Reservoir prediction using multi-wave seismic attributes*

Ye Yuan^{1,2} Yang Liu^{1,2,†} Jingyu Zhang^{1,2}
Xiucheng Wei³ and Tiansheng Chen³

¹ State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum,
Beijing 102249, China

² CNPC Key Laboratory of Geophysical Prospecting, China University of Petroleum,
Beijing 102249, China

³ Petroleum Exploration and Production Research Institute, SINOPEC, Beijing 100083, China

Abstract The main problems in seismic attribute technology are the redundancy of data and the uncertainty of attributes, and these problems become much more serious in multi-wave seismic exploration. Data redundancy will increase the burden on interpreters, occupy large computer memory, take much more computing time, conceal the effective information, and especially cause the “curse of dimension”. Uncertainty of attributes will reduce the accuracy of rebuilding the relationship between attributes and geological significance. In order to solve these problems, we study methods of principal component analysis (PCA), independent component analysis (ICA) for attribute optimization and support vector machine (SVM) for reservoir prediction. We propose a flow chart of multi-wave seismic attribute process and further apply it to multi-wave seismic reservoir prediction. The processing results of real seismic data demonstrate that reservoir prediction based on combination of PP- and PS-wave attributes, compared with that based on traditional PP-wave attributes, can improve the prediction accuracy.

Key words: seismic attribute; multi-wave exploration; independent component analysis; support vector machine; reservoir prediction

CLC number: P315.3 **Document code:** A

1 Introduction

Seismic attribute technology originated from bright spot technology in 1960s. The concept of attribute was firstly introduced into seismic exploration in 1970s. But after a short time development, the seismic attribute technology decayed in 1980s then rose again in 1990s (Chopra and Marfurt, 2006). During recent years, as an effective and stable method, this technology has been deeply applied in 3-D, time lapse and multi-wave seismic exploration (Yuan and Liu, 2010).

The seismic attribute technology includes extraction, optimization and prediction from attributes.

There are some classifications for attribute extraction from seismic data, such as Taner classification (Taner et al., 1994), Brown classification (Brown, 1996), Chen classification (Chen and Sidney, 1997) and Liner classification (Liner et al., 2004; Chopra and Marfurt, 2006). Attribute optimization methods include dimension reduction method (Hotelling, 1933; Jutten and Herault, 1991; Schölkopf et al., 1998; Tenenbaum, 1998; Roweis and Saul, 2000) and selection method (Chen, 1998; Yin and Zhou, 2005; Tang et al., 2009). The methods for reservoir prediction from attributes such as artificial neural network (Hopfield, 1982; Kohonen, 1989; Luo and Wang, 1997), non-linear multiple regression (Xu, 2009), Kriging method (Krige, 1951) and support vector machine (Vapnik, 1998; Yue and Yuan, 2005) are commonly used.

* Received 13 June 2011; accepted in revised form
24 July 2011; published 10 August 2011.

† Corresponding author. e-mail: cuply72@yahoo.com.cn

© The Seismological Society of China and Springer-Verlag Berlin
Heidelberg 2011

This paper applies the seismic attribute technology to multi-wave exploration, including methods of independent component analysis (ICA) for optimization and support vector machine (SVM) for prediction. We develop a processing procedure of multi-wave attribute technology and test the method with model data and real seismic data.

2 Optimization for multi-wave seismic attributes

The seismic attribute optimization relies on interpreters' experience or adopts mathematical method to select the most sensitive (or effective, representative) and the fewest attributes combination, which can improve the quality of seismic data interpretation and the accuracy of seismic reservoir prediction.

The relationship among seismic attributes, reservoir lithological characteristics and fluid properties is complicated. The sensitive attributes change with the predict object, working area and reservoir. But some seismic attributes just reflect the variation of noise and may have no relationship with target layer itself. If these attributes are not distinguished, it will cause the confusion in interpretation. For pattern recognition, when the sample size is fixed, too many attributes will deteriorate the classification effect and lower the prediction accuracy (Yin and Zhou, 2005).

There are various seismic attribute optimization methods, which can be classified into selection methods and dimension reduction methods. The selection methods focus on the data's external characteristics, and select the most representative components to reduce the data's dimension on the premise of maintaining the original data. Whereas the dimension reduction methods concentrate on the data's internal characteristics, and use mathematical transformation to explore the internal information of data. Although they usually destroy data's original geological significance, they can eliminate the internal redundant information more effectively than selection methods and the lost geological significance can be rebuilt by prediction. So dimension reduction methods are the most commonly used optimization methods. Principal component analysis (PCA) is the most representative algorithm of dimension reduction methods.

2.1 Optimization method by principal component analysis

Principal component analysis (PCA), also known as K-L transform, Hotelling transform, is a linear di-

mension reduction method. PCA analyzes the internal characteristics of correlation matrix or covariance matrix of original components by whitening operation, which can make each attribute uncorrelated and turn a number of components into a few composite components so called principal components. The principal components are usually the linear combination of original data, which have fewer dimensions, but can reflect most information of the original data (Wu and Yan, 2009).

The main steps of the algorithm are as follows:

1) Assume the original seismic attribute matrix $\mathbf{X}$ is

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T, \quad (1)$$

where $\mathbf{x}_i$ represents an m -dimension sample which has m seismic attributes. Normalize $\mathbf{X}$ to eliminate the differences among different magnitudes.

2) Compute the covariance matrix $\mathbf{C}_x$ of attribute matrix $\mathbf{X}$,

$$\mathbf{C}_x = E[(\mathbf{X} - \overline{\mathbf{X}})(\mathbf{X} - \overline{\mathbf{X}})^T], \quad \overline{\mathbf{X}} = E[\mathbf{X}], \quad (2)$$

where $E[\]$ represents the mathematical expectation operator.

3) Compute the eigenvalue λ_i and eigenvector $\mathbf{a}_i$ of matrix $\mathbf{C}_x$, and reorder the eigenvectors as the corresponding eigenvalues from small to large to construct eigenvectors set $\mathbf{A}$. Select the first m components to rebuild a new set $\mathbf{A}^*$.

4) Compute matrix $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{A}^T \mathbf{X}, \quad \mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n]^T, \quad (3)$$

and select the first m components of $\mathbf{Y}$ to construct a new matrix $\mathbf{Y}^*$. Thus we can compress $\mathbf{Y}$ to a m -dimension matrix $\mathbf{Y}^*$. The attribute matrix $\mathbf{X}^*$ after dimension reduction optimization can be expressed as

$$\mathbf{X}^* = \mathbf{A}^* \mathbf{Y}^*. \quad (4)$$

5) Each component in $\mathbf{X}^*$ is principal component. Its contribution can be computed by

$$\beta_k = \lambda_k \left(\sum_{i=1}^m \lambda_i \right)^{-1}. \quad (5)$$

The optimization effect depends on the sum of selected eigenvalues' contribution, which represents the effective information of these eigenvalues. Figure 1 shows an example of PCA optimization. The input data has 19 dimensions and threshold is 90%. x axis represents each principal component and y axis represents contribution rate. The curve stands for the cumulative contribution rate. From Figure 1, we can see that the cumulative

contribution rate of first five principal components occupies more than 90% of the total eigenvalues, which means PCA can reduce the data dimension from 19 to 5. The five principal components represent more than 90% of the total information.

Figure 1 Histogram and cumulative contribution rate curve of optimization by PCA. Input data has 19 dimensions, threshold is 90%.

PCA has some shortages as follows:

1) PCA is based on the assumption of linear combination of signals. But actually, most signals are nonlinear. So the optimization effect by PCA becomes imperfect for nonlinear signals.

2) Another assumption of PCA is that signals are all Gaussian signals, which means every two principal components x and y should be uncorrelated. That is to say, x and y should satisfy

$$C_{xy} = E\{(x - E\{x\})(y - E\{y\})\} = 0, \quad (6)$$

which equals to

$$E\{xy\} = E\{x\}E\{y\}. \quad (7)$$

When these two signals are Gaussian signals, the signal information are embodied in the mean and variance, and the higher-order cumulant is 0. When these two signals are non-Gaussian signals, the joint probability density function is computed by second-order or high-order statistic. The concept of uncorrelation which is defined in the assumption of PCA can not express the signal's intrinsic information, but can be solved by independent component analysis (ICA) algorithm.

2.2 Optimization method by independent component analysis

Independent component analysis (ICA), originated from blind signal separation (BSS, Jutten and Herault, 1991), is a new signal processing and analysis method developed in recent two decades. The main idea is to further separate the whitening signals by statistical method so as to make them independent and non-Gaussian. This method may achieve two-order and high-order attribute correlation analysis.

As previously mentioned, each principal component in PCA must be uncorrelated, which means every two principal components x and y should satisfy equation (7). While in ICA, every two independent components x and y must satisfy statistics independent, that is

$$p_{x,y}(x,y) = p_x(x)p_y(y) \quad (8)$$

where $p_{x,y}(x,y)$ is the joint probability density function of x and y , $p_x(x)$ and $p_y(y)$ are the marginal probability density function of x and y , respectively.

Substituting cumulative distribution function for probability density function into equation (8), we have

$$\begin{aligned} E\{g(x)h(y)\} &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x)h(y)p_{x,y}(x,y)dx dy = \\ &= \int_{-\infty}^{+\infty} g(x)p_x(x)dx \int_{-\infty}^{+\infty} h(y)p_y(y)dy = \\ &= E\{g(x)\}E\{h(y)\}. \end{aligned} \quad (9)$$

So every two independent components x and y satisfy equation (8) which equals to

$$E\{g(x)h(y)\} = E\{g(x)\}E\{h(y)\}. \quad (10)$$

Comparing equation (7) with (10), we can find that the uncorrelation between two principal components is a special situation only when functions $g(x)$ and $h(y)$ in equation (10) are linear. In other words, equations (7) and (10) are equivalent when all signals are Gaussian. So if two components are independent of each other, they must be uncorrelated; on the contrary, if two components are uncorrelated, they are not necessarily independent. Therefore, ICA method has much more universality than PCA.

Before ICA processing, it is necessary to whiten the data by PCA. This will make components orthogonal to each other, thereby decrease the complexity and reduce the dimension. The commonly used algorithm is fast ICA, the flow chart of which is shown in Figure 2. The main procedure is as follows (Hyvärinen et al., 2001):

- 1) Remove the mean from original data $\mathbf{X}$

$$\mathbf{X} = \mathbf{X} - E[\mathbf{X}]. \quad (11)$$

- 2) Whiten the data $\mathbf{X}$ by PCA.

- 3) Give the initial weight matrix $\mathbf{w}_1$ randomly, and normalize it as

$$\mathbf{w}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|}. \quad (12)$$

- 4) Iteratively compute $\mathbf{w}$

$$\mathbf{w}_{i+1} = E[\mathbf{X}G(\mathbf{w}_i^T \mathbf{X})] - E[g(\mathbf{w}_i^T \mathbf{X})]\mathbf{w}, \quad (13)$$

where $g(x)$ is the derivative of $G(x)$. If original signals are the mixture of super-Gaussian and sub-Gaussian, then

$$G(x) = \frac{1}{a} \log_{10}[\text{ch}(ax)], \quad g(x) = \text{th}(ax). \quad (14a)$$

If original signals are all super-Gaussian, then

$$G(x) = -\frac{1}{a} \exp\left(-\frac{ax^2}{2}\right), \quad g(x) = x \exp\left(-\frac{ax^2}{2}\right). \quad (14b)$$

If original signals are all sub-Gaussian, then

$$G(x) = \frac{x^4}{4}, \quad g(x) = x^3. \quad (14c)$$

- 5) Orthogonalize and normalize $\mathbf{w}_{i+1}$

$$\mathbf{w}_{i+1} = \mathbf{w}_{i+1} - \sum_{j=1}^i \mathbf{w}_{i+1}^T \mathbf{w}_j \mathbf{w}_j, \quad (15)$$

$$\mathbf{w}_{i+1} = \frac{\mathbf{w}_{i+1}}{\|\mathbf{w}_{i+1}\|}. \quad (16)$$

Figure 2 Flow chart of fast ICA.

- 6) Compare $\mathbf{w}_{i+1}$ with $\mathbf{w}_i$. If their absolute values are equal, $\mathbf{w}$ is convergent and output $\mathbf{w}$. If not, go back to step (4) until $\mathbf{w}$ approaches convergence or the iteration number reaches the maximum.

3 Reservoir prediction using multi-wave seismic attributes

The seismic attributes will lose their geological significance after optimization using mathematical transform. So it is impossible to directly use these optimized attributes to predict the reservoir parameters such as porosity, oil saturation. Fortunately, the reservoir prediction method can rebuild the geological significance of these attributes, and interpret reservoir more accurately by combining the optimized attributes with substructure, petrophysics, and reservoir information.

Artificial neural network (ANN) is the most widely used prediction method. It simulates the information transmission pattern of human brain neurons synapses to memorize and analyze problems. Varieties of different algorithms derive from ANN such as support vector machine (SVM), which is a very advanced algorithm.

3.1 Basic principle of support vector machine

Support vector machine (Vapnik, 1998), based on statistical learning theory (SLT), is a generalized ANN for pattern recognition of small-sample nonlinear classification and regression. In reservoir prediction, it is easily over-fitting when conventional approaches are used to predict the reservoir parameters, because the known wells' data only account for a small proportion of the whole area information. SVM is a good solution to small sample prediction problem so that it is suitable for reservoir prediction. It can eliminate local anomaly by kernel function, structure risk and slack variable technologies, thereby get more stable result and higher accuracy when working together with ICA.

The basic principle of SVM can be described as follows. For given linearly inseparable data, SVM constructs an optimum hyperplane to classify them in a high-dimension linear space which is mapped by the kernel function. The optimum hyperplane should have the maximum geometry intervals for each sample and make the generalization error minimum in extrapolation (Yue and Yuan, 2005). As shown in Figure 3, H is the optimum hyperplane to be computed. H_1 and H_2 are the sample planes which have the maximum geometry intervals to H . Solid circles and squares represent two kinds of support vector samples in computing.

Figure 3 Illustration of the basic principle of SVM.

When using the SVM to predict the reservoir parameters of the target layer, we firstly select some seismic attributes and known wells' reservoir parameters of target layer to get a training model by SVM. Then based on the training model to predict the whole area's reservoir parameters of the target layer qualitatively (classification) or quantitatively (regression) by SVM.

3.2 The data separation method of support vector machine

SVM uses linear classifier to find the optimum hyperplane in classification (regression can be transformed into classification). As the example shown in Figure 4, triangle and circle represent the sample points $X(x_1, x_2)$ located in two-dimensional space, real line represents the optimum hyperplane, expressed as $f(X)=0$. Computing $f(X)=0$ is considered as a convex linear quadratic programming problem which can be solved by Lagrange multiplier method, that is

$$Y = \text{sgn}(\langle \mathbf{W}, \mathbf{X} \rangle + b) = \text{sgn} \left(\sum_{i=1}^l \alpha_i y_i \langle \mathbf{X}_i, \mathbf{X} \rangle + b \right), \quad (17)$$

Figure 4 Linear classification problems.

where $\langle \rangle$ represents the inner product operator, $\mathbf{X}_i$ is the corresponding sample called support vector which is higher than two-dimension, $\mathbf{X}$ is the vector set of $\mathbf{X}_i$, y_i is the label value of $\mathbf{X}_i$, α_i is Lagrange multiplier and the majority of α_i are zero, $\mathbf{W}$ and b are quantities need to be computed.

The classification problem mentioned above is linearly separable. Actually, most problems are linearly inseparable and it is impossible to find a straight line to separate samples. For example, Figure 5a shows a line segment with two kinds of samples represented by solid and hollow circle. The data is linearly inseparable in this space. To solve this problem, we can map points $(x, f(x))$ on the line segment by function $\varphi(x) = (x - m)^2$ to points $(x, \varphi(x))$ on a parabola shown in Figure 5b. Now the samples can be separated by the straight line.

Figure 5 Solution to a simple linearly inseparable problem. (a) The linearly inseparable data before mapping; (b) The linearly separable data after mapping.

Actually all linearly inseparable data can be linearly separated after mapped from low dimension space to high dimension space by a particular function φ . But the particular function φ cannot be constructed systematically. Fortunately a kind of function mapping the data to a higher dimension space has been reported as

$$K(x, y) = \langle x, y \rangle. \quad (18)$$

This kind of functions is called kernel function. Researches have shown that the functions satisfying Mercer theory can be used as kernel functions (Vapnik, 1998). Here we list several commonly used kernel functions:

1) Linear kernel function

$$K(X_i, X) = \langle X_i, X \rangle, \quad (19a)$$

2) q -order polynomial kernel function

$$K(X_i, X) = [\langle X_i, X \rangle + 1]^q, \quad (19b)$$

3) Gaussian radial basis kernel function

$$K(X_i, X) = \exp\left(-\frac{\|X_i - X\|^2}{g^2}\right), \quad (19c)$$

and 4) sigmoid kernel function

$$K(X_i, X) = \tanh[\rho \langle X_i, X \rangle + c]. \quad (19d)$$

Different kernel functions have different effects on different problems. Among them, Gaussian radial basis kernel function is the most widely used because of its stable performance and anti-noise ability in all kinds of problems.

Substitute equation (18) into (17) and eliminate the Lagrange multipliers α_i which are 0, we have

$$Y = \text{sgn}\left(\sum_{i=1}^l \alpha_i y_i \langle \mathbf{X}_i, X \rangle + b\right) = \text{sgn}\left(\sum_j \alpha_j y_j K(\mathbf{X}'_j, \mathbf{X}') + b\right), \quad (20)$$

where $\mathbf{X}'_j$ and $\mathbf{X}'$ are the mapped samples which can be linearly separable in high dimension space.

Mapped data can be linearly separated to a great extent but can not be linearly separated perfectly due to the universal applicability of kernel functions. In most cases, a few samples still cannot be separated. These samples are probably noise or wrong data. We use a slack variable ζ and a loss factor C to neglect these outliers as below:

$$\begin{cases} \min\{\sum_j \alpha_j + C \sum_j \zeta_j\} \\ \text{s.t. } y_i(\sum_j \alpha_j y_j K(\mathbf{X}'_j, \mathbf{X}') + b) - 1 + \zeta_j \geq 0 \end{cases} \quad (21)$$

where $\zeta_j \geq 0$, $\alpha_j \geq 0$. If a sample cannot be linearly separated absolutely, its slack variable $\zeta_i > 0$, otherwise $\zeta_i = 0$. Loss factor C determines the importance of the inseparable outliers and how many of them can be retained.

Larger C means that more outliers are important so they are remained. Smaller C means that more outliers have little importance, so they can be abandoned.

The separation procedure of linearly inseparable data can be summarized as follows:

1) Map the original linearly inseparable data by kernel function. As shown in Figure 6a, the data of Figure 4 added with new solid samples make all data linearly inseparable. Figure 6b shows that the data after mapping can be linearly separated except one solid outlier triangle.

2) Adjust the loss factor C to neglect the solid outlier triangle when training by SVM to get the prediction model. Then the remained samples can be linearly separated perfectly.

Figure 6 Linearly inseparable problems. (a) The linearly inseparable data; (b) The data after kernel function mapping.

3.3 Model example

We use a set of model data to verify the predicting effect of SVM. For given 200 samples' porosity ϕ ($0 \leq \phi \leq 0.4$) and water saturation S ($0 \leq S \leq 1$), we compute their P-wave velocity and S-wave velocity by Gassmann's equation. The rock skeleton is sandrock, $K_{\text{sand}} = 40$ GPa, $\rho_{\text{sand}} = 2.65$ g/cm³, $v_{\text{P,sand}} = 2.5$ km/s, $v_{\text{S,sand}} = 1.26$ km/s, and the fluid is filled with mixture of oil and water, $K_{\text{oil}} = 1.15$ GPa, $\rho_{\text{oil}} = 0.88$ g/cm³,

$K_{\text{water}}=2.38 \text{ GPa}$, $\rho_{\text{water}}=1.089 \text{ g/cm}^3$.

Figure 7 shows porosity and water saturation value at every sample point number. Figure 8 shows curves of P-wave velocity and S-wave velocity computed by Gassmann's equation.

Taking P-wave and S-wave velocities of 200 samples as known data, this paper selects 13 samples for training, we predicts porosity and water saturation by feed-forward back propagation (BP) neural network and SVM, respectively.

Figure 7 The curves of porosity and water saturation versus sample point.

Figure 8 The curves of P-wave velocity and S-wave velocity computed by Gassmann's equation.

Figure 9a shows the predicted result of porosity. Black curve represents the original porosity, red curve represents the porosity predicted by SVM, blue curve represents the porosity predicted by BP neural network, and black squares represent the samples for training.

Figure 9b shows the error of porosity. Red curve represents the error of porosity between original data and predicted value by SVM, and blue curve represents the error of porosity between original data and predicted value by BP neural network.

Figure 10a shows the predicted result of water saturation. Black curve represents the original water saturation, red curve represents the water saturation predicted by SVM, blue curve represents the water saturation predicted by BP neural network, and black squares represent the samples for training.

Figure 10b shows the predicted error of water saturation. Red curve represents the error of water saturation between original data and predicted value by SVM, and blue curve represents the error of water saturation between original data and predicted value by BP neural network.

From Figures 9 and 10 we can see that the predicted error of SVM are smaller than BP neural network, especially the anomalies circled in Figures 9a and 10a. Moreover, the linear classifier used in SVM is simple and can get the unique classification result, which means that SVM has no ambiguity. For small-sample prediction, SVM has incomparable advantages over others.

Figure 9 The predicted result of porosity by SVM and feed-forward BP neural network.

Figure 10 The predicted result of water saturation by SVM and feed-forward BP neural network.

4 Application

4.1 The procedure of multi-wave attribute interpretation

PP-wave and PS-wave attributes from multi-wave and multi-component exploration are able to provide much more information and thus improve the accuracy of reservoir prediction. The main steps are shown in Figure 11 and can be summarized as follows:

- 1) Normalize the PP- and PS-wave attributes to the same magnitude.
- 2) Optimize the PP- and PS-wave attributes by PCA and ICA respectively, and then output the selected independent components.
- 3) Train the optimized independent components and the effective oil-bearing thickness of some known wells by SVM to obtain the multi-wave predicting model.
- 4) Input the optimized independent components of the whole area, use SVM to predict the effective oil-bearing thickness of all wells in the whole working area

based on the multi-wave model, and output the multi-wave prediction slice.

4.2 Extraction and optimization of multi-wave attributes

We extract 17 interformational amplitude attributes from PP- and PS-wave data of the same working area. The attributes are listed in Table 1.

Then optimize PP- and PS-wave attributes respectively by PCA with 90% threshold. Figure 12 shows that the original 17 dimension attributes are optimized to four principal components. Figure 13 shows the first four principal components of PP-wave and PS-wave optimized by PCA.

Optimization by ICA is based on the principal components obtained from PCA. The optimized independent components are shown in Figure 14. ICA can maintain the dimension reduction quantity of PCA. For the same or narrower color bar, the red abnormal area of independents components in Figure 14 is more sensitive than that of the principal components in Figure 13, suggesting ICA is more effective than PCA from the mathematical point of view.

Figure 11 The flow chart of combined prediction from multi-wave attributes.

Figure 12 Contribution histogram of each principal component and cumulative contribution rate curve. (a) PP-wave; (b) PS-wave.

Table 1 List of extracted PP- and PS-wave attributes

No.	Attribute name	No.	Attribute name	No.	Attribute name
1	Arithmetic mean	7	Number of zero crosses	13	Integrated absolute amplitude
2	Geometrical mean	8	RMS amplitude	14	Composite absolute amplitude
3	Standard deviation	9	Amplitude slope	15	Ratio of positive to negative vibration
4	Energy value	10	Number of peaks	16	Ratio of adjacent peak amplitudes
5	Maximum value	11	Number of troughs	17	Isopach
6	Minimum value	12	Average vibration path length		

4.3 The result analysis of oil-bearing property prediction

To predict the effective oil-bearing thickness of target layer by SVM, we select 12 wells as training wells from the total available 83 wells. The rest 71 wells are used to verify the accuracy of prediction. First of all, only input PP-wave attributes to perform prediction. Figure 15a is the prediction slice without any optimiza-

tion. Figures 15b and 15c are the prediction slice optimized by PCA and ICA, respectively. The value of color represents the effective thickness for the oil layer, and the red color represents the abnormality caused by oil. Comparison of these figures indicates that the prediction slices without optimization and optimized by PCA are both under-fitting, and the abnormal areas of them are indistinct. On the contrary, the prediction slice

Figure 13 The first four principal components of PP-wave (a) and PS-wave (b) after attribute optimization by PCA.

Figure 13 The first four principal components of PP-wave (a) and PS-wave (b) after attribute optimization by PCA.

Figure 14 The four independent components after optimized from principal components by ICA. (a) PP-wave; (b) PS-wave.

Figure 14 The four independent components after optimized from principal components by ICA. (a) PP-wave; (b) PS-wave.

Figure 15 The slice of PP-wave oil-bearing property prediction by different optimization method. (a) Prediction without optimization; (b) Prediction after PCA optimization; (c) Prediction after ICA optimization.

optimized by ICA contains large amount of information and is much more sensitive to abnormality caused by oil even with a wider color bar. From what has been discussed above, we can draw a conclusion that ICA can keep the optimized effect of PCA, and get a better predicted effect with SVM.

In order to get a higher accuracy, we input PP- and PS-wave attributes optimized by ICA to do oil-bearing property prediction. Figure 16 shows the PP-wave pre-

diction slice which has been shown in Figure 15c and the multi-wave prediction slice. As we can see, the positions of faults are coincident, the PP-wave prediction slice has better continuity, but multi-wave prediction slice has higher resolution. Next, we will enlarge the same local area of two prediction slices, mark the oil-bearing property of training wells and verify well on them, finally we get Figure 17.

Figure 16 The oil-bearing property prediction slice optimized by ICA. (a) PP-wave prediction slice; (b) Multi-wave prediction slice.

Exclude the two training wells in local area, 20 wells among 27 verifying wells in Figure 17a are in accord with the known oil-bearing property. White crosses enveloped by white lines represent the discordant wells with the known oil-bearing property. Figure 17b shows higher accuracy and sensitivity for oil wells. But the slice does not have the same smoothness as that in Figure 17a. Considering the spatial distribution of oil reservoir is continuous, we regard wells located in dense red area as oil wells, such as the four oil wells at the left bottom corner of Figure 17b. Therefore, only one oil well represented by white cross falls far away from the red abnormal area. And three water wells (represented by white circles) enveloped by white lines fail to achieve the anticipated prediction results. The numbers of cor-

rectly predicted wells in 71 verifying wells are shown in Table 2.

The improvement of accuracy mainly benefits from the adding information of PS-wave attributes, which are particularly important in such a small-sample prediction problem. This is the greatest advantage of reservoir prediction from multi-wave seismic attributes.

5 Conclusions

1) The reservoir prediction method from multi-wave seismic attributes is the same as the traditional reservoir prediction method in theory. The former only makes some improvement and optimization based on the latter.

Figure 17 The local enlarged figure of PP-wave and multi-wave prediction slices. Black circles represent wells for training, crosses represent oil wells for verification, and white circles represent water or dry wells for verification. (a) Local enlarged prediction slice of PP-wave; (b) Local enlarged prediction slice of multi-wave.

Table 2 The statistics of prediction result

Method	Number of correct wells (in number of 71 wells)	Accuracy
Prediction with PP-wave	53	74.6%
Prediction with multi-wave	56	78.9%

2) Traditional seismic attribute extraction usually depends on the individual experiences. The extracted attributes represent rough information which has similarities among different attributes. Attribute optimization by PCA or ICA can eliminate the redundancy of original data, screen out the effective information, get more representative attributes and improve the accuracy. The attributes optimized by ICA are more sensitive than by PCA from the view of mathematics.

3) SVM has incomparable advantages over other

pattern recognition methods for small-sample prediction problem such as reservoir prediction from seismic attributes. It works in eliminating the ambiguity, and local anomaly are more stable.

4) Reservoir prediction from multi-wave attributes can use more information effectively. Therefore, compared with PP-wave reservoir prediction, the accuracy of multi-wave reservoir prediction is higher.

5) Some seismic attributes are sensitive to fault and fracture, and the PCA or ICA optimization meth-

ods can eliminate the background interference and highlight the fracture, such as the blue (or red when the color bar is reversed) banded section in the first principal component and independent component in Figures 13 and 14. So this method may be applied in seismology by periodical detection of the change of complex fracture zone based on the earthquake data.

With the development and wide application of multi-wave and multi-component seismic exploration, new method of optimization and prediction needed to be found to process nonlinear data more efficiently. More attention should be paid to various fields such as artificial intelligence, pattern recognition, and data mining, which may promote the integration of seismic attribute extraction, optimization and prediction. In reservoir prediction field, using pre-stack seismic attributes for reservoir prediction is another restricted approach to hydrocarbon detection.

Acknowledgements This study was supported by China Important National Science & Technology Specific Projects (No. 2011ZX05019-008) and National Natural Science Foundation of China (No. 40839901).

References

- Brown A R (1996). Seismic attribute and their classification. *The Leading Edge* **15**: 1 090.
- Chen Q and Sidney S (1997). Seismic attribute technology for reservoir forecasting and monitoring. *The Leading Edge* **16**: 445–456.
- Chen Z D (1998). The method of rough-set decision analysis and its application to pattern recognition of seismic data. *SEG/EAGE/CPS International Geophysical Conference*. Beijing, China, June 22–25, Expanded Abstracts, 471–475.
- Chopra S and Marfurt K J (2006). 75th Anniversary: Seismic attributes — A historical perspective. *Geophysics* **70**: 3SO–28SO.
- Hopfield J J (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* **79**: 2 554–2 558.
- Hotelling H (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **24**: 417–441.
- Hyvärinen A, Karhunen J and Oja E (2001). *Independent Component Analysis*. Wiley-Interscience, New York, 504pp.
- Jutten C and Herault J (1991). Blind separation of sources: An adaptive algorithm based on a neuromimetic architecture. *Signal Processing* **24**: 1–10.
- Kohonen T (1989). *Self-organization and Associative Memory*. Springer, Berlin, 312pp.
- Krige D G (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa* **52**: 119–139.
- Liner C, Li C F and Gersztenkorn A (2004). Spice: A new general seismic attribute. *72nd Annual International Meeting, SEG*. Lake City, USA, Oct. 6–11, Expanded Abstracts, 433–436.
- Luo L M and Wang Y C (1997). Improvement of self-organizing mapping neural network and the application in reservoir prediction. *Oil Geophysical Prospecting* **32**: 237–245 (in Chinese with English abstract).
- Roweis S T and Saul L K (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**: 2 323–2 326.
- Schölkopf B, Smola A J and Müller K R (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation* **10**: 1 299–1 319.
- Taner M, Schuelke J S and O'Doherty R (1994). Seismic attributes revisited. *64th Annual International Meeting, SEG*. Los Angeles, USA, Oct. 23, Expanded Abstracts, 1 104–1 106.
- Tang Y H, Zhang X J and Gao J H (2009). Method of oil/gas prediction based on optimization of seismic attributes and support vector machine. *Oil Geophysical Prospecting* **44**: 75–80 (in Chinese with English abstract).
- Tenenbaum J B (1998). Mapping a manifold of perceptual observations. In: *Advances in Neural Information Processing Systems*. MIT Press, Cambridge, **10**: 682–688.
- Vapnik V N (1998). *Statistical Learning Theory*. Wiley-Interscience, New York, 736pp.
- Wu X T and Yan D L (2009). Analysis and research on method of data dimensionality reduction. *Application Research of Computers* **26**: 2 832–2 835 (in Chinese with English abstract).
- Xu Q (2009). *The Research on Non-linear Regression Analysis Methods*. [MS Dissertation]. Hefei University of Technology, Hefei, 44pp (in Chinese with English abstract).
- Yin X Y and Zhou J Y (2005). Summary of optimum methods of seismic attributes. *Oil Geophysical Prospecting* **40**: 482–489 (in Chinese with English abstract).
- Yuan Y and Liu Y (2010). New progress on seismic attribute optimizing and predicting. *Progress in Exploration Geophysics* **33**: 229–238 (in Chinese with English abstract).
- Yue Y X and Yuan Q S (2005). Application of SVM method in reservoir prediction. *Geophysical Prospecting for Petroleum* **44**: 388–392 (in Chinese with English abstract).