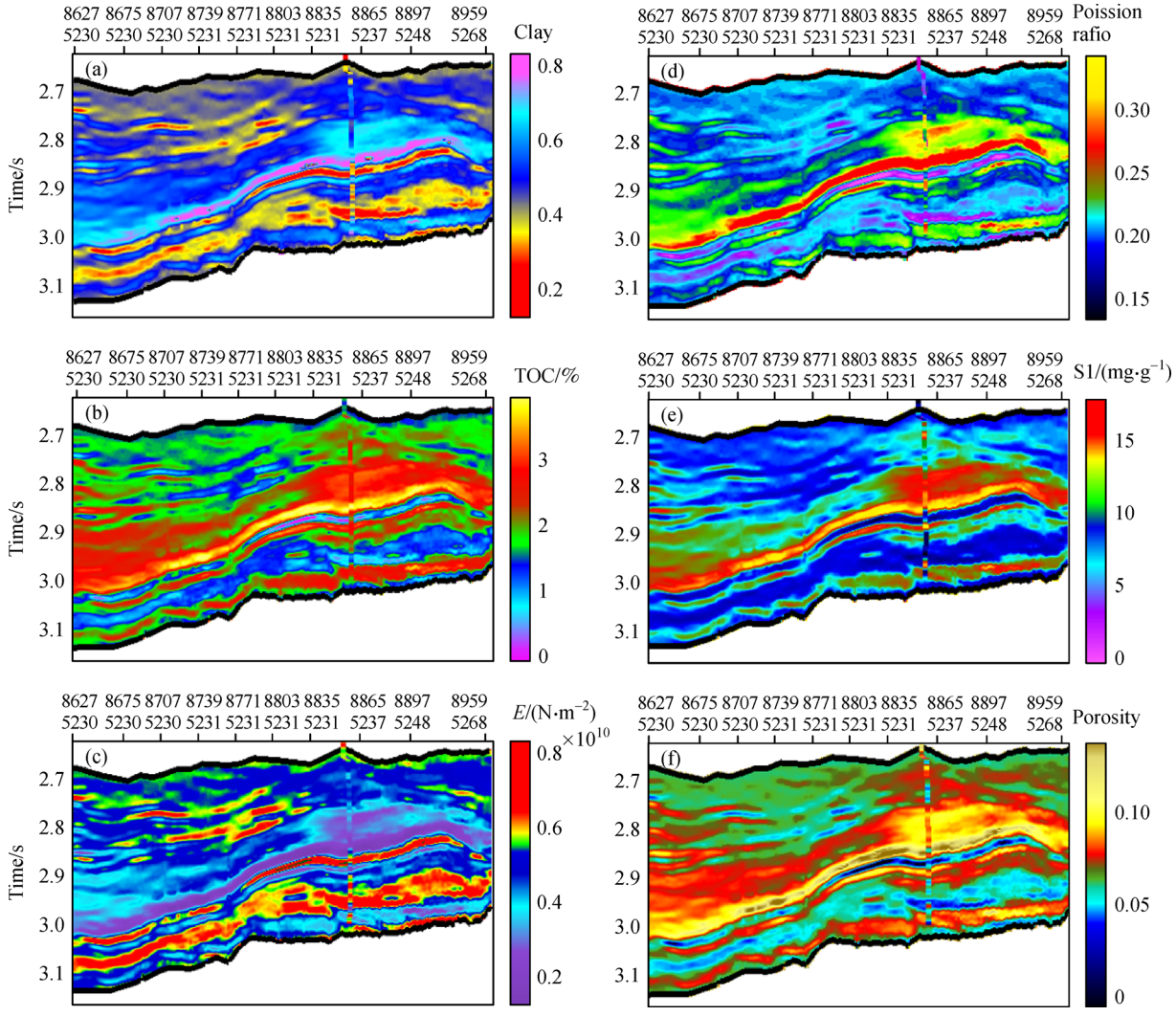


Fig. 6 The inversion results of six key shale oil parameters, (a) clay content, (b) total organic carbon, (c) Young's modulus, (d) pyrolysis S1, (e) Poisson's ratio, and (f) porosity.

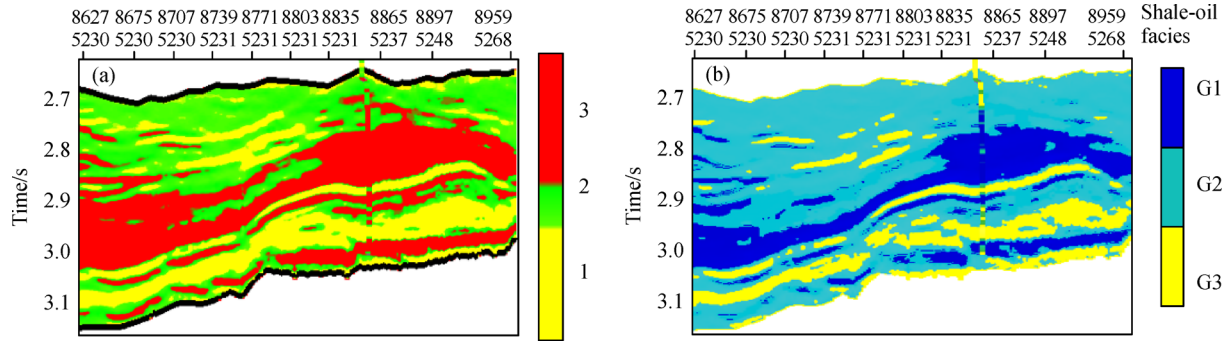


Fig. 7 The comparison of the delineation of shale oil sweet spots from (a) single TOC attribute and (b) the delineation obtained by FPDA method from multiple attributes.

Acknowledgements We would like to acknowledge the sponsorship of the National Natural Science Foundation of China (Grant Nos. 41974119 and 42030103) and Science Foundation from Innovation and Technology Support Program for Young Scientists in Colleges of Shandong Province and Ministry of Science and Technology of China.

References

AbuZeina D, Al-Anzi F S (2018). Employing fisher discriminant analysis for Arabic text classification. Comput Electr Eng, 66: 474–486

Bao Y S, Zhang L Y, Zhang J G, Li J Y, Li Z (2016). Factors influencing mobility of Paleogene shale oil in Dongying Sag, Bohai Bay Basin. Oil & Gas Geo, 37(3): 408–414

Chen S, Zhao W Z, Zeng Q C, Yang Q, He P, Gai S H, Deng Y (2018). Quantitative prediction of total organic carbon content in shale-gas reservoirs using seismic data: a case study from the Lower Silurian Longmaxi Formation in the Chang Ning gas field of the Sichuan Basin, China. Interpretation (Tulsa), 6(4): SN153–SN168

Chen J Q, Pang X Q, Wang X L, Wang Y X (2020). A new method for assessing tight oil, with application to the Lucaogou Formation in the Jimusaer depression, Junggar Basin, China. AAPG Bull, 104(6): 1199–1229

Connolly P (1999). Elastic impedance. Leading Edge (Tulsa Okla), 18(4): 438–452

Espitalie J, Laporte J L, Madec M, Marquis F, Leplat P, Paulet J, Boutefeu A (1977a). Rapid method for source rock characterization and for determination of their petroleum potential and degree of evolution. Oil & Gas Sci Techn Revue, 32(1): 23–42

Fisher R A (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7: 179–188

Foley D H, Sammon J W (1975). An optimal set of discriminant vectors. IEEE Trans Comput, C-24(3): 281–289

Gorynski K E, Tobey M, Enriquez D, Smagala T, Dreger J, Newhart R (2019). Quantification and characterization of hydrocarbon-filled porosity in oil-rich shales using integrated thermal extraction, pyrolysis, and solvent extraction. AAPG Bull, 103(3): 723–744

Grana D, Della Rossa E (2010). Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion. Geophys, 75(3): O21–O37

Grana D, Fjeldstad T, Omre H (2017). Bayesian Gaussian mixture linear inversion for geophysical inverse problems. Math Geosci, 49(4): 493–515

Guo R Y, Bai Y Q, Li C N, Shao Y H, Ye Y F, Jiang C Z (2021). Reverse nearest neighbors Bhattacharyya bound linear discriminant analysis for multimodal classification. Eng Appl Artif Intell, 97: 104033

Hu R, Vernik L, Nayvelt L, Dicman A (2015). Seismic inversion for organic richness and fracture gradient in unconventional reservoirs: Eagle Ford Shale, Texas. Leading Edge (Tulsa Okla), 34(1): 80–84

Huang R Z (1984). A model for predicting formation fracture pressure. J China U Petrol (Nat Sci), 8: 335–347

Jiang C R, Chen L H (2020). Filtering-based approaches for functional data classification. Wiley Interdiscip Rev Comput Stat, 12(4): 1–15

Jarvie D M (2012). Shale resource systems for oil and gas: part 2—shale-oil resource systems. AAPG Mem, 97: 89–119

Luo K, Zong Z Y (2020). Pyrolysis S1 discrimination of shale gas with Pre-stack seismic inversion. In: Annual Meeting of Chinese geoscience union 2020, Beijing

Li K, Yin X Y, Zong Z Y (2020). Facies-constrained pre-stack seismic probabilistic inversion driven by rock physics. Scientia Sinica Terrae, 50(6): 832–850

Luo K, Zong Z Y (2019). Sweet spots discrimination of shale gas with Pre-stack seismic inversion. In: Annual Meeting of Chinese Geoscience Union 2019, Beijing, 53: 29–31

Ma L, Lin Z L, Hu H F, Zhou D (2020). Seismic prediction method of fracture pressure in a shale formation. In: SEG International Exposition and 90th Annual Meeting, 1068–1072

Ouadfeul S, Aliouane L (2016). Total organic carbon estimation in shale-gas reservoirs using seismic genetic inversion with an example from the Barnett Shale. Leading Edge (Tulsa Okla), 35(9): 790–794

Ogiesoba O, Hammes U (2014). Seismic-attribute identification of brittle and TOC-rich zones within the Eagle Ford Shale, Dimmit County, South Texas. J Pet Explor Prod Technol, 4(2): 133–151

Raji M, Gröcke D R, Greenwell C, Cornford C (2015). Pyrolysis, porosity and productivity in unconventional mudstone reservoirs: free and adsorbed oil. In: Unconventional Resources Technology Conference, 11: 270–279

Rickman R, Mullen M J, Petre J E, Grieser W V, Kundert D (2008). A practical use of shale petrophysics for stimulation design optimization: all shale plays are not clones of the Barnett Shale. In: SPE Technical Conference and Exhibition

Rao C R (1948). The utilization of multiple measurements in problems of biological classification. J R Stat Soc B, 10(2): 159–193

Sena A, Castillo G, Chesser K, Voisey S, Estrada J, Carcuz J, Carmona E, Hodgkins P (2011). Seismic reservoir characterization in resource shale plays: "sweet spot" discrimination and optimization of horizontal well placement. In: 81st Annual International Meeting, SEG, Expanded Abstracts, 1744–1748.

Song G Z, Lin Y, Lu S F (2013). Resource evaluation method for shale oil and its application. Earth Sci Front, 20(4): 221–228

Verma S, Zhao T, Marfurt K J, Devegowda D (2016). Estimation of total organic carbon and brittleness volume. Interpretation (Tulsa), 4(3): T373–T385

Wang H J (2017). Prediction of geological dessert in shale gas with pre-stack seismic inversion. Dissertation for the Master's Degree. Qingdao: China University of Petroleum (East China)

Wu L, Shen C H, Hengel A (2017). Deep linear discriminant analysis on fisher networks: a hybrid architecture for person re-identification. Pattern Recognit, 65: 238–250

Wang S, Feng Q, Javadpour F, Xia T, Li Z (2015). Oil adsorption in shale nanopores and its effect on recoverable oil-in-place. Int J Coal Geol, 147–148: 9–24

Wang H, Yan S C, Xu D, Tang X O, Huang T (2007). Trace ratio vs. ratio trace for dimensionality reduction. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE, 1–8

Wang Z, Ruan Q, An G (2016). Facial expression recognition using sparse local Fisher discriminant analysis. Neurocomputing, 174: 756–766

Yin X Y, Cui W, Zong Z Y, Liu X J(2014). Petrophysical property inversion of reservoirs based on elastic impedance. Chinese J

Geophys, 57(12): 4132–4140

Yu J Q, Yu Z Q, Mao Z Q,Gao G, Luo K, Lei T, Zong Z Y(2020). Prediction of total organic carbon content in source rock of continental shale oil using pre-stack inversion. Geophys Prospect Petrol, 59(5): 823–830

Yin X Y, Zong Z Y, Wu G C (2015). Research on seismic fluid identification driven by rock physics. Sci China Earth Sci, 58(2): 159–171

Yin X Y, Yuan S H, Zhang F C (2004). Rock elastic parameters calculated from elastic impedance. CPS/SEG Technical Program Expanded Abstracts

Zong Z Y, Yin X Y, Wu G C (2013). Elastic impedance parameterization and inversion with Young's modulus and Poisson's ratio. Geophysics, 78(6): N35–N42

Zong Z Y, Yin X Y, Zhang F,Wu G C (2012a). Reflection coefficient and pre-stack seismic inversion with Young's modulus and Poisson's ratio. Chinese J Geophys, 55(11): 3786–3794

Zong Z Y, Yin X Y, Wu G C (2012b). Pre-stack inversion for rock brittleness indicator in gas shale. In: Annual meeting of Chinese Geoscience Union 2012. Beijing, 17: 495