Application research of attribute fusion technology based on principal component analysis in fracture identification

Rui Wang¹, Qiong Li*  
¹ College of Geophysics, Chengdu University of Technology, Chengdu, Sichuan, China  
*Corresponding author’s e-mail: liqiong@cdut.cn

Abstract. The seismic interpretation and analysis of a single attribute has multiple solutions and limitations, so the article proposes the technical research on the feature identification of fault structure based on the attribute fusion of principal component analysis (PCA). The results indicate that, compared with a single attribute, the integrated seismic attributes obtained by the fusion of the principal component analysis (PCA) method can more clearly reflect the development direction and boundary range of the fault, and the small fractures distributed around it can also be more obvious. The characterization proves that this technology has great application potential in fracture identification.

1. Introduction
Seismic attributes can be widely used in various aspects such as geological structure interpretation, reservoir prediction, and analysis of geological anomalies. However, seismic attributes have a large category with many types, and there are extremely complicated relationships between each attribute and the geological objects studied. As a result, a single seismic attribute must have multiple solutions and limitations for seismic interpretation and analysis, and comprehensive analysis of multiple seismic attributes is required[1].

Seismic attribute fusion technology is based on the joint analysis of seismic attributes and has been continuously developed. Reservoir prediction and analysis technology has been continuously developed. Principal component analysis performs dimensionality reduction processing on data before interpretation to achieve seismic attribute fusion and reduce the calculation of comprehensive analysis. So it has been widely used in image fusion[2], data compression[3], feature extraction[4] and other fields.

2. Method
2.1. Principal component analysis, PCA
Principal component analysis (PCA) is the use of matrix dimensionality reduction ideas to transform multiple attributes into several comprehensive attributes (ie principal components), and use comprehensive attributes to replace original attributes for data interpretation and analysis. Each comprehensive attribute obtained by principal component analysis (PCA) can reflect most of the information of each original attribute, and the information contained in each comprehensive attribute does not repeat each other[5].

Standardize the original data matrix X of the P-dimensional random vector:

\[ x_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (i=1,2,\ldots,n; j=1,2,\ldots,p) \]  

(1)
Where $\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}$, \( \text{var}(x_j) = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \) (j = 1,2,\ldots, p)

Find the covariance of the original data matrix X after normalization processing; the obtained covariance matrix is the correlation matrix of matrix X, and the correlation coefficient matrix R is

$$R = \begin{bmatrix}
    r_{11} & r_{12} & \cdots & r_{1p} \\
    r_{21} & r_{22} & \cdots & r_{2p} \\
    \vdots & \vdots & \ddots & \vdots \\
    r_{p1} & r_{p2} & \cdots & r_{pp}
\end{bmatrix} \quad (2)$$

Where $r_{ij} = \frac{1}{n-1} \sum_{t=1}^{n} x_{ti}x_{tj}$ (i, j = 1,2,\ldots, p)

By using the method of matrix diagonalization, the eigenvalues ($\lambda_1$, $\lambda_2$, \ldots $\lambda_p$) of the correlation coefficient matrix R corresponding to the original data matrix X and the eigenvalues corresponding to the eigenvalues $\alpha_{i1}$, $\alpha_{i2}$, \ldots, $\alpha_{ip}$ can be obtained, where i=1,2,\ldots,p.

Sort the eigenvalues ($\lambda_1$, $\lambda_2$, \ldots $\lambda_p$) of the correlation coefficient matrix R from large to small, which indicates that the amount of information contained in the principal components formed is from large to small. In practical applications, the top n principal components whose cumulative contribution of each principal component exceeds 85% are generally selected.

$$G(n) = \frac{\sum_{k=1}^{n} \lambda_k}{\sum_{k=1}^{p} \lambda_k} \quad (n < p) \quad (3)$$

Calculate the scores of the sample on n principal components:

$$F_i = \alpha_{i1}X_1 + \alpha_{i2}X_2 + \cdots + \alpha_{ip}X_p \quad (i = 1,2,\ldots,n) \quad (4)$$

Generally, the coefficient values of each variable are different, and the principal component with a large coefficient value indicates that it carries a large amount of effective information in the original data set. When the coefficients of each principal component are close, it can be considered that this principal component is a comprehensive reflection of these variables. When analyzing and interpreting the geological significance represented by the principal components, it should be combined with the actual geological structure, reservoir information, and logging data to make a reasonable interpretation.

2.2. Attribute calculation and extraction

2.2.1. Dip and azimuth attributes

The improved GST algorithm which is based on the gradient structure tensor can extract the inclination and azimuth attributes from the seismic data volume. For a three-dimensional data volume, by calculating the directional derivative of each point and establishing the gradient structure tensor, the gradient structure tensor matrix $T_G$ can be obtained, which can be expressed as:

$$T_G = \begin{bmatrix}
    G * g_x g_x G * g_x g_y G * g_x g_z \\
    G * g_y g_x G * g_y g_y G * g_y g_z \\
    G * g_z g_x G * g_z g_y G * g_z g_z
\end{bmatrix} \quad (5)$$

Decompose the eigenvalue of formula (5):

$$T_G v = \lambda v \quad (6)$$

In formula (6), $\lambda$ is the eigenvalue of the matrix $T_G$, and $v$ is the eigenvector of the matrix $T_G$ belonging to the eigenvalue $\lambda$. Among them, the eigenvalues $\lambda_1$, $\lambda_2$, and $\lambda_3$ correspond to the eigenvectors $v_1$, $v_2$, $v_3$, respectively. Through the feature vectors $v_1$, $v_2$, $v_3$, the inclination and azimuth attributes can be constructed as follows:

Inclination attributes:

$$\text{Dip} = \arctan\left(\frac{v_1(x)}{\sqrt{v_1(x)^2 + v_1(y)^2}}\right) \quad (7)$$

Azimuth attributes:

$$\text{Azimuth} = \arctan\left(\frac{v_1(y)}{v_1(x)}\right) \quad (8)$$
2.2.2. Configuration tensor attributes

According to the calculation of inclination angle and azimuth angle, the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ and eigenvectors $v_1, v_2, v_3$ obtained by eigenvalue decomposition contain rich geological structure information.

According to the relationship between the eigenvalues and the three-dimensional image structure, the following expressions describing the fault can be obtained\cite{6}:

Chaos attributes:

\[
C_{\text{fault}} = \frac{2\lambda_2(\lambda_2 - \lambda_3)}{(\lambda_1 + \lambda_2)(\lambda_2 + \lambda_3)} \quad (9)
\]

In addition to identifying cracks on slices, this attribute has also achieved good results in identifying cracks on seismic profiles, and has a more obvious role in describing river edges and weak energy cluttered reflections.

Lamda transverse gradient attributes:

\[
C_{\text{lamda}} = \lambda_2 + \lambda_3 \quad (10)
\]

The horizontal gradient attribute has good noise resistance and has a good characterization effect on the edge of the river.

2.2.3. The third-generation coherent algorithm

The calculation of the coherence value of the C3 algorithm uses the covariance matrix $C$ in the C2 algorithm, and $C$ is expressed as follows:

\[
C = RTR = \sum_{m=n-M}^{n+M} \begin{bmatrix}
R_{1m}R_{1m} & R_{1m}R_{2m} & \cdots & R_{1m}R_{jm} \\
R_{2m}R_{1m} & R_{2m}R_{2m} & \cdots & R_{2m}R_{jm} \\
\vdots & \vdots & \ddots & \vdots \\
R_{jm}R_{1m} & R_{jm}R_{2m} & \cdots & R_{jm}R_{jm}
\end{bmatrix} \quad (11)
\]

In formula (11), $T$ represents the transpose of the matrix.

Perform eigenvalue decomposition on the covariance matrix $C$ to obtain the eigenvectors and the eigenvalues corresponding to each eigenvector. The maximum eigenvalue obtained is stored in 95% of the information of the covariance matrix $C$. Selecting the largest eigenvalue as the estimation of the coherence of the feature structure, the third-generation coherence algorithm expression can be obtained as

\[
C_3 = \max(C(p, q)) = \frac{\lambda_{\text{max}}}{\sum_{m=1}^{M}} \quad (12)
\]

In formula (12), $(p, q)$ is the apparent inclination angle, $\lambda_{\text{max}}$ is the maximum eigenvalue of the covariance matrix $C$, and $\lambda_m$ is the mth eigenvalue of the covariance matrix $C$.

Compared with the first and second-generation coherent algorithms, the third-generation coherent algorithm has the advantages of strong anti-interference ability and high signal-to-noise ratio of the output result, and is more suitable for the identification of faults and cracks in seismic data volumes with low signal-to-noise ratio.

3. Results & Discussion

3.1. Background of the work area

The main research is the Jurassic Sangonghe Formation strata in the Moxizhuang area. This strata has a large burial depth, mainly fluvial facies deposits, with a large number of sand bodies. The physical properties of the sand body are not uniformly distributed, and the spatial distribution of the sand body varies greatly in the vertical and horizontal directions. It is a multi-phase resetting deposit, so it is difficult to accurately obtain its spatial distribution. The structure of the entire stratum is characterized by a wide and gentle paleo-uplift structure, and low-amplitude anticlines and nose-like structures are locally developed on the background of the uplift structure. At the same time, normal faults with an east-west or north-east direction are mainly developed.

The 3D seismic data with a sampling interval of 2ms in the study area is shown in Figure 1 as the seismic profile of main survey line No. 210. The target horizon is marked with a red line in the figure,
which is 2ms above the s21 horizon. In the entire section, there is a f1 fault. However, it is difficult to observe the internal micro-faults on the section. It is necessary to study the characteristics of the fault structure of the target horizon and detect the development of the fractures.

3.2. Single attribute analysis

3.2.1. Azimuth and dip attributes

Figure 2 shows a plan view of the azimuth and dip attributes of the target horizon.

The azimuth attribute can generally describe the fault development system of the study area on a large scale, and describe the strike and trend characteristics of the strata. As shown in Figure 2(a), in the area marked by the red curve, the azimuth value changes drastically and the drop is obvious. This shows that, on a large scale, the area marked by the red curve is a concentrated area of stratum faults and fractures. The dip angle attribute reflects the plane distribution of the fracture development zone. As shown in Figure 2(b), in the area marked by the red ellipse, the dip angle changes drastically and is distributed with fine line bands, indicating that this is the main fault development zone. There are some tiny cracks around it.

3.2.2. Configuration tensor attributes

Because the reflected energy of the target horizon is weak, the chaotic attribute in the configuration tensor attribute has a significant role in describing the edge of the channel and the weak energy.
cluttered reflection. Therefore, the chaotic attribute is extracted to reflect the fracture development of
the target horizon, as shown in Figure 3.

In the area marked by the red ellipse, the distribution of fault factors presents clumpy, short and
curved linear features. Connecting the characteristics of these fault factors can clearly identify the
boundary range of the main fracture development area and depict the fractures. The developed edge
structure. The more concentrated black area indicates the high degree of fracture development; while
the wide-spread white area indicates that the fracture degree is poor and the continuity of the
formation is better.

3.2.3. Coherence attribute

The seismic coherence attribute characterizes the continuity of the seismic waveform event axis. In
geological regions with large lithological changes and developed fractures, the seismic waveform
event axis is discontinuous and the coherence value is small; while in the formation lithology with
small changes and no cracks, the seismic waveform is continuous on the phase axis and the coherence
value is high. Therefore, extracting the coherent attributes of the target horizon to identify faults and
fractures is a great effective attribute analysis method.

The third-generation coherence algorithm based on feature structure decomposition is used to
extract the seismic coherence attributes of the target horizon, as shown in Figure 4. In the area marked
by the red ellipse, the coherence value is small and distributed in strips, indicating that the continuity
of the formation is poor, and it can be inferred as a concentrated distribution area of fractures.

3.3. Fusion attribute analysis

Using principal component analysis (PCA), the azimuth angle, inclination angle attribute, chaos
attribute, and coherence attribute are used for attribute fusion. And comprehensive attributes including
the first principal component, the second principal component, and the third principal component are
obtained, as shown in Figure 5.

The first principal component (see Figure 5(a)) contains most of the information of each seismic
attribute participating in the attribute fusion. On the entire slice along the layer, the development
direction and boundary range of the six main faults (As shown in Figure 5, the yellow ellipse marked
area) has reached a very clear description, and some small cracks existing on the edge have also been
described more clearly. The overall effect is more prominent than the description of the fault
caracteristics reflected by each single attribute. The purpose of comprehensive analysis of fusion
attributes. The second principal component (see Figure 5(b)) contains part of the seismic attribute
information. The development direction and boundary range of the six main faults that exist are
described clearly, but it contains less seismic attribute information. The depiction of some tiny cracks
is not clear enough, and it is not clearly distinguished from some continuity formations. Due to the
information of the new comprehensive attributes (principal components) obtained by the principal
component analysis (PCA) is not related to each other, the third principal component (see Figure 5(c)) only contains very little seismic attribute information. The description of the 6 faults is not clear, and only the local development and boundary range of the A, B, and E faults can be described clearly.

4. Conclusions
The use of principal component analysis (PCA) to fuse individual attributes has achieved very good results, and the integrated attribute analysis obtained by fusion can more clearly reflect the development direction, direction of faults, boundary area, and the distribution of tiny cracks than single attribute analysis.

The use of comprehensive attributes for analysis can effectively avoid data redundancy, reduce the computational workload in the interpretation and analysis process, and improve the accuracy of structural interpretation and reservoir prediction, which is extremely important for the interpretation and analysis of massive seismic data.
Acknowledgements
This research was funded by CNPC Chuanqing Drilling Engineering Company Drilling and Production Engineering Technology Research Institute "Drilling Fault Complex Processing System and UAV Image Intelligent Identification Development", National Science and Technology Major Project (2016ZX05026001-004), National Natural Science Foundation of China (41274129), in 2018, the central government supported the local co-construction of disciplines (80000-18Z0140504) and in 2019, geophysics, the joint construction and development of universities (double first-class central government, 80,000-19Z0204).

References
[1] Cheng Y. (2016) Application of multi-attribute fusion technology in 3D seismic data interpretation [J]. China Coal Geology, 28(10):67-70. (In Chinese)
[2] Ding F, Nian Y J, Wang Z G, et al. (2010) Application research of seismic multi-attribute RGBA color fusion technology [J]. Geophysical Prospecting for Petroleum, 49(03): 248-252. (In Chinese)
[3] Chen P. (2014) Research on principal component analysis and its application in feature extraction [D]. Xi'an: Shanxi Normal University. (In Chinese)
[4] Zhao X, Wang W J, Zeng Y Y, et al. (2015) Improved modular PCA new algorithm for face recognition [J]. Computer Engineering and Applications, 51(02):161-164. (In Chinese)
[5] Fan X L, Feng H H, Yuan M. (2013) Principal component analysis feature selection algorithm based on mutual information [J]. Control and Decision, 28(06): 915-919. (In Chinese)
[6] Zhang J H, Dong M, Zhou Z X, et al. (2007) Research and application of coherence cube method based on GST [J]. Natural Gas Industry, (S1):381-383. (In Chinese)