IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

High-resolution velocity analysis and program design based on Bootstrap algorithm

To cite this article: Dechao Cai *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **227** 042029

High-resolution velocity analysis and program design based on Bootstrap algorithm

Dechao Cai[1,3], Guifang Fa[1], Jia Li[1], Xing Xing[2] and Yanjing Yi[1]

[1] Research Institute of Petroleum Exploration & Development, PetroChina, Beijing, China;
[2] Research Institute of Petroleum Exploration & Development, Dagang Oil Field, Tianjin, China.

[3] Email: caidechao@petrochina.com.cn

Abstract. Seismic wave velocity is applied in all processes of seismic data processing and interpretation, and it is one of the most important parameters in seismic exploration. How to obtain high-resolution stack **velocity spectrum** [1], and then understand accurate velocity distribution of underground media is the key to seismic data processing. Conventional velocity analysis is based on multi-track cross-correlation within a time window defined along a hyperbola on a CDP gather, so it reflects the changes in signal stack energy within the window. The Bootstrap method is derived from mathematical statistics and is a mathematical method based on matrix transformation, which can accurately estimate the statistical distribution of given data. Based on conventional velocity analysis, this paper introduces how to use the **Bootstrap** method [2] to greatly improve the resolution of stack velocity spectrum, especially in processing data of shallow layers, large offsets and thin layers, based on old data and without additional investment. The application has provided high-resolution results and can effectively compensate many shortcomings of conventional methods. Based on theoretical derivation and process design, this paper presents the software designing process based on the data structure of an orthogonal list, then verifies the results using model and actual data, and finally compares the results with those from popular commercial **software** to verify the stability of the program and the reliability of the method.

1. Introduction

Seismic wave velocity is involved in all processes from seismic acquisition, processing to interpretation and it is one of the most important parameters in seismic exploration. Conventional velocity analysis is based on the multi-track cross-correlation in the time window defined along a hyperbola on a CDP gather. The analysis reflects the changes in signal stack energy in the window, but does not consider the influences of noises associated with similar or interfering events, residual static correction, non-hyperbolic move-out, etc. and non-random noises. In addition, conventional velocity analysis can't take full use of seismic data, so the resolution is greatly limited. Especially when analyzing the velocity of shallow seismic data which is characterized by quick changes in the lateral velocity of seismic wave, limited CDP [3] gathers and low SNR, how to use a mathematical method to maximize the accuracy of velocity analysis based on old data and without additional efforts is an important direction in researching seismic exploration methods.

The essence of the Bootstrap velocity analysis algorithm based on matrix calculation is to increase the number of logical samples by a mathematical means, expand the sample space, greatly increase the actual number of stacks and improve the resolution of seismic data. It is a very meaningful attempt in studying the processing of shallow seismic data.

2. Velocity analysis algorithm and process design

2.1. Conventional velocity analysis

Seismic records are multi-track records [1]. The propagation velocity of seismic wave, an important parameter, is hidden in the normal move-out of multi-track signals. When a underground medium is horizontally layered, the normal move-out of reflected wave Δt_i is a function of offset x_i , echo time t_0, and RMS velocity v_σ.

$$\Delta t_i = \Delta t_i(x_i, t_0, v_\sigma) = \sqrt{t_0^2 + \frac{x_i^2}{v_\sigma^2}} - t_0 \qquad (i = 1,\ 2,\ N. \tag{1}$$

If reflected signals can accurately picked up from seismic records and the normal move-out can be obtained, the interval velocity will be calculated. This is the first step of velocity analysis. The concept of velocity spectrum is modeled after the concept of frequency spectrum, which is used to represent how seismic wave energy changes with velocity. At a given echo time t_0, the reflected wave time-distance curve is calculated (a hyperbola) at a certain velocity step. Based on this, the values of the hyperbola are defined and stacked on CSP or CMP gathers to get stacked amplitude. If the curve corresponding to a velocity point coincides with the events of the reflected wave at t_0, the stack will be done in the same phase, and the stacked amplitude will be extremely large. Therefore, in the calculated stacked amplitude values, the velocity corresponding to the maximum amplitude value is deemed as the velocity at t_0. This is the calculation principles of stack velocity spectrum.

2.2. Bootstrap method

The initial application of the Bootstrap method was to statistically estimate the accurate distribution of given data. The essence of this method is to randomly re-sample actual data and obtain multiple data sets. Based on the individual errors of these data sets, the standard error can be calculated to find the optimal data. The related technique of the Bootstrap velocity analysis is derived from the eigenvalue of data co-variance. Using the Bootstrap method, and according to relevant standard errors, interval velocities will be accurately picked up [1].

Principle of conventional velocity analysis:

Under the concept of RMS velocity [4], the equation of P-P reflected wave time-distance curve in a CMP gather is expressed as

$$t_p = \sqrt{t_0^2 + \frac{4h^2}{v_p^2}} \tag{2}$$

Where, t_p is P wave travel time; t_0 is P-P wave two-way travel time at zero offset; h is a half offset; v_p is the P wave velocity at the same reflection depth.

Assuming a velocity v, using the time-distance curve expressed by equation (2-2), a data time window can be obtained, which is centered on intercept time t_0 and velocity v, and which length is 2m+1 expressed by equation (2-3).

$$Y(t_0, v) = (X_1, X_2, ..., X_i, ..., X_N) \tag{3}$$

The spatial section is regarded as the column 2D matrix [1], $X_i = (x_{i,t_0-m\Delta t}, ..., x_{i,t_0-m\Delta t}, ..., x_{i,t_0+m\Delta t})^T$ where i=1,2,...,N; N is the total number of columns, namely the folds of a CDP; $x_{i,t_0-m\Delta t}$ represents the amplitude of the i_{th} seismic trace at the time sample $t_0+k\Delta t$.

Then the co-variance matrix is

$$R(t_0,v)=\frac{1}{2m+1}X(t_0,v)X^T(t_0,v)=(R_1,R_2,...,R_i,...,R_N)^T \tag{4}$$

Where $R_i=(r_{i,t_0-m\Delta t},...,r_{i,t_0-m\Delta t},...,r_{i,t_0+m\Delta t})^T$,(i=1,2,...,N;j=-m,-m+1,...,m).

After randomly sampling the data in R_i,

$C_{i,t_0+k\Delta t}(t_0,v)=r_{i,t_0+k\Delta t}$, $(i=1,2,...,B)$

Where k is a random natural number with a range of [-m,m]. There are B groups expressed by matrix:

$$Q(t_0,v)=(C_1,C_2,...,C_i,...,C_B)^T \tag{5}$$

where $C_i=(c_{i,t_0-m\Delta t},...,c_{i,t_0-m\Delta t},...,c_{i,t_0+m\Delta t})^T$,(i=1,2,...,B;j=-m,-m+1,...,m).

Using the data, B velocities can be estimated: $v^*(t_0)_1,v^*(t_0)_2,...v^*(t_0)_B$, and consequently average velocity $v(t_0)_{av}$ and velocity error $\sigma(t_0)$.

$$v(t_0)_{av}=\frac{1}{B}\sum_{i=1}^{B}v_i(t_0) \tag{6}$$

$$\sigma(t_0)=\sqrt{\frac{1}{B-1}\sum_{i=1}^{B}[v_i^*(t_0)-v_{av}]^2} \tag{7}$$

The velocity from the Bootstrap method can be replaced by average energy $C(t_0)_{av}$.

$$C(t_0)_{av}=\frac{1}{B}\sum_{i=1}^{B}C_i(t_0,v) \tag{8}$$

The velocity analysis method for P wave is summarized as follows:

(1) Given CDP data, the stack velocity of P wave v and sampled groups B.

(2) Use different apparent velocities to hyperbolically scan the traces in the CDP gather from small to large velocity increments [5].

(3) When using a fixed velocity to scan the traces on a CDP, calculate the zero-offset travel time t_0, of the P wave on any point on any trace using equation (1).

(4) After obtaining the amplitude corresponding to t_0, use time window 2m+1 to randomly sample and set up B groups of data $Q(t_0,v)$.

(5) Use equations (5) and (6) to find the standard error $\sigma(t_0)$ for each row (totally N) of the matrix $Q(t_0,v)$.

(6) Among N standard errors, identify minimum error $\sigma(t_0)_{\min}$, then use $v^*(t_0)_{av}$ corresponding to the minimum standard error to replace the original value.

(7) Repeat steps (4) to (6).

(8) Repeat steps (3) and (7).

(9) Repeat steps (2) and (8).

(10) Display the energy values UC_{SC} corresponding to stack velocity v and t_0 in the form of contours or a curve, and the P wave velocity spectrum will be obtained.

2.3. Principle of velocity analysis updated by Bootstrap method

The Bootstrap method was established by Efron in the late 1970s. With the significant advances in modern computers, the method is being used more and more in practice. Assuming the overall distribution F unknown, but there is already a data sample with a capacity of N. From this sample, a sample with a capacity of N is taken by sampling with replacement[5]. The new sample is called a Bootstrap sample [6]. After separately and successively extracting many Bootstrap samples from the original sample, use these samples to statistically infer the F. This method is called non-parameter Bootstrap method, or Bootstrap method [1].

In the following, according to the conventional velocity analysis method, the feasibility of actual data, and the principle of the Bootstrap method in mathematical statistics, the detailed derivation from the principle to the specific implementation is given for the entire process based on the Bootstrap velocity analysis.

Under the concept of RMS velocity, the equation of the P-P reflected wave time-distance curve in a CMP gather is expressed by equation (2-9) [7].

$$t_p = \sqrt{{t_0}^2 + \frac{4h^2}{{v_p}^2}} \tag{9}$$

where t_p is P wave travel time; t_0 is the P-P wave two-way travel time at zero offset (converted waves not considered); h is a half offset; v_p is the P wave velocity at the same reflection depth.

Assuming a test velocity v, using the time-distance curve expressed by equation (1), a data time window can be obtained, which is centered on the intercept time t_0 and velocity v, and which length is $2m+1$.

$$Y(t_0, v) = (X_1, X_2, ..., X_i, ..., X_N) \tag{10}$$

Similarly, the spatial section is treated as a large 2D matrix.

In the above formula, $X_i = (x_{i,t_0-m\Delta t}, ..., x_{i,t_0-m\Delta t}, ..., x_{i,t_0+m\Delta t})^T$ is the column vector; N represents the total number of columns, that is, the folds of a CDP; $x_{i,t_0+m\Delta t}$ represents the amplitude of the i_{th} seismic trace at the time sample $t_0 + m\Delta t$.

A specified velocity is used to correct the data of any trace, or wave-forms, or more specifically, a sequence of wave-forms of column vectors arranged from top to bottom, using the NMO correction formula. After the NMO correction, the interval velocity should be normalized because there are multiple folds, in other words, there are many reflection wave-forms from the same reflector. According to a conventional stacking method, all sub-amplitudes of the same reflector should be stacked and averaged to obtain more reliable post-stack amplitude of the reflector. However, when the sub-amplitudes from the same reflector are stacked, if they are understood as a sample sequence of the same random variable, which satisfies distribution F , the F can be statistically inferred based on these samples, and the non-parametric Bootstrap idea in probability theory and mathematical statistics. Finally the true amplitude value with the highest confidence, of the reflector will be found [1].

The velocity analysis method for P wave is as follows [6]:

(1) Given CDP data, the range of p wave stack velocity v value, and sampled groups B.

(2) Use different apparent velocities to hyperbolically scan the traces in the CDP gather from small to large velocity increments.

(3) When using a fixed velocity to scan the traces on the CDP, calculate the zero-offset travel time t_0, of the P wave on any point on any trace using equation (1).

(4) Use the t_0 to determine the NMO $\Delta t = t - t_0$ of the amplitude value. Then determine the x_i of the Y_i in the $Y_{CDP} = (Y_1, Y_2, ..., Y_i, ..., Y_N)^T$ which expresses the reflector after NMO correction.

(5) In respect of the row vector $Y_i = (x_1, x_2, ..., x_N)$ corresponding to the sampling point at a certain depth[3], the Bootstrap method is applied to estimate the reflection amplitude of the reflector at this depth.

(6) Repeat steps (4) and (5)

(7) Repeat steps (3) to (6).

(8) Repeat steps (2) and (7).

(10) Display the energy values UC_{SC} corresponding to stack velocity v and t_0 in the form of contours or a curve, and the P wave velocity spectrum will be obtained.

Step (5) includes the implementation process of the Bootstrap method [2] [3]:

(a) A sample $\overrightarrow{X^*} = (x_1^*, x_2^*, ..., x_N^*)$ with a capacity of n can be obtained the original sample $\vec{x} = (x_1, x_2, ..., x_n)$ by sampling with replacement. The new sample is called a Bootstrap sample.

(b) B(B ≥ 1000) Bootstrap samples with a capacity of n each is estimated separately and successfully. Each sample is a vector $\overrightarrow{x^*} = (x_1^{*i}, x_2^{*i}, ..., x_N^{*i})$,i=1,2,...,B. To the i_{th} Bootstrap sample, calculate the corresponding $\hat{\theta}_i^* = \hat{\theta}(x_1^{*i}, x_2^{*i}, ..., x_n^{*i})$,i=1,2,...,B ($\hat{\theta}_i^*$ is the i_{th} estimate of θ).

2.4. Algorithm flowchart

According to the process derivation and step analysis of the Bootstrap-based velocity analysis method II, considering with the characteristics of the programming platform and language used, the flow chart used in the programming process is shown in Figure 1:

3. Program design

Since this work depends on a specific project, detailed program implementation won't be described, instead only a few core technologies and some explanations will be emphasized here[4].The program is based on Windows Visual Studio, and original data are SEG-Y data. For CDP gathers, it is recommended to store and read data through an orthogonal list. The data structure is as follows in Figure 2:

In the process of Bootstrap-based matrix calculation, the mathematical logic is clear and the calculation is efficient. Using Seismic Unix, a model composed of 5 horizontal reflectors was designed. The acquisition parameters include 140 source points, 120 traces per point, 20m trace spacing, 2ms sample interval, record length, and maximum offset 3000m. The parameters such as are shown in Table 1 lists the interval thickness and velocity of the model.[1]

4. Theoretical model experiment and application results

4.1. Theoretical model experiment

Figure 3 shows a CDP gather based on the theoretical model. There are a total of 120 traces. In the model, layers 2 to 4 are thin [8].

Based on the model, two velocity spectra were obtained using conventional and Bootstrap velocity methods, respectively (Figure 4).

On the velocity spectrum from the conventional method, the energy masses of the first and fifth layers are relatively large, but not well concentrated, so the velocity accuracy is low[4]; the energy of the first layer is weak; the energy masses of the middle three thin layers are all stuck together, and it is difficult to distinguish[5]. On the velocity spectrum from the Bootstrap method [1] (Figure 5), the energy mass of any layer is well concentrated. Compared with the conventional method, the Bootstrap method provided strong spectral energy [3] [7] (as shown by the energy curve on the right side), so the velocity accuracy is guaranteed [8]. Especially for the middle three thin layers, the energy masses from the conventional method cannot be distinguished, and but those from the Bootstrap method can, and each energy mass is strong [5].

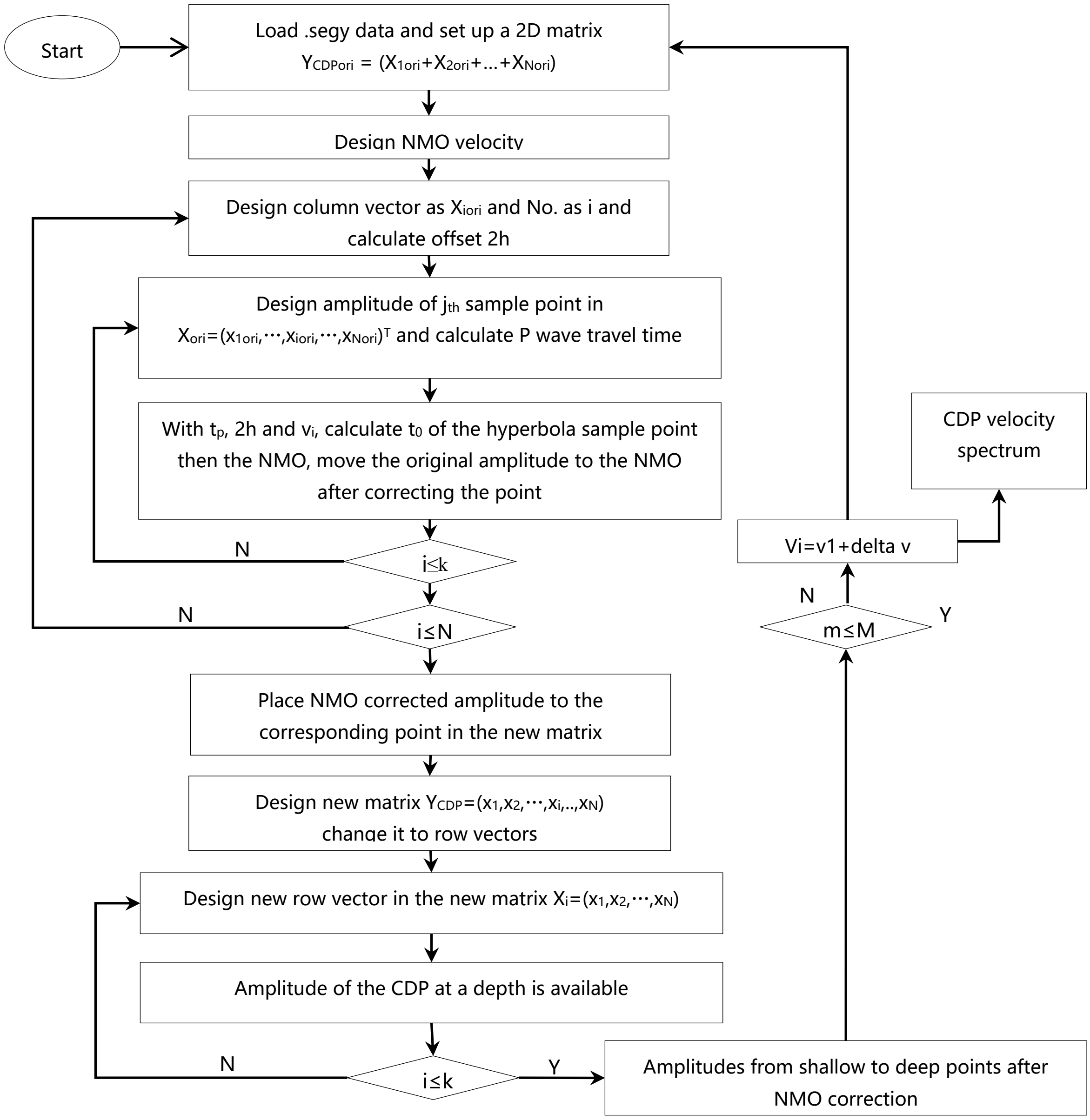

Figure 1. Bootstrap algorithm flowchart.

Row No.	Column No.	Amplitude	Go to next row	Go to next column
row	Column	Value	next row	next column

Figure 2. The internal structure on a node of the orthogonal list.

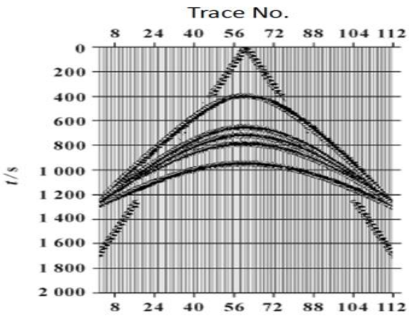

Figure 3. A modeled CDP gather.

Table 1. A theoretical model.

Interval	Thickness (m)	Velocity (m/s)
1	400	2000
2	310	2600
3	90	3000
4	110	3400
5	290	3800

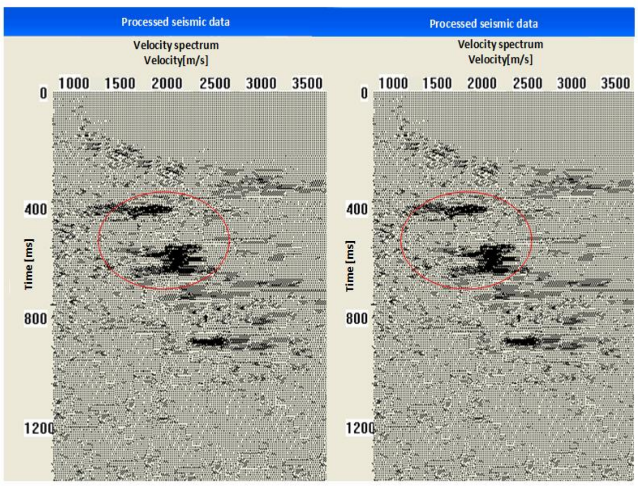

Figure 4. Velocity spectra based on the same experimental model: conventional method (left), and Bootstrap method (right).

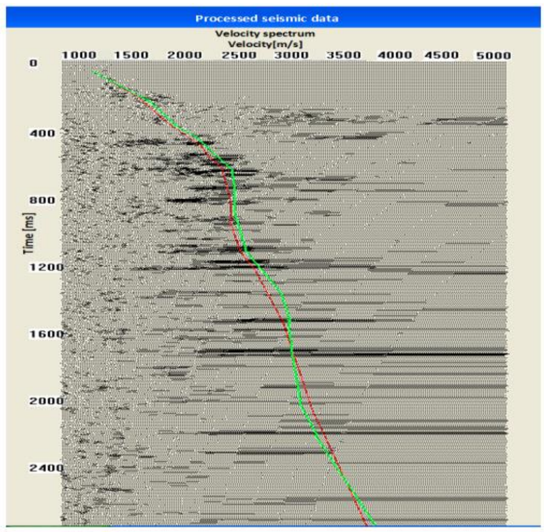

Figure 5. Velocity curve from Bootstrap method (red) and that from a commercial software (green).

In the program, the number of the Bootstrap traces B is not very large [2]. To get more ideal results, B is better large enough, but it depends on your computer. A larger B will spend a longer calculation time, and need more memory [5] [6].

4.2. Application results

Real seismic data were processed using different special software, and the results were compared by taking CGG result as the standard [9].

After appropriate bench-marking operations, the velocity spectrum from commercial software was projected onto our software [10]. Comparison demonstrates that the two spectra from the commercial software and our software are basically consistent, especially in the target layer [11].

5. Conclusions and suggestions

The Bootstrap velocity analysis method has obvious advantages. Using old seismic data, and without additional investment, the Bootstrap method, depending on mathematical statistic and combined with conventional velocity analysis, can significantly improve velocity spectra, and provide unexpected results, especially in processing data of shallow and thin layers with lateral changes. Application has proved the method meaningful and practical.

The method is still in the experimental stage, and there are many problems that have not been solved perfectly. For example, the error is unstable, the velocity spectrum can be improved only to a limited extent, the computer is highly demanded in computing ability, and the memory must be large. More researches should be carried out on the Bootstrap method. Based on the existing experimental experience, future works are suggested to focus on how to optimize the algorithm flowchart, reduce the number of iterations and adapt seismic data in various formats.

References

[1] Liming Wang, Baohua Liu, Qingchun Li 1998 An effective velocity analysis method 30 May 2008
[2] Mauricio D S 2013 A Bootstrap Procedure for High-Resolution Velocity Analysis[J] *Geophysics* **63** 1716-1725
[3] Liming Wang, Qingcuhn Li 2006 Correlation method Selected to improve the accuracy of wave velocity analysis *Progress in Geophysics* 1 December 2006
[4] H M Deitel, P J Deitel 2004 Programming tutorial (version 4) Tsinghua
[5] Silverman B W 1986 Density Estimation for Statistics and Data Analysis[M] New York: Chapman and Hall
[6] Key S C, Smithson S B 1990 New Approach to Seismic-Reflec-tion Event Detection and Velocity Deter-Mination[J] *Geo-physics* **55(10)** P1057-1069
[7] Efron B, Tibshirani R 1993 An Introduction to the Bootstrap Method[M]. New York: Chapman and Hall
[8] Effon B 1979 Bootstrap Methods: Another Look at the Jackknife[J]. *Ann Statist* **7** P1-26
[9] Neidell N S, Taner M T 1971 Semblance and Other Coherency Measures for Multichannel Data[J] *Geophysics* **36** P482-497
[10] Yilmaz O 1993 Seismic data processing *Soc.Explor. Geoghys* 1993
[11] Jinggui Lv, Jinyao Gao 2011 Improved stacking velocity analysis method *Geophysical prospecting for petrleum*