The Attribute Optimization Method Based on the Probability Kernel Principal Component Analysis and Its Application

Jingjing Zheng[*], **Yun Wang, Chunying Yang**

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

Email address:

zhjjm1983@163.com (Jingjing Zheng), wangyun@mail.gyig.ac.cn (Yun Wang), yangchychn@cugb.edu.cn (Chunying Yang)

[*]Corresponding author

Abstract: The principal component analysis (PCA) is the most common attribute optimization analysis techniques, but it is a linear method and exists the problem of lack of probability model and the absence of higher-order statistics information. It has poor comprehensive ability to complex non-linear attributes. Therefore, in order to overcome two shortcomings of the principal component analysis (PCA) and improve the effect of attribute optimization, this paper studies the probability kernel principal component analysis (PKPCA) method which is based on Bayesian theory and kernel principal component analysis (KPCA). First, the sample data are mapped to the high dimensional feature space, then define probability model of the data in high-dimensional space, and finally, expectation maximization (EM) estimated is used to get the best results. This method has both the advantage of probability analysis and kernel principal component analysis (KPCA). It is able to effectively adapt to more complex reservoir conditions and can realize the non-linear probability analysis. The probability kernel principal component analysis (PKPCA) method is applied to reservoir prediction of the Southern oil fields in China. The predicted results show that the method can improve the precision of the attribute optimization, while improving the accuracy of the forecasts of reservoir.

Keywords: Kernel Principal Component Analysis, The Probability Analysis, Kernel Function, Attribute Optimization Analysis, Reservoir Prediction

1. Introduction

In recent years, seismic attribute analysis has become an important mean for people to understand and monitor oil and gas reservoirs, mainly because it can effectively reflect the lithology and physical properties (such as porosity, permeability, shale content, oil saturation, etc.) information contained in seismic data. The sensitivity of different seismic attributes to different lithologic reservoir parameters is different, and the role of different seismic attributes in describing different objects is also different. In the process of seismic reservoir prediction, various seismic attributes related to reservoir prediction are usually introduced, but the infinite increase of attribute types will also bring adverse effects on reservoir prediction [1-2]. Therefore, it is necessary to select the least number of seismic attributes or their combinations which are most sensitive (or most effective and representative)

to solve specific problems, enhance the prediction accuracy of seismic reservoirs, and improve the effect of processing and interpretation methods related to seismic attributes, which is called "seismic attribute optimization".

Among many attribute optimization methods, principal component analysis (PCA) is the most widely used statistical analysis technology, such as data compression, feature extraction and pattern recognition. However, PCA has two drawbacks [3, 4]: (1) lack of probabilistic model structure, which is often very important for many research objects, such as hybrid simulation and Bayesian decision, (2) linear nature, discarding the information of higher-order statistics.

In order to overcome these two shortcomings of PCA, people have proposed a series of improvement methods. For example, Tipping et al. [5] first proposed the Probabilistic Principal Component Analysis (PPCA) method. It allowed the noise components to have an isotropic structure, and used the maximum likelihood estimation (MLE) method to add the

parameter probability learning stage to the model, which overcomes its first shortcoming. Subsequently, Schölkopf [6] proposed the kernel principal component analysis (KPCA) method to overcome its second shortcomings. The key idea of KPCA is to use kernel function to avoid calculating inner product in high-dimensional feature space directly. Although it uses implicit form to do nonlinear mapping, it can capture high-order statistical information [7].

For the two shortcomings of PCA, this paper focuses on the method of probability kernel principal component analysis (PKPCA), which captures the statistical features of nonlinear high-dimensional feature space by the Bayesian probability model. And it combines the advantages of PPCA and KPCA. Based on this theory, the attribute feature optimization analysis of actual data is carried out. The results fully prove the superiority and practicability of PKPCA method.

2. The Basic Principle of Probabilistic Kernel Principal Component Analysis (PKPCA) Method

2.1. Kernel Principal Component Analysis (KPCA)

The basic idea of kernel principal component analysis (KPCA) is to project each sample vector x_k in the input space R^N into the high-dimensional feature space F by introducing a nonlinear function φ, and then Principal Component Analysis is performed in high dimensional space [8-10]. Duo to the high dimensional characteristics of the feature space, the nonlinear mapping and the decomposition of characteristic variables become very difficult, but the nonlinear mapping can be avoided by introducing the kernel technique [11-13], and the point product of the original space variables is only needed to be calculated [6].

Setting up {x1, …, xN} is a spatial sample of Rd. Nonlinear mapping is Rd→Rf, The spatial sample Rf is as follows: Φf×N=[φ1, …, φN], φn=φ(xn)∈Rf. The mean of samples is defined as φ0 =Φe, $e_{N \times 1} = N^{-1}\mathbf{1}$. Then the covariance matrix of the high-dimensional space data φ1, …, φN can be expressed as:

$$\Sigma = \Phi J J T \Phi T = \Psi \Psi T \qquad (1)$$

Where, $J = N^{-1/2}(I_N - eI^{\mathrm{T}})$; $\psi = \Phi J$; The first principal eigenvector of Σ can be expressed as matrix form:

$$U_q = [u_1, u_2, \cdots, u_q] = \Phi J V_q \Lambda_q^{-1/2} \qquad (2)$$

Where, $V_q = [v_1, v_2, \cdots, v_q]$, $\Lambda_q = \mathrm{diag}(\lambda_1, \cdots, \lambda_q)$, And $\{(\lambda_n, v_n)\}_{n=1}^{q}$ is the first q eigenvector and eigenvalue of matrix $\bar{K} = J^{\mathrm{T}} K J$. $K = \Phi^{\mathrm{T}} \Phi$.

2.2. Probabilistic Kernel Principal Component Analysis

Probabilistic principal component analysis (PPCA) is an extension of traditional principal component analysis (PCA).

It defines an appropriate probability model for data (Figure 1), which effectively overcomes the shortcoming that the PCA transform assumes that the data satisfy ellipsoidal distribution. With the constraint of probability model, PPCA can find the direction of principal component more effectively from high-dimensional data or a large number of data points. Then the optimal probabilistic model (optimal solution) is obtained by estimating the model parameters with the maximum likelihood function or expectation maximization (EM) algorithm. PCA or PPCA is still an effective optimization method to deal with linear relation data, but it is not effective to deal with nonlinear relation variable data. Probabilistic Kernel Principal Component Analysis (PKPCA) defines the PPCA model in the high-dimensional feature space by nonlinear mapping (Figure 1 below), which is the extension of PPCA in the kernel space [14]. It effectively overcomes the lack of probability model and high-order statistics information in PCA. Therefore, PKPCA analysis can effectively deal with data variables with nonlinear relationship.
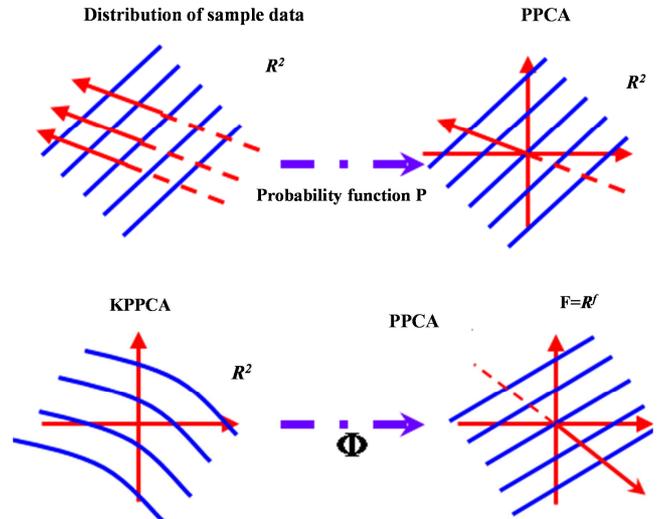

Figure 1. *Basic thinking of PKPCA.*

In probability analysis, it is assumed that the data $\phi(x)$ in the high dimensional feature space conform to a special factor analysis model [15], which makes the relationship between the f-dimensional data and the implied q-dimensional variable z be expressed as:

$$\phi(x) = Wz + \mu + \varepsilon \qquad (3)$$

Where: $z \sim N(0, I_q)$, $\varepsilon \sim N(0, \rho I_f)$, W is a $f \times r$ dimensional load matrix. Then the random probability distribution of high dimensional feature space data $\phi(x)$ varying with implied variable Z can be obtained by factor analysis model.

$$P(\phi(x) \mid z) = (2\pi\sigma^2)^{-f/2} \exp\left\{-\frac{1}{2\sigma^2} \|\phi(x) - Wz - \mu\|^2\right\} \qquad (4)$$

Further, suppose the prior model of implied variable Z is Gauss distribution.

$$P(z) = (2\pi)^{-q/2} \exp\left\{-\frac{1}{2}z^{\mathrm{T}}z\right\} \qquad (5)$$

The random probability distribution of data $\phi(x)$ is

$$P(\phi(x)) = \int P(\phi(x)|z)P(z)$$
$$= (2\pi\sigma^2)^{-f/2}|S|^{-1/2}\exp\left\{-\frac{1}{2}(\phi(x)-\mu)^{\mathrm{T}}S^{-1}(\phi(x)-\mu)\right\} \qquad (6)$$

Here S is the covariance matrix of the model.

$$S = WW^{\mathrm{T}} + \rho I_f$$

Then the maximum likelihood function of data $\phi(x)$ is

$$L = \sum_{n=1}^{N}\ln\{p[\phi(x)]\}$$
$$= -\frac{N}{2}\{f\ln(2\pi) + \ln|S| + \mathrm{tr}(S^{-1}K)\} \qquad (7)$$

Here, K is the sampling covariance matrix of data $\phi(x)$.

Estimated values of μ、W and ρ are obtained from maximum likelihood estimation.

$$\hat{\mu} = \bar{\phi}_0$$
$$\hat{W} = U_q(\Lambda_q - \hat{\rho}I_q)^{1/2}R \qquad (8)$$
$$\hat{\rho} = (f-q)^{-1}[\mathrm{tr}(S) - \mathrm{tr}(\Lambda_q)]$$

R is an arbitrary orthogonal matrix $q \times q$

Because the dimension f of kernel space is usually unknown, the $\hat{\rho}$ value is difficult to obtain accurately. In general, the f value is very large and the value $\hat{\rho}$ is very small, so it can be given a small fixed value according to experience[16].

Substituting the U_q expression (formula 2) into $\hat{W}$:

$$\hat{W} = \Phi JQ \qquad (9)$$

Where, $Q = V_q(I_q - \hat{\rho}\Lambda_q^{-1})^{1/2}R$

Because Q links the load matrix with the empirical data, it is called the empirical load matrix. Finally, the EM iterative algorithm is applied to get the estimated values Q and ρ.

$$Q^{(j+i)} = \bar{K}Q^{(j)}(\rho^{(j)}I_q + M^{-1}Q^{(j)\mathrm{T}}\bar{K}^2Q^{(j)})^{-1}$$
$$\rho^{(j+i)} = \frac{1}{f}\mathrm{tr}(\bar{K} - \bar{K}Q^{(j)}M^{(j)^{-1}}Q^{(j+i)\mathrm{T}}\bar{K}) \qquad (10)$$

Where, $M = \hat{\rho}I_q + \hat{W}^{\mathrm{T}}\hat{W}$, Substituting the expression of $\hat{W}$ to get M: $M = R^{\mathrm{T}}\Lambda_q R$

Similar to KPCA expression, the following formula can be applied in PKPCA to obtain the eigenvector $\hat{z}$ of the high-dimensional space.

$$\hat{z} = M^{-1}\hat{W}^{\mathrm{T}}(\phi(x)-\mu) \qquad (11)$$

The eigenvector of the observed sample $\phi(x)$ can be represented by a posteriori mean corresponding to the z variable. Then substituting $M, \hat{W}$ and $\hat{\mu}$, the formula can be got as following:

$$\hat{z} = R^{\mathrm{T}}L_q^{-1}R(JQ)^{\mathrm{T}}(K^t - Ke) \qquad (12)$$

Where the ith element of K^t is $\phi(x_i)^{\mathrm{T}}\phi(x) = k(x_i, x)$ $(i = 1, 2, \cdots, N)$. The dimension q of the eigenvector $\hat{z}$ can be smaller than N, so that the low-dimensional vector can be used to represent the high-dimensional original eigenvector, thereby achieving the purpose of data dimensionality reduction.

2.3. PKPCA Algorithm Implementation

The implementation of the PKPCA algorithm has the main steps as following:

(1) Standardized preprocessing of data: To overcome the differences in dimension and magnitude between various seismic attributes, the original seismic data is standardized and preprocessed to ensure the objectivity and accuracy of the optimization results. There are many methods for standardizing data. Here, standard deviation methods are used to preprocess seismic attribute parameters.

(2) Selecting kernel function: The choice of kernel function in practical applications is very important. The success of the kernel function selection directly affects the prediction results of the KPCA and PKPCA methods. In this paper, a polynomial kernel function and a Gaussian radial basis kernel function are used. In addition, the choice of the kernel function parameters is also very important, and sometimes the choice of kernel function parameters is even more important than the choice of the type of the kernel function itself.

(3) High dimensional space covariance matrix and its decomposition: Under the action of the selected kernel function, a covariance matrix is obtained by applying formula $\bar{K} = J^{\mathrm{T}}KJ$. And then he feature value $\Lambda_q = \mathrm{diag}(\lambda_1, \cdots, \lambda_q)$ and the feature vector $V_q = [v_1, v_2, \cdots, v_q]$ is obtained by eigenvalue decomposition.

(4) Estimating Q and ρ: The MLE algorithm can be used to calculate the initial value by using the formula (8), and then the eigenvalue and the feature vector obtained in step (3) are substituted into the iterative formula (10) of the EM algorithm. At last, the estimated values of A and B can be calculated quickly and accurately. Or directly apply the EM algorithm, that is, first give the initial value, and then gradually iterate to obtain the exact sum and estimate.

(5) Feature extraction: Apply (12) to extract the feature vector $\hat{z}$, and obtain a new component that characterizes the main features of the original attribute. Based on the characteristics of

these new attributes, the comprehensive geological interpretation analysis of the target reservoir is carried out.

3. Case Study

The above method is applied to real data to verify the advantages of PKPCA algorithm. The actual data is from Southern oil fields in China. Reservoirs in this block are poor or tight reservoirs, but they have good porosity and permeability in local areas (such as grain chips, oolitic and so on) and become favorable reservoir areas. Well logging data of J1 well in this area (Figure 3 black spots) is available. The lithology of the selected interval is grey dolomite with lime-bearing silty sand cuttings, which is comprehensively interpreted as gas reservoir by logging. In case, the amplitude kurtosis, amplitude torsion, amplitude arc length, effective bandwidth and average instantaneous frequency are extracted along the target layer of the block (Figure 2). PCA, KPCA and PKPCA are used to optimize the attributes, and the final comparison is shown in Figure 3. Where, the two kernel functions used in the PKPCA optimization method are:

Kernel function 1 $k(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\dfrac{1 \times \|\boldsymbol{x} - \boldsymbol{y}\|^2}{50}\right)$,

Kernel function 2 $k(\boldsymbol{x}, \boldsymbol{y}) = \left(0.5 \times (\boldsymbol{x} \cdot \boldsymbol{y}) + 4\right)^2$


(a)

(b)

(c)

(d)

(e)
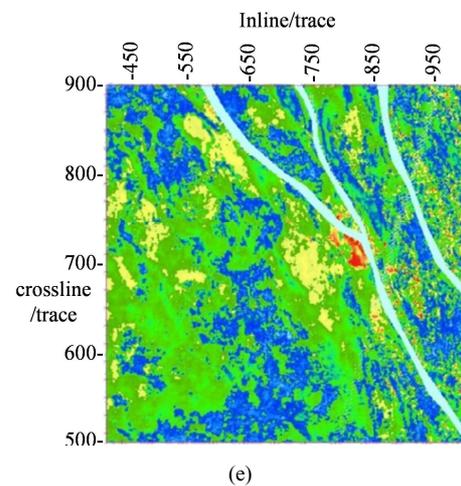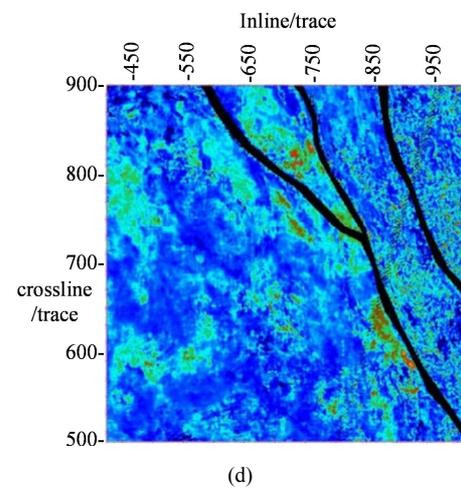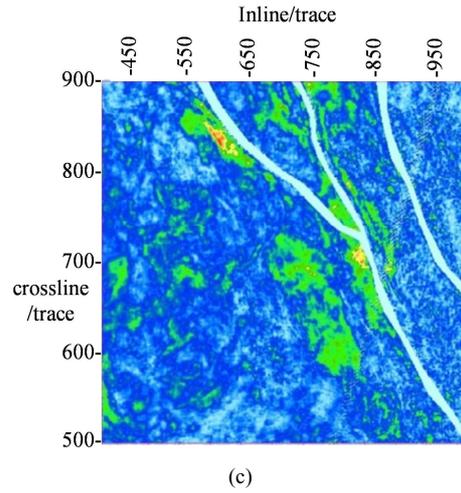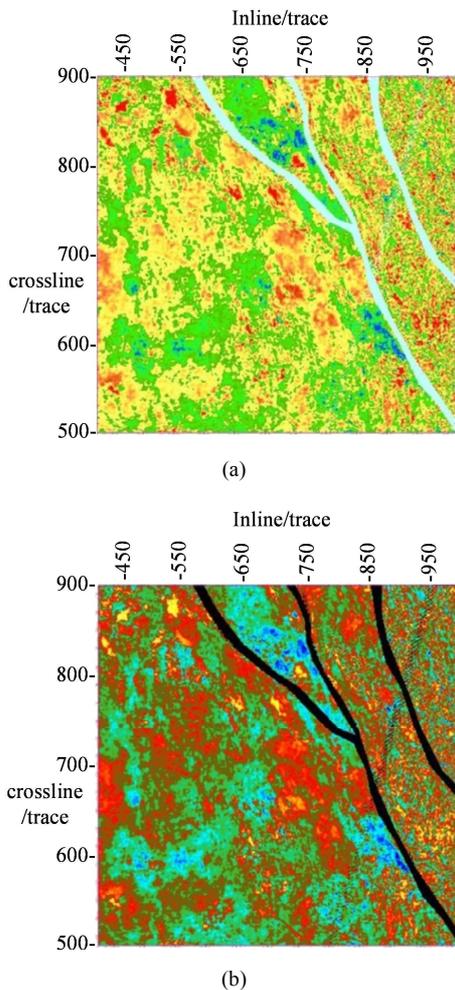
Figure 2. *Original seismic attribute (a) Amplitude Kurtosis, (b) amplitude torsion, (c) amplitude arc length, (d) effective bandwidth, (e) average instantaneous frequency.*

Figure 3 is a comparison of attribute optimization results obtained by various methods. It can be clearly seen that the effect of feature extraction by directly using PCA algorithm for attribute optimization is not ideal, which can not effectively integrate the commonalities of a large number of attributes, and does not achieve the purpose of effective attribute optimization.

KPCA algorithm based on radial basis function (Figure 3b) can extract the approximate range of reservoir, but it is not accurate enough. The PKPCA algorithm (Figure. 3C and Figure. 3d) gets the best probability solution by adding probability model and introducing kernel function, and using the discarded principal component as noise variance estimation to constrain factor load matrix W. It can delineate accurately the reservoir range and coincide with the result of J1 logging interpretation, thus achieving the effective attribute optimization.


(a)

(b)

(c)

(d)
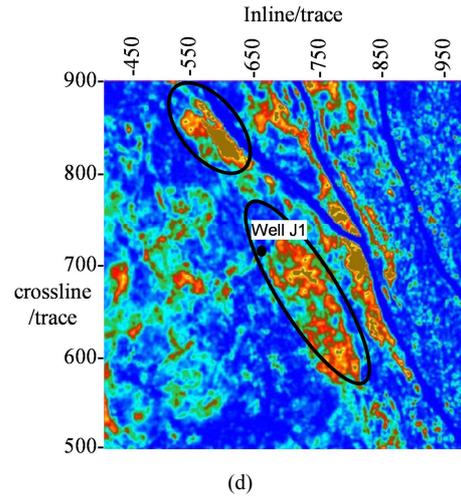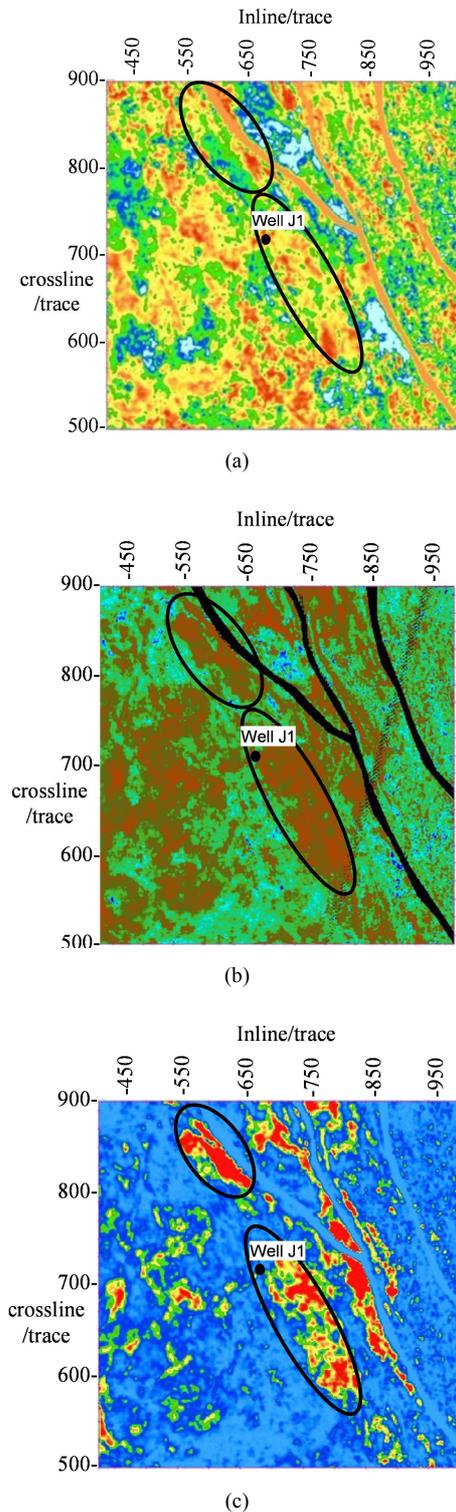
Figure 3. *Comparison of attribute optimization results by different methods (a)PCA, (b)KPCA(RBF kernel), (c)PKPCA(RBF kernel), (d)PKPCA (Polynomial basis——kernel function 2).*

It can be seen from case study that the optimization results of the polynomial kernel function and the radial basis function are similar, which can describe the reservoir range well. But the optimization results of polynomial kernel function describe the range more accurately, and the calculation efficiency is higher than that of radial basis function. Therefore, the kernel function selection of PKPCA algorithm is very important. If not properly selected, it will not only waste time, but also seriously affect the accuracy of reservoir prediction.

4. Conclusions

(1) As an extension of PPCA in high dimensional space, PKPCA defines PPCA probability model in high dimensional feature space by nonlinear mapping. It not only overcomes the shortcoming of lacking probability statistics for PCA linear algorithm, but also captures the high-order statistical information of attributes, which effectively improves the ability of non-linear simulation of attributes and the accuracy of attribute optimization. Its additional probability structure also provides an effective classification method.

(2) PKPCA method is applied to optimize and analyze a southern oilfield data. Under the same parameters, its processing result is better than PCA and KPCA algorithm, which is more sensitive to oil and gas reservoirs, can describe the favorable distribution range of reservoirs more precisely, and is consistent with logging interpretation results, and has high accuracy of reservoir prediction. This provides a new idea and approach for multi-attribute optimization and complex geological reservoir prediction.

(3) Compared with KPCA and PCA, PKPCA has higher computational complexity and lower efficiency, and the effect of attribute optimization depends more on the kernel function. However, in the case of complex data, the optimization effect of PKPCA can not be replaced by other methods. In practical application, the optimization algorithm and parameters should be reasonably selected according to the complexity of the

target area in order to save time and effort and achieve the desired results.

Acknowledgements

References

[1]   X. Y. Yin, J. Y. Zhou. Review of seismic attribute optimization methods, OGP, 2005, vol.4, pp. 258-262.

[2]   S. Chopra and K. J. Marfurt. Seismic facies classification using some unsupervised machine-learning methods, SEG Technical Program Expanded Abstracts 2018: pp. 2056-2060.

[3]   L. F Liu, S. Z. D. Sun, H. Y. Wang, and H. J. Yang, J. F. Han, and B. Jing. 3D Seismic attribute optimization technology and application for dissolution caved carbonate reservoir prediction, SEG Technical Program Expanded Abstracts 2011: pp. 1968-1972.

[4]   S. K. Zhou, R. Chellappa, W. Zhao. Unconstrained face recognition. Springer-Verlag, New York Inc, 2006.

[5]   M. E. Tipping, C. M. Bishop, Probabilistic principal component analysis. Journal of the Royal Statistical Society, Series B (Statistical Methodology), 1999, vol.61, pp.611-622.

[6]   B. Schölkopf, A. Smola, K. R. Müller. Nonlinear component analysis as a kernel eigenvalue problem, Neural computation, 1998, vol.10, pp.1299-1319.

[7]   S. S. Dehsoukhteh. Compare and Evaluate the Performance of Gaussian Spatial Regression Models and Skew Gaussian Spatial Regression Based on Kernel Averaged Predictors, American Journal of Theoretical and Applied Statistics, 2015, Vol. 4, pp. 368-372.

[8]   P. Dejtrakulwong, T. Mukerji, and G. Mavko. Using kernel principal component analysis to interpret seismic signatures of thin shaly-sand reservoirs, SEG Technical Program Expanded Abstracts 2012: pp. 1-5.

[9]   Z. N. Liu, C. Y. Song, H. P Cai, X. M Yao, and G. M. Hu. "Enhanced coherence using principal component analysis." Interpretation, 2017, vol.5, pp. T351-T359.

[10]   G. Phelps, C. Scheidt, and J. Caers . Exploring viable geologic interpretations of gravity models using distance-based global sensitivity analysis and kernel methods, GEOPHYSICS, 2018, vol.83, pp. G79-G92

[11]   P. Dejtrakulwong, T. Mukerji, and G. Mavko. Using kernel principal component analysis to interpret seismic signatures of thin shaly-sand reservoirs, SEG Technical Program Expanded Abstracts 2012: pp. 1-5.

[12]   G. Li, H. F. Li , L. Zhang. Novel Model Using Kernel Function and Local Intensity Information for Noise Image Segmentation, Tsinghua Science and Technology, 2018, vol 23, pp. 303-314.

[13]   Guangzhi Zhang, Guoying Kong, and Jingjing Zheng (2009) Seismic attribute optimization based on kernel principal component analysis. Beijing 2009 International Geophysical Conference and Exposition, Beijing, China, 24–27 April 2009: pp. 57-57.

[14]   A. Pradhan and T. Mukerji. Seismic estimation of reservoir properties with Bayesian evidential analysis. SEG Technical Program Expanded Abstracts 2018: pp. 3166-3170.

[15]   T. Li, R. Wang, Z. Z. Wang, M. Y. Zhao, and L. Li (2018). Prediction of fracture density using genetic algorithm support vector machine based on acoustic logging data, GEOPHYSICS, 2018, vol 83, pp. D49-D60.

[16]   R. Gross, I, Matthews, S. Baker, Appearance-based face recognition and light-fields, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2004, vol.26, pp.449-465.