International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 7, Issue 1 (January 2018), PP. 73-95

Quick porosity prediction for carbonate reservoirs using modified N-way analysis of variance

Yufeng Gu[a, b],Zhidong Bao[a, *],Zhenhua Rui[c]
[A]*college Of Geosciences, China University Of Petroleum-Beijing, Beijing, China*
[B]*petrochina Research Institute Of Petroleum Exploration And Development, Beijing, China*
[C] *University Of Alaska Fairbanks, USA*
[*]*Corresponding Author: Yufeng Gu[a,]*

Abstract: Physical characteristic models are generally considered as the best way to evaluate reservoirs, whereas require some essential geological parameters involved in calculation. On purpose of promoting the efficiency of logging interpretation, a new combined method of N-way analysis of variance and multivariate linear fitting, thus, is proposed based on the statistical theories. The combined method has the capability of predicting porosity for the carbonate reservoirs without aiding any geological parameters. Additionally, according to the statistical results of multiple comparisons, the fitted values can be accurately corrected, further. The whole processing procedures are completed in only three steps. Using N-way analysis of variance to optimally select logging curves is the first step. Then figuring out values by applying the established multivariate linear fitting model. Finally attempting to improve the accuracy of fitted results with the creative algorithm of fitting correction. Five cases are comprehensively analyzed in the verification of proposed method. As a results, the mean absolute errors of all the fitted values calculated by N-way analysis of variance are relatively smaller than that only by the multivariate linear fitting method. After the fitting correction, almost all the fitted results are getting much closer to the core data, especially in the case that the accuracy can be raised up by nearly 30% when the robustness of the fitted coefficients is enhanced. All the calculated results and analytical conclusions convincingly prove the fact that the combined method can predict the porosity of lacustrine carbonate formations in a cost-efficient way, and the results are reliable enough to sever as the reference data for other relative work.

Keywords: reservoir evaluation; logging interpretation; porosity prediction; N-way analysis of variance; multivariate linear fitting; fitting correction

---

Date of Submission: 19 -01-2018 Date of acceptance: 31-01-2018

I. INTRODUCTION

Generally, porosity prediction is considered as one of the most important contents for reservoir evaluation. With deep analyzing of porosity data, many reservoir characteristics can be comprehended in the aspects of storing capability of fluid or diagenesis processes, etc. At present, theories and methods system of logging interpretation is mainly composed of physical model, artificial intelligence and statistical analysis (Down, P.A., 1992; Bassiouni, Z., 1994; Asoodeh, M., Bagheripour, P., 2013a). Physical models established based on reservoir physical properties usually applies geological statistical parameters which are normally acquired from core experiments. On early developing stage of logging interpretation, one of the most practical models is average time equation (Wyllie, M.R.J., et al., 1956). Such equation can effectively calculate the porosity but is inappropriate for formations with high shale content and developed fractures. Furthermore, Raymer (1980) studied the relationship between shale content and acoustic time through doing a great deal of core experiments, and proposed an improved model. Although giving satisfactory prediction for shale-sand reservoirs, this model neglects the influence of fractures which will cause cycle skip of acoustic waves. Under researching of transmission properties of acoustic waves in sand-shale formations, Xu and White (1995) put forward an important velocity model as tools to assist other models in solving reservoir evaluation. Keys (2002) simplified the velocity model in accordance with geophysical theories to expand application range. In view of development of both fractures and faults formed in carbonate formations, more geological parameters are required to involve in the physical model establishment for reservoir characterization (such as pore fluid pressure, fluid viscosity coefficients, geostress, etc.) (Gupta, A., Civan, F., 1994; Hamilton, E.L., et al., 1982; Han, D.A., et al., 1986; Kahraman, S., Yeken, T., 2008; Boadu, F.K., 2001; He, J., et al., 2016). Therefore, the workload and cost of projects are increased, and logging interpretation becomes more challenging.

In some regions, due to complex depositional situation or great depth of target layers and limited number of exploratory wells and well logging data, many parameters of physical models cannot be estimated. However, artificial intelligence algorithms have the capability of evaluating reservoir only by applying their exclusive prediction system. As humans learn knowledge and recognize things, such systems are generally built up by analyzing and matching basic data. (Emilson, P.L., ALexander, C.V., 2011; Nahser, M.A., Wang, Y.H., 2011; Jamialahmadi, M., Javadpour, F.G., 2000; Helle, H.B., et al., 2001; Ou, C., et al., 2016). However, in most cases these algorithms still need larger amounts of basic data to establish high-quality prediction system in order to ensure the validity of results, indicating that such kind of methods are also unsuitable for solving logging interpretation problems on condition of lacking of basic data (Rui, Z., et al., 2013a, 2013b; Sun, J., et al., 2016; Zhao, X., et al., 2014, 2016).

Linear fitting or regression is one of the most widely used statistical methods for reservoir evaluation (Ehrenberg, S.N., 1993; Moraes, M.A.S., De, Ros, L.F., 1990; Rui, Z., et al., 2011a, 2011b). This method often builds up the fitting equations with little logging data and uses fitted coefficients to predict parameters. In General, results obtained from the

multivariate linear fitting models are much more accurate than that from the linear fitting ones. The reason is that the interrelations and responses between variables and targets are considered and analyzed in the multivariate models (Byrnes, A.P., Wilson, M.D., 1991; Zhao, X., et al., 2015a, 2015b). Whereas those irrelevant variables will bring about passive impacts on calculation of the multivariate linear fitting, which leads to the unreliability for the fitted results. Analysis of variance, a kind of statistic method, select useful data sets based on their significance (Fisher, R.A., 1936). Based on the function of analysis of variance, those critical influence factors can be validly picked out from the abundant raw logging data, and then the multivariate linear fitting will be used in a reasonable way (Lindley, D.A., 1957; Berger, J.O., Delampady, M., 1987; Berger, J.O., Sellke, T., 1987; Yi, K.N., Cheng X.J., 2013; Rui, Z., et al., 2012a, 2012b). As each kind of the analytical data must be graded into several levels before operating analysis of variance, the special relationship between targets and so-called "level combinations" consisting of graded data can be found. Thus according to the statistical results of the level combinations, each fitted value obtained by the multivariate linear fitting model can be accurately corrected in accordance with such relationship, to some degree.

The purpose of the study is to acquire the porosity of lacustrine carbonate formations in a cost-efficient way without aiding any other geological parameters. Furthermore, applying the results of multiple comparisons to improve the accuracy of fitted values in a special way. Therefore, a new combined method of N-way analysis of variance and multivariate linear fitting is proposed.

II. METHODOLOGY

At present, analysis of variance has been developed into the different types such as the one-way, two-way, or three-way, etc. (Delampady, M., Berger, J., 1987; Dempster, A.P., 1971; Dickey, J.M., 1971). For predicting porosity usually needs two or more logging curves, the type of analysis of variance should be N-way but not in interaction. No detailed deductions of N-way analysis of variance are available to use, so the mathematic analysis is decided to start by referencing the theory of the three-way.

2.1. Model Building

Suppose that there are three factors respectively labeled with $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$. Factor $\mathbf{A}$ can be divided into r levels, and marked with A_1, A_2, $\cdots$, A_r; factor $\mathbf{B}$ divided into s levels B_1, B_2, $\cdots$, B_s; factor $\mathbf{C}$ divided into t levels C_1, C_2, $\cdots$, C_t. Target sample is set as $\mathbf{X}$. If all combinations (A_i, B_j, C_k) corresponded to X_{ijk} are independent and comply with the normal distribution, then the following equations can be given as:

$$X_{ijk} \sim N(\mu_{ijk}, \sigma^2) \tag{1}$$

$$\begin{cases} \mu = \dfrac{1}{rst}\sum\limits_{i=1}^{r}\sum\limits_{j=1}^{s}\sum\limits_{k=1}^{t}\mu_{ijk} \\ \mu_{i\cdot\cdot} = \dfrac{1}{st}\sum\limits_{j=1}^{s}\sum\limits_{k=1}^{t}\mu_{ijk} \\ \mu_{\cdot j\cdot} = \dfrac{1}{rt}\sum\limits_{i=1}^{r}\sum\limits_{k=1}^{t}\mu_{ijk} \\ \mu_{\cdot\cdot k} = \dfrac{1}{rs}\sum\limits_{i=1}^{r}\sum\limits_{j=1}^{s}\mu_{ijk} \end{cases} \tag{2}$$

$$\begin{cases} \alpha_i = \mu_{i\cdot\cdot} - \mu \\ \beta_j = \mu_{\cdot j\cdot} - \mu \\ \gamma_k = \mu_{\cdot\cdot k} - \mu \end{cases} \tag{3}$$

where μ is the sample mean, $\mu_{i\cdot\cdot}$ the mean of *i*'th level of factor $\mathbf{A}$, $\mu_{\cdot j\cdot}$ the mean of *j*'th level of factor $\mathbf{B}$, $\mu_{\cdot\cdot k}$ the mean of *k*'th level of factor $\mathbf{C}$, α_i the deviation of $\mu_{i\cdot\cdot}$ for the μ, β_j the deviation of $\mu_{\cdot j\cdot}$ for the μ, γ_k the deviation of $\mu_{\cdot\cdot k}$ for the μ, $i = 1, 2, \cdots, r$, $j = 1, 2, \cdots, s$, $k = 1, 2, \cdots, t$.

Obviously, α_i, β_j, γ_k satisfy the expressions below according to the Eqs. (2) and (3):

$$\begin{cases} \sum\limits_{i=1}^{r}\alpha_i = 0 \\ \sum\limits_{j=1}^{s}\beta_j = 0 \\ \sum\limits_{k=1}^{t}\gamma_k = 0 \end{cases} \tag{4}$$

if $\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$, then the model of three-way analysis of variance can be written as below:

$$\begin{cases} X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk} \\ \sum_{i=1}^{r} \alpha_i = 0, \sum_{j=1}^{s} \beta_j = 0, \sum_{k=1}^{t} \gamma_k = 0 \\ \varepsilon_{ijk} \sim N(0, \sigma^2) \end{cases} \tag{5}$$

where ε_{ijk} is the error.

2.2. Hypothesis Test

The judging equations of hypothesis test are set below:

$$\begin{cases} \mathrm{H}_{01} : \alpha_1 = \alpha_2 = \cdots = \alpha_{\mathrm{r}} = 0 \\ \mathrm{H}_{02} : \beta_1 = \beta_2 = \cdots = \beta_{\mathrm{s}} = 0 \\ \mathrm{H}_{03} : \gamma_1 = \gamma_2 = \cdots = \gamma_{\mathrm{t}} = 0 \end{cases} \tag{6}$$

where H_{01} is the first hypothesis condition that all levels of factor **A** equal to the sample mean; H_{02} the second hypothesis condition that that all levels of factor **B** equal to the sample mean; H_{03} the third hypothesis condition that all levels of factor **C** equal to the sample mean.

2.3. Sum of Square of Deviation and Decomposition Equation

Several essential variables are designed as shown below:

$$\begin{cases} \overline{X} = \dfrac{1}{rst} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} X_{ijk} \\ \overline{X}_{i\cdot\cdot} = \dfrac{1}{st} \sum_{j=1}^{s} \sum_{k=1}^{t} X_{ijk} \\ \overline{X}_{\cdot j\cdot} = \dfrac{1}{rt} \sum_{i=1}^{r} \sum_{k=1}^{t} X_{ijk} \\ \overline{X}_{\cdot\cdot k} = \dfrac{1}{rs} \sum_{i=1}^{r} \sum_{j=1}^{s} X_{ijk} \end{cases} \tag{7}$$

if S_{T} is written as:

$$S_{\mathrm{T}} = \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (X_{ijk} - \overline{X})^2 = S_{\mathrm{E}} + S_{\mathrm{A}} + S_{\mathrm{B}} + S_{\mathrm{C}} \tag{8}$$

then the expressions of sum of deviation squares for each term in Eq. (8) are derived as follows:

$$\begin{cases} S_{\mathrm{A}} = st \sum_{i=1}^{r} (\overline{X}_{i\cdot\cdot} - \overline{X})^2 = \sum_{i=1}^{r} \dfrac{1}{st} \overline{X}_{i\cdot\cdot}^2 - n\overline{X}^2 \\ S_{\mathrm{B}} = rt \sum_{j=1}^{s} (\overline{X}_{\cdot j\cdot} - \overline{X})^2 = \sum_{j=1}^{s} \dfrac{1}{rt} \overline{X}_{\cdot j\cdot}^2 - n\overline{X}^2 \\ S_{\mathrm{C}} = rs \sum_{k=1}^{t} (\overline{X}_{\cdot\cdot k} - \overline{X})^2 = \sum_{k=1}^{t} \dfrac{1}{rs} \overline{X}_{\cdot\cdot k}^2 - n\overline{X}^2 \\ S_{\mathrm{E}} = S_{\mathrm{T}} - S_{\mathrm{A}} - S_{\mathrm{B}} - S_{\mathrm{C}} = \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (X_{ijk} - \overline{X}_{i\cdot\cdot} - \overline{X}_{\cdot j\cdot} - \overline{X}_{\cdot\cdot k} + 2\overline{X})^2 \end{cases} \tag{9}$$

where S_{A} is the sum of deviation squares of factor **A**, S_{B} the sum of deviation squares of factor **B**, S_{C} the sum of deviation squares of factor **C**, S_{E} the sum of deviation squares of error.

On the basis of Eq. (9), S_{A}, S_{B}, S_{C} respectively obey the distribution as:

$$\begin{cases} S_{\mathrm{A}} \sim \chi^2(r-1) \\ S_{\mathrm{B}} \sim \chi^2(s-1) \\ S_{\mathrm{C}} \sim \chi^2(t-1) \end{cases} \tag{9}$$

2.4. Non-interactive Influence Analysis

Aiming at the independent properties of analytical factors, non-interactive influence should be analyzed in the following processes. In this step, those variables that the degree of freedom of the factor **A**, **B**, **C**, error **E**, total sum T, the mean square of the factor **A**, **B**, **C**, error **E** and the F value of the factor **A**, **B**, **C** are all needed to be calculated out. The equations used to compute the degree of freedom for all terms are presented as below:

$$\begin{cases} A_{\mathrm{d.f.}} = r-1 \\ B_{\mathrm{d.f}} = s-1 \\ C_{\mathrm{d.f}} = t-1 \\ E_{\mathrm{d.f}} = (r-1)(s-1)(t-1)+(r-1)(s-1)+(r-1)(t-1)+(s-1)(t-1) \\ T_{\mathrm{d.f}} = rst-1 \end{cases} \tag{10}$$

The sum square and F value are expressed as:

$$\begin{cases} MS_{\mathrm{A}} = S_{\mathrm{A}}/(r-1) \\ MS_{\mathrm{B}} = S_{\mathrm{B}}/(s-1) \\ MS_{\mathrm{C}} = S_{\mathrm{C}}/(t-1) \\ MS_{\mathrm{E}} = S_{\mathrm{E}}/(r-1)(s-1)(t-1)+(r-1)(s-1)+(r-1)(t-1)+(s-1)(t-1) \end{cases} \tag{11}$$

$$\begin{cases} \mathrm{F}_{\mathrm{A}} = MS_{\mathrm{A}}/MS_{\mathrm{E}} \\ \mathrm{F}_{\mathrm{B}} = MS_{\mathrm{B}}/MS_{\mathrm{E}} \\ \mathrm{F}_{\mathrm{C}} = MS_{\mathrm{C}}/MS_{\mathrm{E}} \end{cases} \tag{12}$$

2.5. Testing Rules

In order to reveal the significance of each factor, F value should be compared with the confidence level α. The testing rules are given as the follows:

if $\mathrm{F}_{\mathrm{A}} > \mathrm{F}_{1-\alpha}(r-1,(r-1)(s-1)(t-1)+(r-1)(s-1)+(r-1)(t-1)+(s-1)(t-1))$, H_{01} is rejected, which indicates that each level of factor **A** has the significance.

if $\mathrm{F}_{\mathrm{B}} > \mathrm{F}_{1-\alpha}(s-1,(r-1)(s-1)(t-1)+(r-1)(s-1)+(r-1)(t-1)+(s-1)(t-1))$, H_{02} is rejected, which indicates that each level of factor **B** has the significance.

if $\mathrm{F}_{\mathrm{C}} > \mathrm{F}_{1-\alpha}(t-1,(r-1)(s-1)(t-1)+(r-1)(s-1)+(r-1)(t-1)+(s-1)(t-1))$, H_{03} is rejected, which indicates that each level of factor **C** has the significance.

2.6. N-way analysis of variance

Assume that there are N factors with the labels of N_1, N_2, $\cdots$, N_{N}. All factors are independent and each of them has r levels. Before establishing the variance model, some marks should be defined as below in order to simply the expression for the following analysis:

if $i1 = 1,2,\cdots,r$, $i2 = 1,2,\cdots,r$, $\cdots$, $i\mathrm{N} = 1,2,\cdots,r$,

then set $\Lambda = \underbrace{i1,\cdots,i\mathrm{N}}_{\mathrm{N}}$, $\Delta_{1j} = \underbrace{i1\cdots\cdots}_{\mathrm{N}}$, $\Delta_{2j} = \underbrace{i2\cdots\cdots}_{\mathrm{N}}$, $\cdots$, $\Delta_{\mathrm{N}j} = \underbrace{\cdots\cdots i\mathrm{N}}_{\mathrm{N}}$, $\Sigma_{\mathrm{Q}_N} = \overbrace{\sum_{i1=1}^{r},\cdots,\sum_{i\mathrm{N}=1}^{r}}^{\mathrm{N}}$,

$\Sigma_{\mathrm{Q}_{N-1}} = \overbrace{\sum_{i1=1}^{r},\cdots,\sum_{i\mathrm{N}=1}^{r}}^{\mathrm{N-1}}$, $\Sigma_{P_N} = \overbrace{\sum_{j=1}^{r},\cdots,\sum_{j=1}^{r}}^{\mathrm{N}}$.

According to the whole deductions of the three-way analysis of variance and the defined marks above, some fundamental equations can be obtained as below:

$$X_{\Lambda} \sim N(\mu_{\Lambda}, \sigma^2) \tag{13}$$

$$\begin{cases} \mu = \dfrac{1}{r^N}\Sigma_{Q_N}\mu_\Lambda \\ \mu_{\Delta 1j} = \dfrac{1}{r^{N-1}}\Sigma_{Q_{N-1}}\mu_\Lambda \\ \quad\vdots \\ \mu_{\Delta Nj} = \dfrac{1}{r^{N-1}}\Sigma_{Q_{N-1}}\mu_\Lambda \end{cases} \tag{14}$$

$$\begin{cases} \alpha_{\Delta 1j} = \mu_{\Delta 1j} - \mu \\ \alpha_{\Delta 2j} = \mu_{\Delta 2j} - \mu \\ \quad\vdots \\ \alpha_{\Delta Nj} = \mu_{\Delta Nj} - \mu \end{cases} \tag{15}$$

Thus , the model of N-way analysis of variance is:

$$\begin{cases} X_\Lambda = \mu + \sum_{k=1}^{N}\alpha_{\Delta kj} + \varepsilon_\Lambda \\ \sum_{j=1}^{r}\alpha_{\Delta 1j} = 0 \\ \quad\vdots \\ \sum_{j=1}^{r}\alpha_{\Delta Nj} = 0 \\ \varepsilon_\Lambda \sim N(0,\sigma^2) \end{cases} \tag{16}$$

The judging conditions of hypothesis test are:

$$\begin{cases} \mathrm{H}_{01} = \alpha_{11} = \alpha_{12} = \cdots = \alpha_{1r} = 0 \\ \mathrm{H}_{02} = \alpha_{21} = \alpha_{22} = \cdots = \alpha_{2r} = 0 \\ \qquad\vdots \\ \mathrm{H}_{0N} = \alpha_{N1} = \alpha_{N2} = \cdots = \alpha_{Nr} = 0 \end{cases} \tag{17}$$

Based on the processing procedures of non-interactive analysis, the equations of those which are the sum of deviation squares, degree of freedom, sum square and F value are derived out as shown below:

$$\begin{cases} \overline{X} = \dfrac{1}{r^N}\Sigma_{Q_N}X_\Lambda \\ \overline{X}_{\Delta 1j} = \dfrac{1}{r^{N-1}}\Sigma_{Q_{N-1}}X_\Lambda \\ \overline{X}_{\Delta 2j} = \dfrac{1}{r^{N-1}}\Sigma_{Q_{N-1}}X_\Lambda \\ \qquad\vdots \\ \overline{X}_{\Delta Nj} = \dfrac{1}{r^{N-1}}\Sigma_{Q_{N-1}}X_\Lambda \end{cases} \tag{18}$$

$$\begin{cases} S_\mathrm{T} = \Sigma_{Q_N}(X_\Lambda - \overline{X})^2 = S_\mathrm{E} + \sum_{i=1}^{N}S_i \\ S_i = r^{N-1}\sum_{j=1}^{r}(\overline{X}_{\Delta ij} - \overline{X})^2 = \dfrac{1}{r^{N-1}}\sum_{j=1}^{1}X_{\Delta ij} - n\overline{X}^2 (i = 1,\cdots,N) \\ S_\mathrm{E} = \Sigma_{P_N}(X_\Lambda - \sum_{i=1}^{N}\overline{X}_{ij} + (N-1)\overline{X})^2 \end{cases} \tag{19}$$

$$S_i \sim \chi^2(r-1) \tag{20}$$

The degree of freedom for each factor is $r-1$, the error $(r-1)^N + N(r-1)^{N-1}$, the total sum r^{N-1}. The sum square and F value, hence, can be obtained as:

$$\begin{cases} MS_i = S_i / (r-1) \\ MS_{\mathrm{E}} = S_{\mathrm{E}} / (r-1)^N / N(r-1)^{N-1} \end{cases} \tag{20}$$

$$\mathrm{F}_i = MS_i / MS_{\mathrm{E}} \tag{21}$$

If F_i satisfies the judging condition $\mathrm{F}_i > \mathrm{F}_{1-\alpha}(r-1,(r-1)^N + N(r-1)^{N-1})$, H_{0i} is rejected, which indicates that all levels of factor N_i have the significance.

2.7. Fitting Correction

Based on the theories of linear fitting, if p value of constant term is much larger than that of all coefficient terms, the variation of the fitted results will be primarily controlled by the constant, which will cause all the fitted values to be nearly same. Thus, the constant term should be removed in order to improve the impact degree of the coefficient terms on the fitted results. Accordingly, the fitting equation is advised to change into the form as below:

$$\tilde{y} = C + \sum_{i=1}^{n} a_i N_i \xrightarrow{C=0} \sum_{i=1}^{n} b_i N_i \tag{22}$$

where C is the constant term, a_i and b_i the fitted coefficient, $\tilde{y}$ the fitted value, n the number of coefficient term.

As all data sets must be graded into several levels before processing the N-way analysis of variance, the particular relationships between the fitted values acquired by Eq. (22) and those so-called "level combinations" mainly consisting of graded analytical data can be build up. Then the fitted results, to some degree, will be further and detailedly corrected in an effective way with using some defined computational rules. In this paper, a creative algorithm of fitting correction is proposed. The first step of this algorithm is to confirm the deviation trend of target sample to be larger or smaller compared with the mean, which can be realized by statistical methods such as maximum likelihood estimation. Define that C_{great} and C_{small} that the number of larger and smaller values are respectively determined by the expressions $\mathrm{C}_{\mathrm{great}} = \mathbf{X} > \mu$ and $\mathrm{C}_{\mathrm{small}} = \mathbf{X} > \mu$. If $C_{\mathrm{great}} > C_{\mathrm{small}}$, all values of the sample tend to be larger. Then all X_Λ are decided to sort according to the formula as:

$$\mathbf{X}_\Delta = \mathrm{sort}_{\mathrm{descend}}(X_\Lambda) \tag{23}$$

where $\mathbf{X}_\Delta$ is the inverted order of all X_Λ.

Based on vector $\mathbf{X}_\Delta$, those special relationships between sample values and the related level combinations can be revealed well. Note that if one level combination corresponds to several different target values X_Λ, the mean will be assigned as the new target value to this level combination. However, the degrees of deviation of all X_Λ do not determinate in the vector $\mathbf{X}_\Delta$, or in other words, $\mathbf{X}_\Delta$ does not present the significant degree of all level combinations. So Eq. (23) has to be transformed into another form as the expression below:

$$\mathbf{X}_{\Delta\mathrm{con\_level}} = \mathbf{X}_\Delta / X_{\Delta\mathrm{MAX}} \tag{24}$$

where $\mathbf{X}_{\Delta\mathrm{con\_level}}$ includes the significant degrees of all X_Λ, $X_{\Delta\mathrm{MAX}}$ the maximum of $\mathbf{X}_\Delta$.

According to the changing trend of the target sample, if the significant degree in $\mathbf{X}_{\Delta\mathrm{con\_level}}$ is higher, the corresponding fitted value can be considered to be more reliable. As little logging data are available to evaluate reservoirs, $\mathbf{X}_{\Delta\mathrm{con\_level}}$, the target sample cannot reflect the actual variation of the totality, or in other words, partial values of $\mathbf{X}_{\Delta\mathrm{con\_level}}$ are incredible. With the purpose of applying the significant degree in a reasonable way, $\mathbf{X}_{\Delta\mathrm{con\_level}}$ is intended to divide into the credible and incredible two sections with the defined boundary α_Δ. This parameter can be respectively described by the expressions $\mathbf{X}_{\Delta\mathrm{con\_level}} \geq \alpha_\Delta$ and $\mathbf{X}_{\Delta\mathrm{con\_level}} < \alpha_\Delta$. In order to make the range of significant degree be same with that of the credible sections from 1 to α_Δ, the incredible section should be recalculated by the following formula:

$$\mathbf{X}_{\Delta\alpha\_\mathrm{level}} = \mathbf{X}_{\Delta\mathrm{less}\_\alpha} / \alpha_\Delta \tag{25}$$

where $\mathbf{X}_{\Delta\mathrm{less}\_\alpha}$ is the incredible section, $\mathbf{X}_{\Delta\alpha\_\mathrm{level}}$ the significant degree vector of the incredible section.

The aim of transformation of Eq. (25) is to let the incredible sections can be analyzed and processed in a similar or even same way with the credible one in order to simply the processing procedures of fitting correction.

As the significant degrees within the credible section are all relatively reliable, the fitted values in this step can be directly corrected by the standard errors which are obtained from the multiple comparisons and included in the level

combinations. If the fitted value is larger or smaller than the corresponded X_{Λ}, it should be corrected by reducing or adding the standard error. The equations for this kind of fitting correction are given as below:

$$\tilde{y}_{\text{new}\_i} = \begin{cases} \tilde{y}_i - X_{\Lambda\text{std}}, (\tilde{y}_i > X_{\Lambda}) \\ \tilde{y}_i + X_{\Lambda\text{std}}, (\tilde{y}_i \le X_{\Lambda}) \end{cases} \tag{26}$$

where $\tilde{y}_{new\_i}$ is the fitting corrected value, $\tilde{y}_i$ the fitted value, $X_{\Lambda\text{std}}$ the standard error of *i*'th level combination.

For the incredible section, several circumstances should be discussed so the processes will be complicated. If the distance (difference) between X_{Λ} and the corresponding fitted value is less than $X_{\Lambda\text{std}}$, the fitted value also can be directly corrected by $X_{\Lambda\text{std}}$. Because all X_{Λ} are relatively unreliable in this section, the distance should be expanded, or in other words, the fitted values must be changed far away instead of getting closer to the X_{Λ}. The fitting correction, therefore, will be processed inversely in comparison with Eq. (26), which implies that the fitted value should be added with correction volume if the fitted value is larger, or subtracted on the contrary. The fitting correction equations in this case can be written as:

$$\tilde{y}_{new\_i} = \begin{cases} \tilde{y}_i + \text{abs}(\tilde{y}_i - X_{\Lambda}) * (1 - X_{a\_\text{level}\_i}), (\tilde{y}_i > X_{\Lambda}) \\ \tilde{y}_i - \text{abs}(\tilde{y}_i - X_{\Lambda}) * (1 - X_{a\_\text{level}\_i}), (\tilde{y}_i \le X_{\Lambda}) \end{cases} \tag{27}$$

$$s.t. \quad \text{abs}(\tilde{y}_i - X_{\Lambda}) \le X_{\Lambda\text{std}}$$

where $X_{\alpha\_\text{level}\_i}$ is *i*'th significant degree of $\mathbf{X}_{\Delta\alpha\_\text{level}}$.

In Eq. (27), the significant degree of the fitted value is viewed as $X_{\alpha\_\text{level}\_i}$. On account of the definition of correction volume described above, the degree that corresponds to the expanded part of the distance should be $1 - X_{\alpha\_\text{level}\_i}$, which explains the calculation principle of fitting correction for this situation.

If distance is larger than $X_{\Lambda\text{std}}$, the fitted value will be considered as the most unreliable one in all circumstances. So the value must be recalculated again to decrease the distance in order that it can processed in a reasonable way with the correction equations as Eq. (27). There are a variety of recalculated methods, e.g. image method which can reasonably modify the fitted values with the defined upper or lower boundaries based on the reflection principle. If the distance between the modified value and X_{Λ} is less than $X_{\Lambda\text{std}}$, the modified value will be used in the following processes instead of the original fitted value. Otherwise, if the distance is still rather larger, the modified value cannot be accepted until it satisfies the requirement of the fitting correction for this situation.

The modified value is defined as $\tilde{y}'$, then the corresponding fitting correction equations can be written as the follows:

$$\tilde{y}_{new\_i} = \begin{cases} \tilde{y}'_i - \text{abs}(\tilde{y}'_i - X_{\Lambda}) * (1 - X_{\alpha\_\text{level}\_i}) (\tilde{y}'_i > X_{\Lambda} \,\&\, \text{abs}(\tilde{y}'_i - X_{\Lambda}) \le X_{\Lambda\text{std}}) \\ \tilde{y}_i + \text{abs}(\tilde{y}'_i - X_{\Lambda}) * (1 - X_{\alpha\_\text{level}\_i}) (\tilde{y}'_i \le X_{\Lambda} \,\&\, \text{abs}(\tilde{y}'_i - X_{\Lambda}) \le X_{\Lambda\text{std}}) \end{cases}$$

$$s.t. \begin{cases} \tilde{y}_i > X_{\Lambda} \\ \text{abs}(\tilde{y}_i - X_{\Lambda}) > X_{\Lambda\text{std}} \end{cases} \tag{28}$$

$$\tilde{y}_{new\_i} = \begin{cases} \tilde{y}'_i + \text{abs}(\tilde{y}'_i - X_{\Lambda}) * (1 - X_{\alpha\_\text{level}\_i}) (\tilde{y}'_i > X_{\Lambda} \,\&\, \text{abs}(\tilde{y}'_i - X_{\Lambda}) \le X_{\Lambda\text{std}}) \\ \tilde{y}'_i - \text{abs}(\tilde{y}'_i - X_{\Lambda}) * (1 - X_{\alpha\_\text{level}\_i}) (\tilde{y}'_i \le X_{\Lambda} \,\&\, \text{abs}(\tilde{y}'_i - X_{\Lambda}) \le X_{\Lambda\text{std}}) \end{cases}$$

$$s.t. \begin{cases} \tilde{y}_i \le X_{\Lambda} \\ \text{abs}(\tilde{y}_i - X_{\Lambda}) > X_{\Lambda\text{std}} \end{cases} \tag{29}$$

2.8 Computational flow

According to the algorithm and processing details concerning the N-way analysis of variance, multivariate linear fitting and fitting correction, the computational flow of the proposed combined method can be designed as shown in Fig. 1. All the essential processing procedures are listed and connected. Based on the collected data, using N-way analysis of variance to select significant logs. Then establishing the predicted equations by the principle of multivariate linear fitting. Finally, under the fitting correction algorithm, the results will be processed again to improve their accuracy. The automatic program can be coded by following such flowcharts.

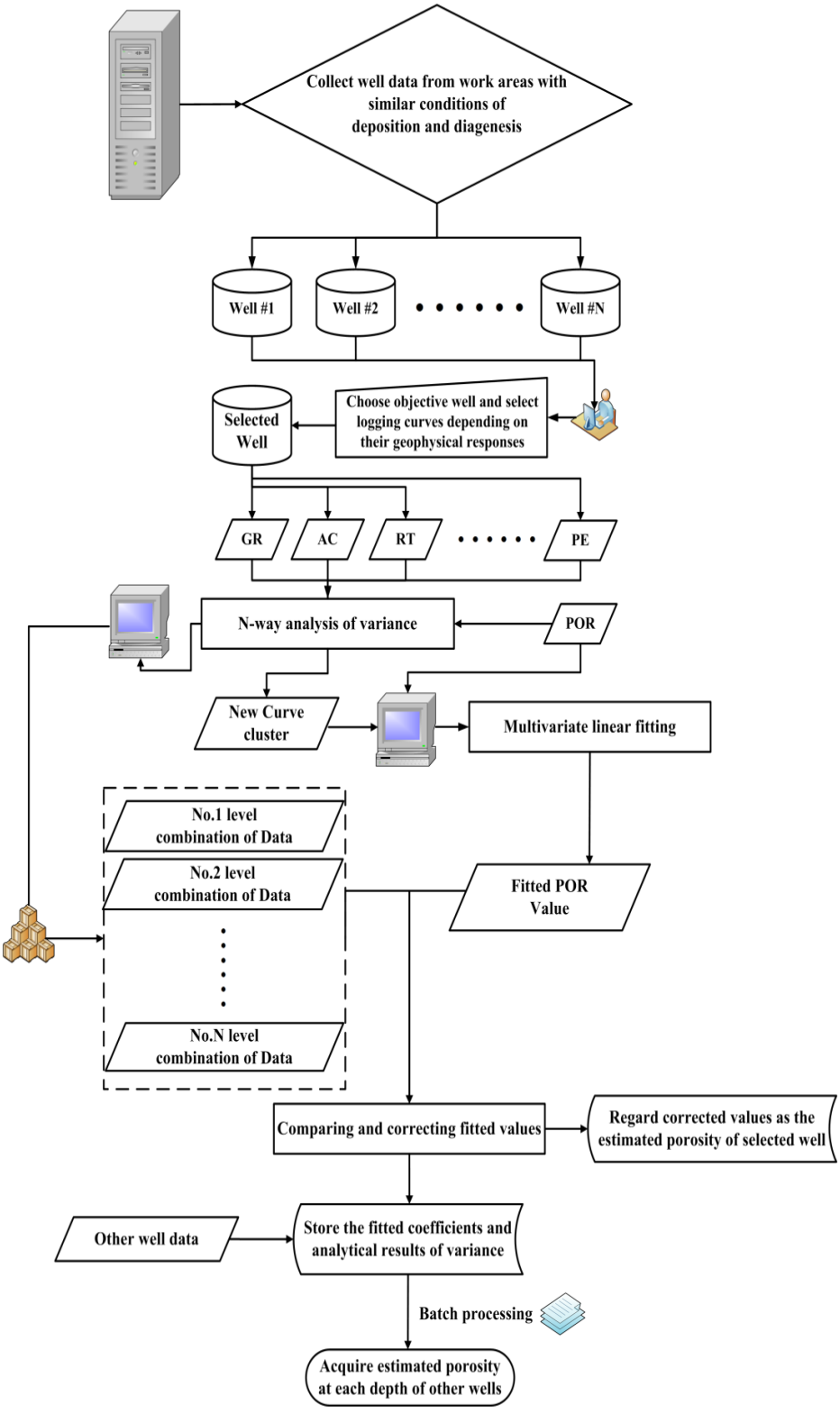


Fig.1. Computational flowchart of combined method of N-way analysis of variance and multivariate linear fitting.

III. DATA SOURCE

Verified data are selected from the Brazilian pre-salt lacustrine carbonate reservoirs. These reservoirs are primarily distributed both in Santos and Campos Basin as presented in Fig. 2, and all of them are regarded as the large-scale petroleum-bearing belts. Among those basins surrounding the Atlantic Ocean , Santos is the largest one in scales and its recoverable oil-gas reserves can reach up to 39.2 bboe which almost shares 86% of the total quantities. The numbers indicate that it is the most important exploration region within the Brazil pre-salt lacustrine carbonate formations (Fainstein, R., et al., 2001; Mohriak, W.U., 2001).

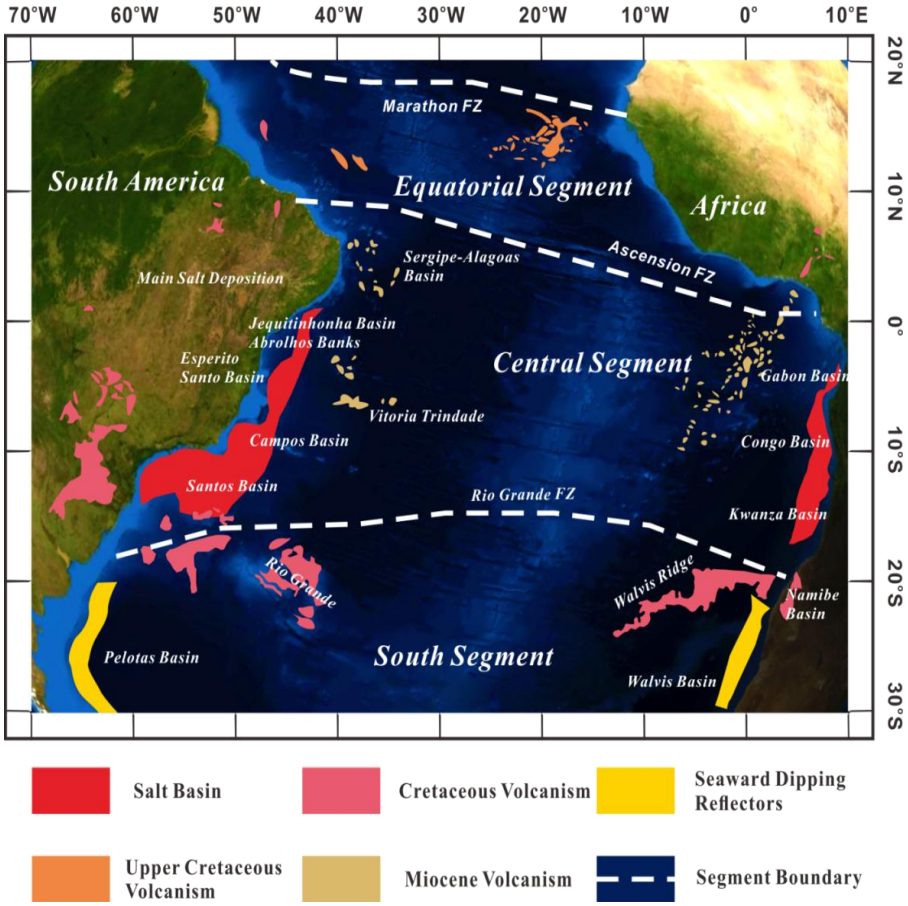


Fig.2. Distribution of Brazilian pre-salt lacustrine carbonate reservoirs.

Based on the deep analysis of regional seismic data and casting lamellas, the lacustrine carbonate formations within Santos Basin perhaps have suffered halokinesis, dolomitization, or hydrothermal erosion during the complex diagenesis evolutions (Mohriak, W.U., 2005; Mohriak, W.U., et al., 2008a; Quirk, D.G., et al., 2012; Jurgen A, et al., 2014; Montaron, B., Tapponier, P., 2010; Rui, Z., et al., 2012c, 2012d). Hence, those formations generally present with the strong fluid storing capability owing to the developed pore-throat and fractured systems. For the reason that the large depth of target layers brings about the high cost for project, only fewer exploration wells have been drilled in the Santos Basin, i.e. IARA oil field which only includes 12 wells. Inadequate data of IARA cannot meet the processing requirement of physical models, and accordingly proposed combined method is adopted to use for porosity prediction.

IV. Results And Discussion

4.1. Curve Selection

Distribution of wells within the work area of IARA oil field is presented in Fig. 3. The study region is mainly consisted by three structural high points which are formed severally with the different tectonic movements and sedimentary cycles (Moczydlower, B., et al., 2012). In order to prove the prediction capability of the proposed method with less words, the logging data derived from any one high point are chose as the target samples for verification. For instance, four wells located in the middle structural high point are selected as samples, namely RJS-715, RJS-656, RJS-726, RJS-682. RJS-715 is a horizontal well which is out of analysis scope, and thus only three wells provide the available logging data, but enough. The sample used for establishing the fitting models can be selected from any wells, i.e. RJS-682. The relative logging curves are shown in Fig. 4. Thereinto, CAL and BS are related to well caliper; GR and SGR have something to do with shale content; all of RXO, RI and RT response to the resistivity from undisturbed to flushed zone; AC, CNL and DEN are all concerned with porosity; PE can be used to analysis lithologies and lithofacies. On account of geophysical responses of each

curve regarding the porosity, GR, SGR, RT, AC, CNL, DEN and PEFZ that seven curves are determined to set as the test data. μ

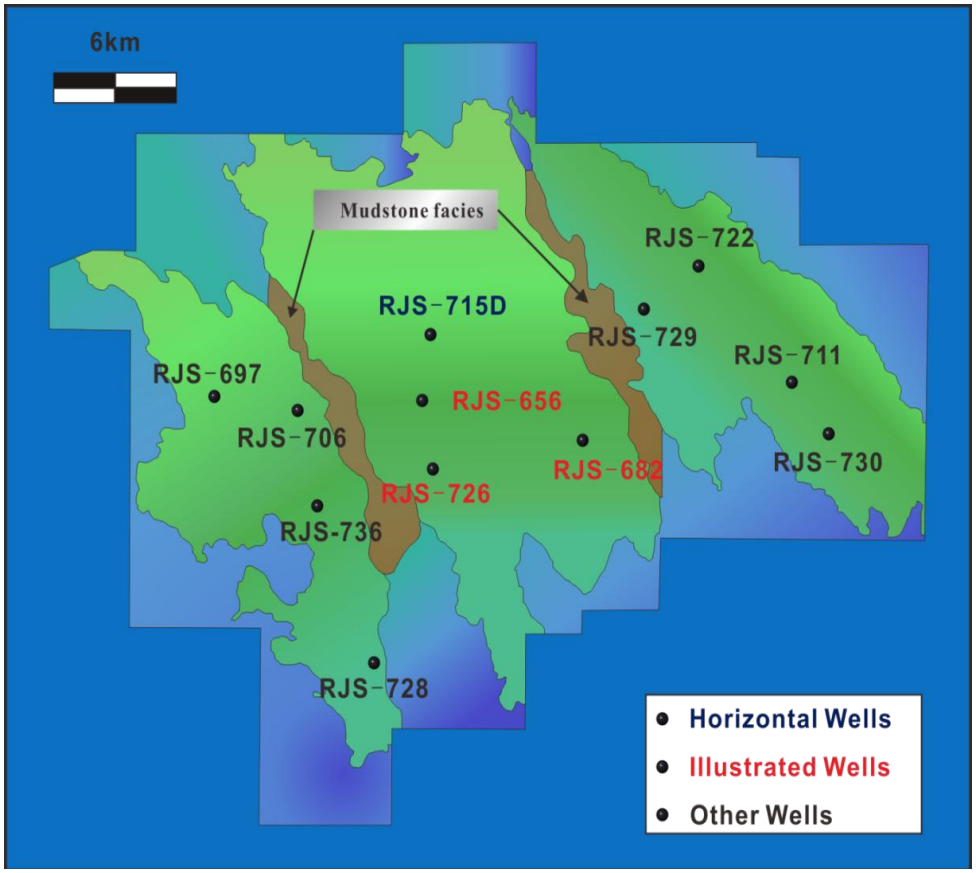


Fig.3. Work area of IARA oil field.

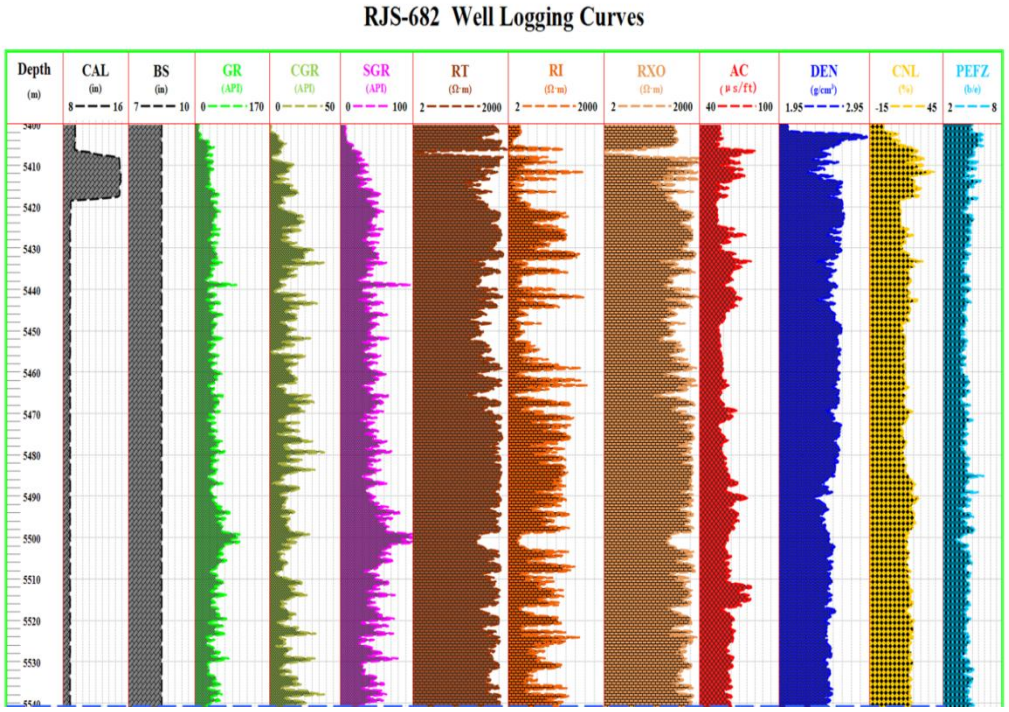


Fig.4. RJS-682 well logging curves.

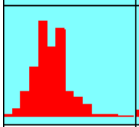
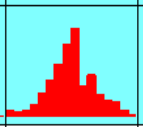
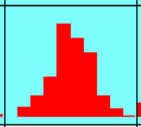
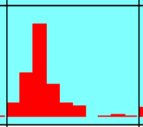
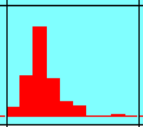
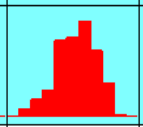
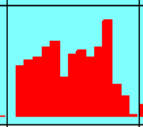
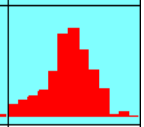
AC	DEN	CNL	GR	SGR	PE	RT	POR
μm/ft	g/cm^3	%	API			Ω•m	%
							
53.496	2.645	10.651	55.666	57.436	4.027	763.088	2.5
60.203	2.608	12.616	37.665	38.979	4.359	780.416	4.2
55.534	2.631	11.953	41.373	42.721	4.281	829.523	2.1
75.83	2.375	26.317	32.373	33.46	3.978	152.337	1.9
68.077	2.54	16.746	43.931	45.202	4.141	764.934	17.5
65.547	2.521	17.23	35.582	36.759	4.122	1091.501	8.3
62.623	2.519	17.959	34.207	35.214	4.729	221.594	5.2
64.416	2.485	19.157	24.529	25.293	4.776	284.515	15.2
58.004	2.587	16.371	82.468	85.112	3.855	165.69	1.8
65.528	2.526	16.19	30.21	31.279	4.794	883.151	8.2
67.897	2.505	17.685	44.499	46.047	3.999	615.038	9.5
70.93	2.491	17.923	41.061	42.447	4.26	1050.18	7.5
65.607	2.476	22.271	44.078	45.536	4.938	430.203	25.2
68.579	2.519	17.398	33.335	34.421	4.341	937.578	3
58.726	2.589	16.439	66.305	68.384	4.208	428.695	8.4
59.241	2.568	16.515	42.495	43.706	4.302	156.183	12
56.781	2.609	14.718	41.517	42.937	4.271	435.881	9.5
55.081	2.582	13.922	66.958	69.058	4.142	120.988	8
56.618	2.613	13.428	31.228	32.17	3.992	497.931	1.6
57.907	2.608	14.116	42.692	44.049	4.295	1002.151	1.3
56.132	2.605	13.224	46.854	48.404	4.346	374.194	1.1
56.376	2.6	15.548	48.401	50.044	4.126	606.344	4.4
60.083	2.566	16.954	38.984	40.045	4.596	547.536	10.2
53.693	2.639	10.725	37.31	38.429	4.329	1156.968	6.7

Fig.5. Partial logging and core porosity data (acquired from MCIP) of RJS-682 well.

4.2. N-way Analysis of Variance

The numeric data of both curves and core porosity are partially presented in Fig. 5. Direct and indirect geophysical responses of those seven logging curves are not all effective in the porosity prediction. So N-way analysis of variance is adopted to filter the irrelative data. Before operating the process, all the curves should be graded into several levels in accordance with the theories. Nothing of the gradation should be complied with but only one point must be noted that both theoretical and actual freedom degree of each curve should be equivalent. According to the gradation principle above, all curves are decided to grade into 3, 4, or 5 levels in order to find out the rational gradation through comparisons. By applying the Eqs. (13) to (21), the analytical results of variance are respectively presented in Tables 1 to 3. In Table 1, freedom degree of both GR and SGR do not reach to the theoretical ones. That implies that these two curves are almost entirely graded into only two levels so the three-tier gradation is unsuitable. Moreover, the freedom degree of DEN in Table 3 is less than the demanded value in theory, and thus five-tire gradation also cannot be accepted. All freedom degrees in Table 4, however, are eligible and thus the four-tier is set as the standard gradation which will be adopted in the follow-up calculations. Assuming that the confidence level is 0.11, those curves of GR, SGR and RT present with small significance due to their larger p values and then should be removed. Therefrom, the logging data have been completely filtered by N-way analysis of variance, and four significant curves are preserved as the fitting variables.

Table 1: Summary of calculated results for the N-way analysis of variance under three-tier gradation

Source	Sum Sq.	d.f.	Mean Sq.	F Value	Prob>F
AC	38.49	2	19.24	1.2	0.3082
CNL	112.97	2	56.48	3.51	0.0348
DEN	176.9	2	88.49	5.49	0.0059
GR	3.13	1	3.12	0.19	0.6607
SGR	11.62	1	11.62	0.72	0.3982
PE	14.73	2	7.36	0.46	0.6346
RT	7.34	2	3.67	023	0.7966
Error	1239.85	77	16.11	/	/
Total	2215.57	90	/	/	/

Table 2 Summary of calculated results for the N-way analysis of variance under four-tier gradation

Source	Sum Sq.	d.f.	Mean Sq.	F Value	Prob>F
AC	88.15	3	29.38	2.1	0.1036
CNL	162.67	3	54.22	3.87	0.0109
DEN	82.71	3	27.56	1.97	0.10219
GR	76.16	3	25.38	1.81	0.148
SGR	64.24	3	21.43	2.2	0.2099

PE	92.53	3	30.84	0.85	0.0909
RT	35.75	3	11.92	1.53	0.4682
Error	1749.15	125	13.99	/	/
Total	3170.72	146	/	/	/

Table 3 Summary of calculated results for the N-way analysis of variance under five-tier gradation

Source	Sum Sq.	d.f.	Mean Sq.	F Value	Prob>F
AC	103.67	4	25.92	1.75	0.1425
CNL	261.58	4	65.39	4.41	0.0021
DEN	66.83	3	22.27	1.5	0.2164
GR	29.45	4	7.36	0.5	0.7386
SGR	77.47	4	14.84	1.31	0.2705
PE	147.62	4	36.91	2.49	0.0457
RT	13.95	4	3.48	0.23	0.9183
Error	2330.1	157	14.84	/	/
Total	3900.42	184	/	/	/

Table 4 Summary of calculated results for the N-way analysis of variance under four-tier gradation

Source	Sum Sq.	d.f.	Mean Sq.	F Value	Prob>F
AC	105.34	3	35.31	2.23	0.0858
CNL	240.97	3	80.32	5.06	0.002
DEN	111.44	3	37.14	2.34	0.0739
PE	166.65	3	55.54	3.5	0.0161
Error	3822.9	241	15.86	/	/
Total	5363.24	253	/	/	/

Based on the standard gradation, the analytical results of these four curves are obtained and presented in Table 4. Compared with the defined confidence level, all p values manifest that the four selected curves have obvious significance on porosity, which proves once again that the results of N-way analysis of variance is correct and the adopted gradation is suitable for processing data.

4.3. Multivariate Linear Fitting

According to the analytical results of variance, four selected logging curves that AC, DEN, CNL, PE are assigned as the variables in the multivariate linear fitting and core POR data as the targets. Therefrom, the fitted equation is derived as shown below:

$$\mathrm{POR} = -0.254 * \mathrm{AC} + 0.439 * \mathrm{CNL} - 30.55 * \mathrm{DEN} + 1.1 * \mathrm{PE} + 91.414 \tag{30}$$

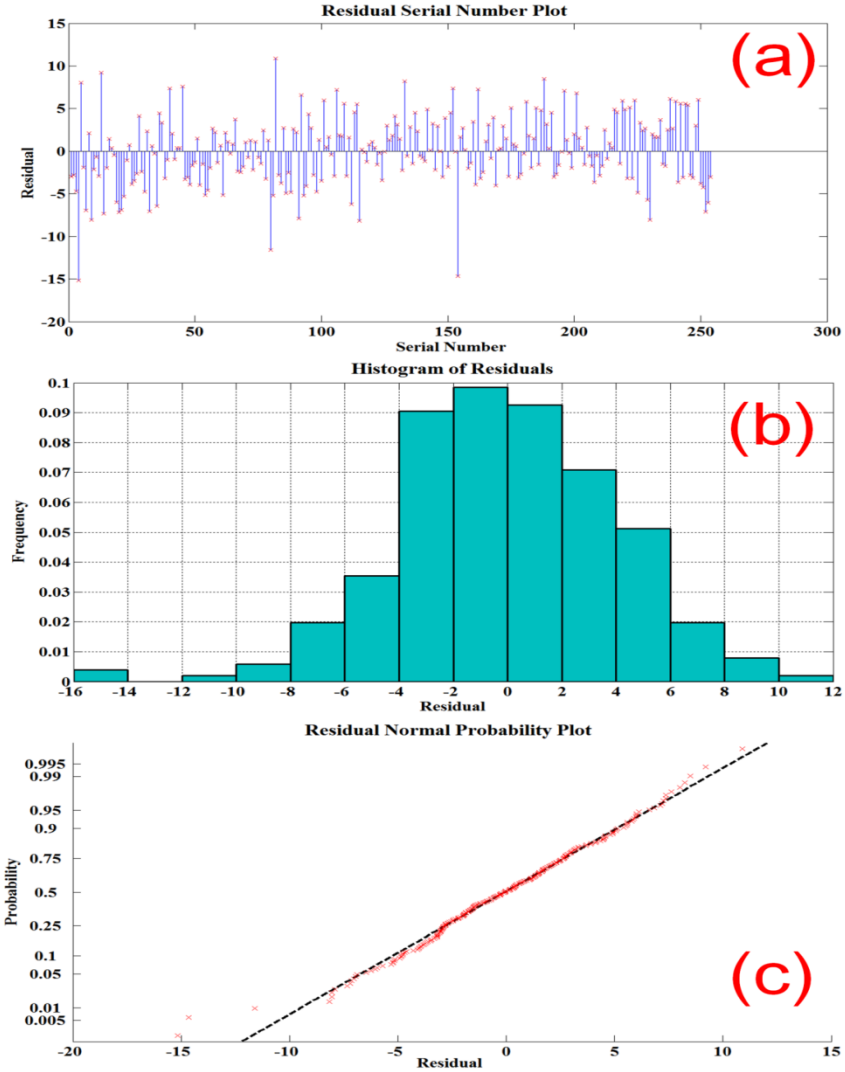
In Eq. (30), the p value of constant is larger than that of all coefficient terms, which indicates that the variation of fitted values is dominantly controlled by the constant rather than others. So the constant term must be removed. The new fitted equation is given below:

$$\mathrm{POR} = -0.119 * \mathrm{AC} + 0.889 * \mathrm{CNL} - 1.431 * \mathrm{DEN} + 1.529 * \mathrm{PE} \tag{31}$$

The analytical results of the multivariate linear fitting model are presented in Fig. 6. In Fig. 6 (a), the residuals distribute randomly on both sides of $y = 0$, manifesting that the residuals are independent for each other. As the distributions of residuals displayed in Fig. 6 (b) and (c) comply with the normal pattern, the fitted equation can be considered that it is correct to use as the fitting model.

4.4. Fitting Correction

In this step, the fitted values will be processed by Eqs. (22) to (29) with the results of the analysis of variance and multiple comparisons. Aiming at the Eqs. (28) to (29), the image method is adopted for recalculating the fitted values. There are two situations probably occur in the processes. The first one is presented in Fig. 7 (a). In this case, as the distance d between the fitted value and defined boundary is less than $X_{\Lambda \mathrm{std}}$, the modified value can be worked out at only once time. The second one is shown as Fig. 7 (b). This figure illustrates that if d is larger than $X_{\Lambda \mathrm{std}}$, the modified value will not be eligible at the first processing time, so the calculation will not stop until the final result meets the requirement of fitting correction.



(a) Distribution of the residuals; (b) Histogram of the residuals; (c) Normal probability distribution of the residuals

Fig.6. Analytical results of multivariate linear fitting model.

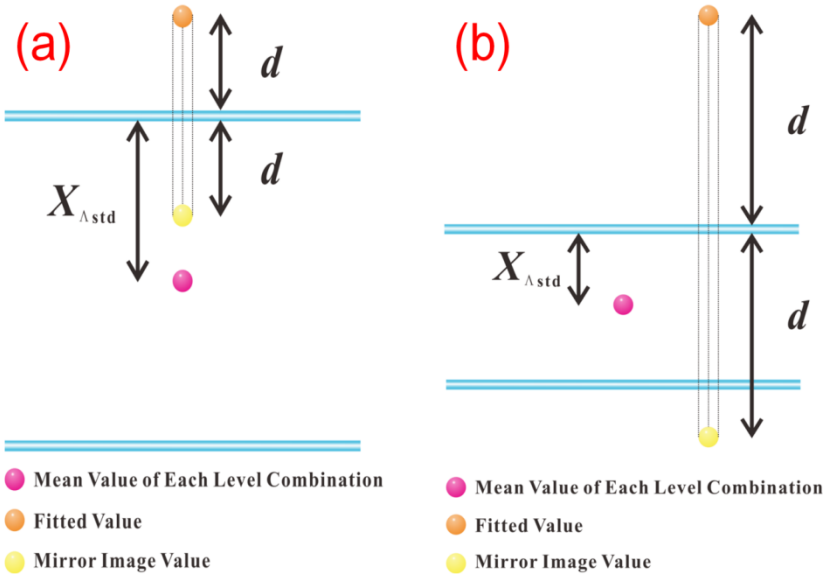
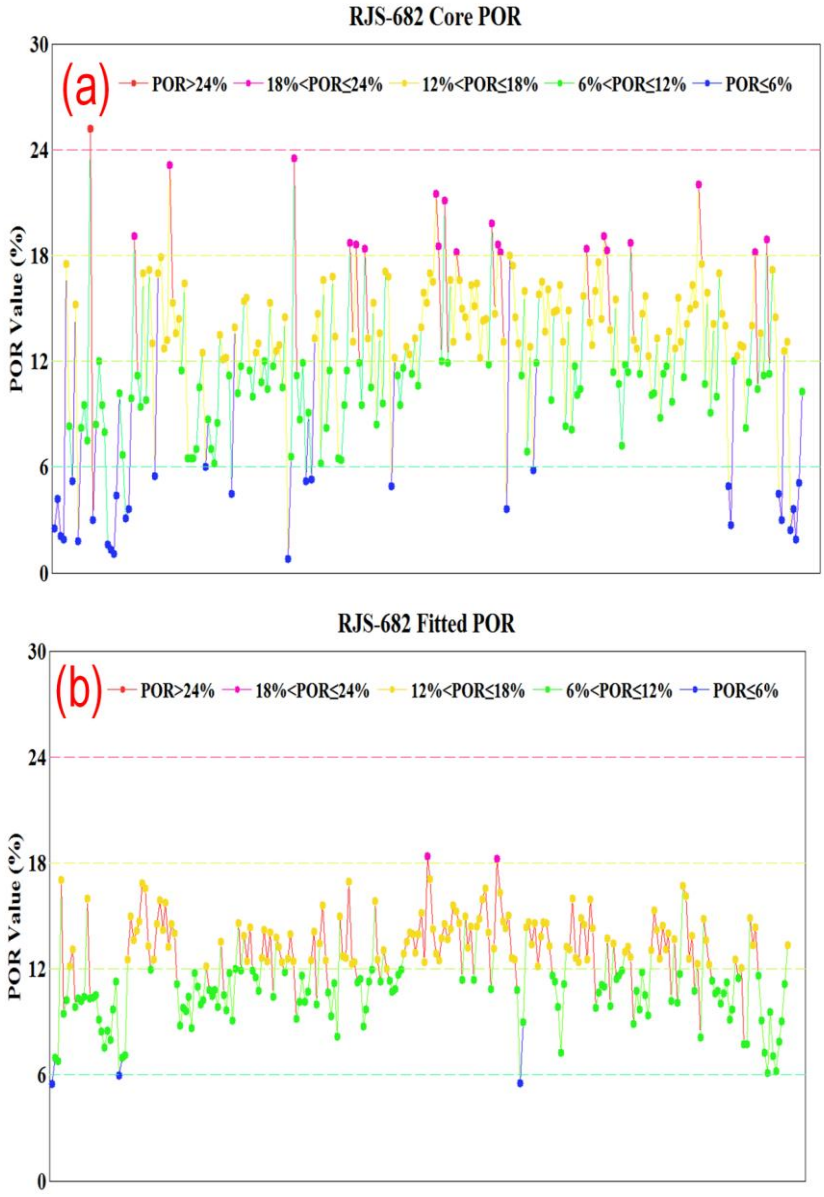


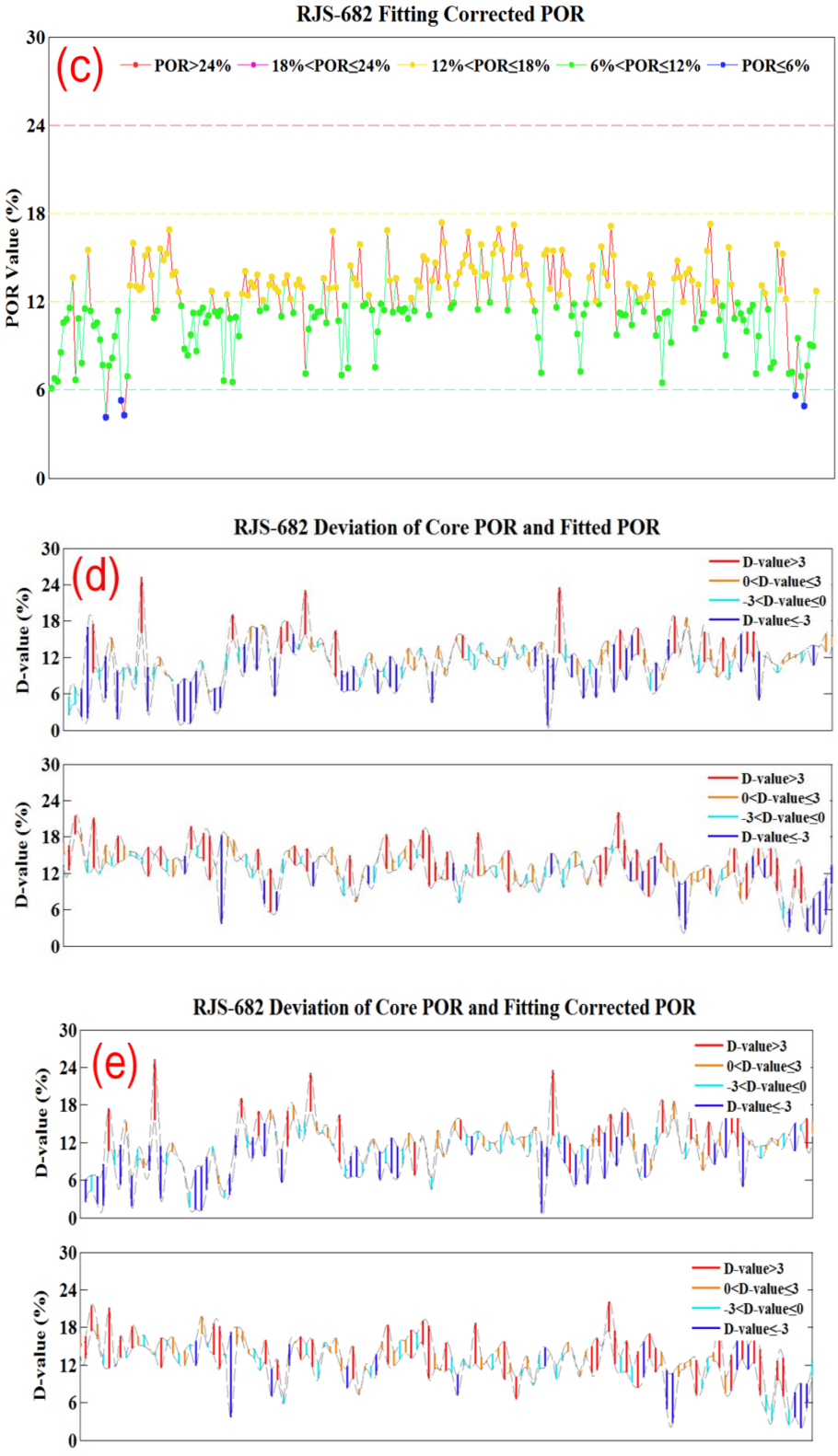
Fig.7. Using image method to correct fitted value in the first (a) or second situation (b).

4.5. Discussion

According to the processes of N-way analysis of variance and fitting correction, the fitted and fitting corrected porosity can be respectively obtained. For other two wells, the prediction was conducted in the same way as the RJS-682 well. The accuracy of all the results are used to verify with the core data. The mean absolute error (MAE) is the indicator to evaluate the accurate degree. Now, the results from five different cases will be discussed below.

The processed results of RJS-682 well are all presented in Fig. 8. Among of them, Fig. 8 (d) and (e) respectively illustrate the deviation of the fitted or corrected porosity compared with the core data. Generally, the fitted values fall within the deviation range between -2 and 2 are reliable and acceptable. But due to only limited data used for prediction, the overfitting problem of the fitting is possible, and thus most of the fitted results does not achieve such accurate degree. Therefore, the commonly used deviation range should be enlarged in order to reasonably weigh the accuracy of the fitted results. In this case, if the value is included in the deviation between -3 and 3, it will be view as the reliable one. In Fig. 8(e), the ratio of deviation between -3 and 3 seems to be larger than that in Fig. 8 (d), so the corrected values, to some degree, can be considered much closer to the core data. Besides, according to the mean absolute error (MAE) of the corrected and fitted values of 3.12 and 3.17 respectively, the accuracy of fitted results are improved by nearly 2% after fitting correction. So based on the analytical results, the creative algorithm of fitting correction really takes effects on the accuracy improvement of the fitted results.





(a) Distribution of core porosity; (b) Distribution of fitted porosity; (c) Distribution of fitting corrected porosity; (d) Deviation of core and fitted porosity; (e) Deviation of core and fitting corrected porosity

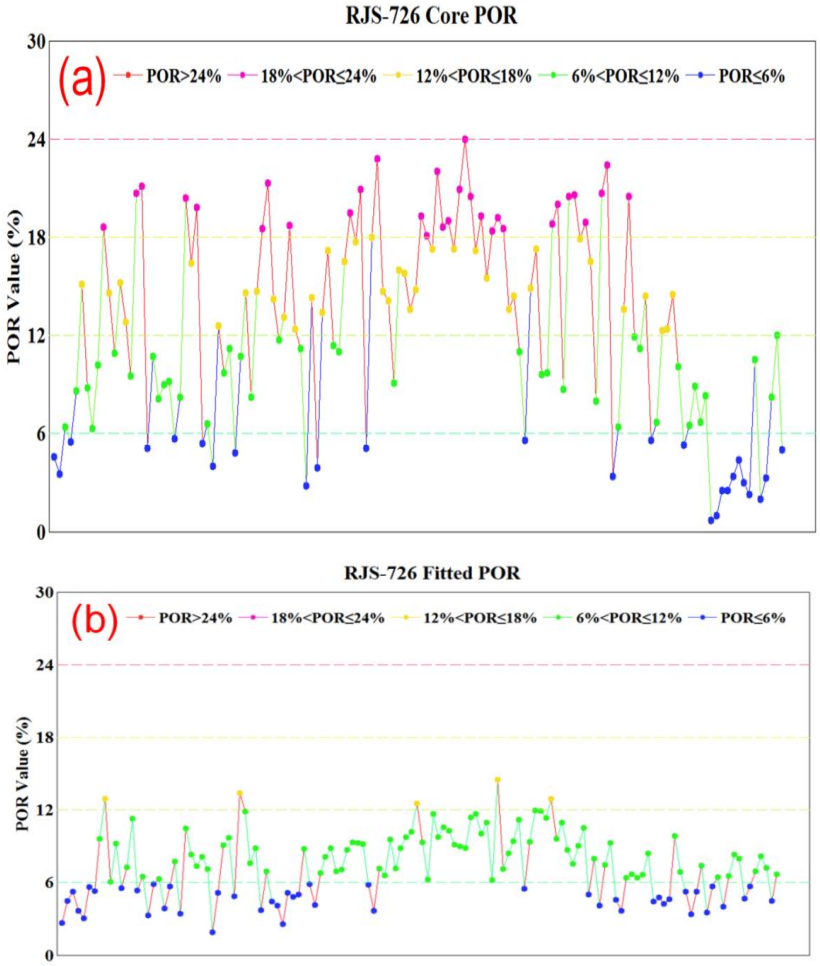
Fig.8. Comprehensive processed results of RJS-682 well.

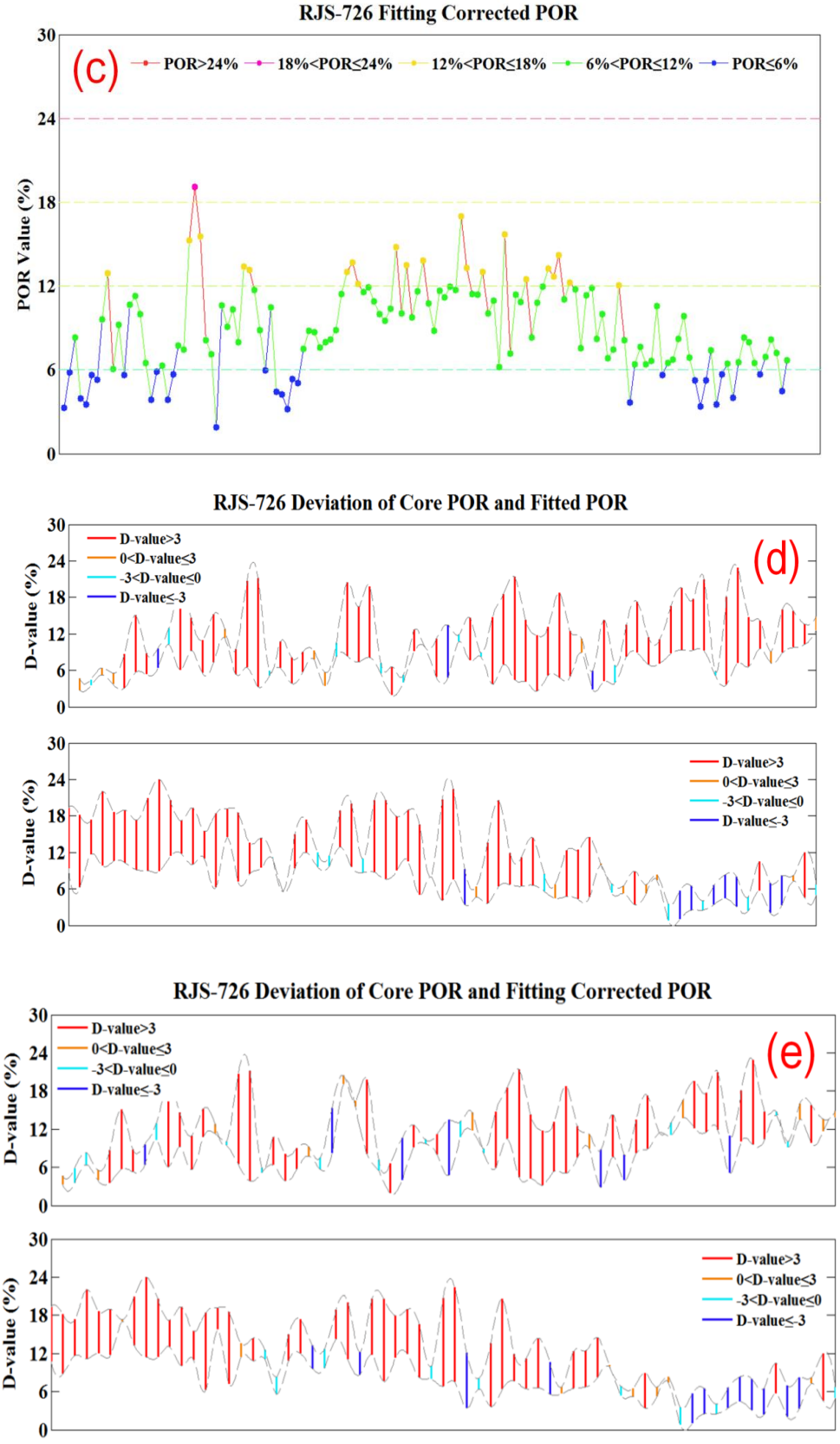
In the porosity prediction of RJS-726 and 656 wells, the raw data will be processed in the same way as the previous case. The comprehensive results are displayed in Figs. 9 and 10. For the RJS-726 well, the ratio of deviation between -3 and 3 in Fig. 9 (e) also seems to be larger than that Fig. 9 (d). Accordingly, the MAE of corrected values 5.43 is less than that of the fitted ones 5.63. As the accuracy raises up by almost 3.6% after fitting correction, the fact that creative

algorithm can properly correct the fitted results is convincingly proved again. For the RJS-656 well, although the deviations of the results presented in Figs. 10(d) and 10(e) are rather similar, the MAE of the corrected porosity 4.71 is unexpectedly larger than that of the fitted ones. The bad effect of fitting correction is mainly caused by the overfitting, or in other words, the used fitted coefficients and the statistical results of "level combination" are not suitable for any cases but only RJS-682 and 726 wells. Improving robustness of the fitted coefficients is normally viewed as the most efficient way to eliminate the overfitting impacts of the fitting models. The process can be operated by adding more logging data as the basic data used for fitting. So in the second verification of RJS-656 well, the compounding data composed of that of the RJS-682 and 726 wells will be applied to establish the new fitting model. The related results are presented in Fig. 11. Compared to Fig. 10 (e), the ratio of deviation between -3 and 3 in the Fig. 11 (b) seems to be larger, and correspondingly the MAE of corrected porosity is reduced to 3.34. Nearly 30% raise of the accuracy of fitted results strongly proves that the combined method can predict porosity in a more accurate way through improving the robustness of the fitted coefficient even though only two wells' data are available.

Aiming at the twice verifications of RJS-656 well, the case of RJS-726 is interpreted again to confirm the practicability of the robustness improvement. The compounding data used for fitting consists of that of the RJS-682 and 656 wells. Fig. 12 displays the distribution and deviation of the corrected results. In Fig. 12 (b), the ratio of deviation between -3 and 3 is undoubtedly larger than that in Fig. 9(e). As the MAE is 4.39, the accuracy is greatly raised by nearly 20% with comparison to that in the first verification. Likewise, the increase of accuracy prove the fact again that under condition of enhancing the robustness of the fitted coefficients, the proposed method can calculate out more accurate porosity values.

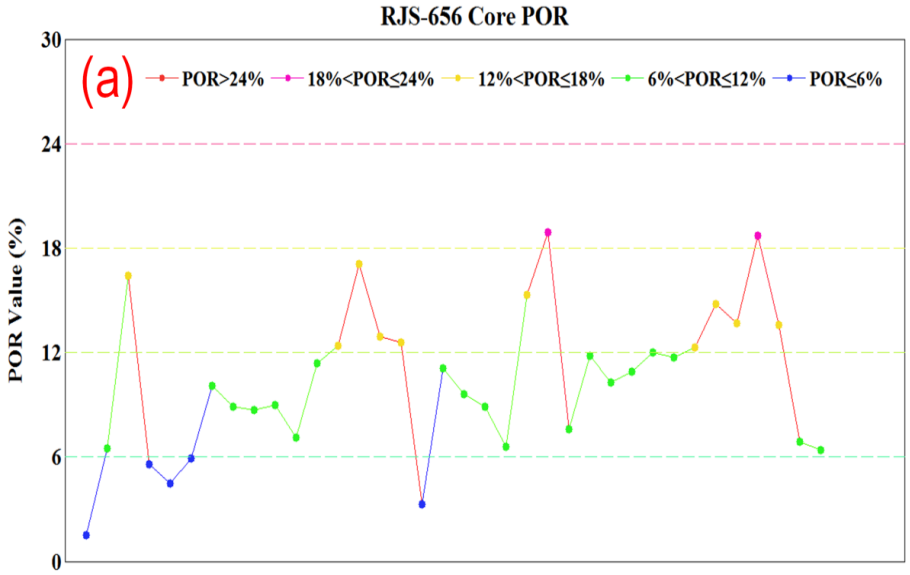
Consequently, the proposed combined method has the capability of predicting porosity for the lacustrine carbonate reservoirs with limited basic data, and after process of fitting correction, the fitted values will be more accurate, generally. But if the overfitting problem of fitting model passively influence the prediction effects of the combined method, enhancing robustness of the fitted coefficients will be better way to improve the accuracy of the fitted results. Predicting porosity with little logging data, so the mean of improving robustness will not be considered firstly unless the fitting corrected results are really unreliable.

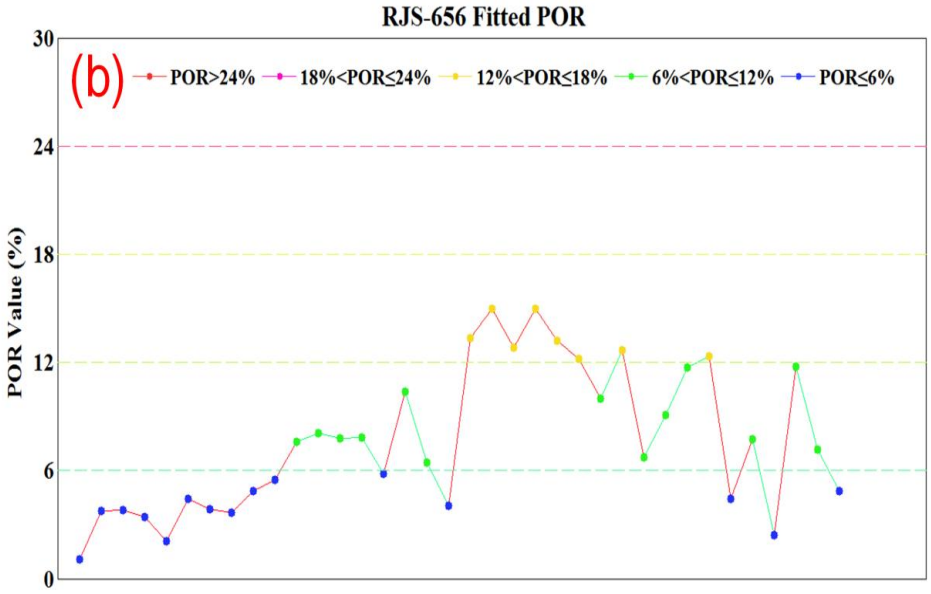


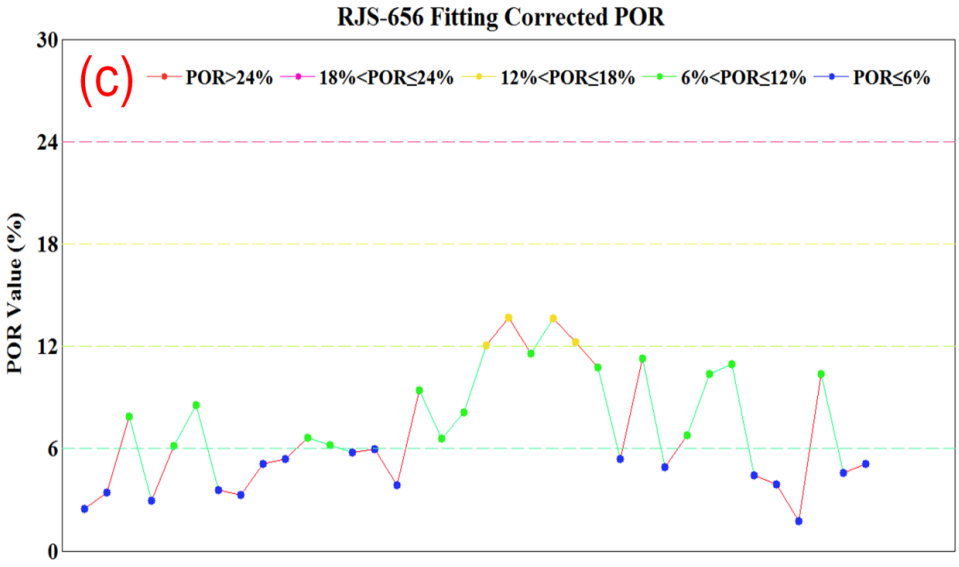


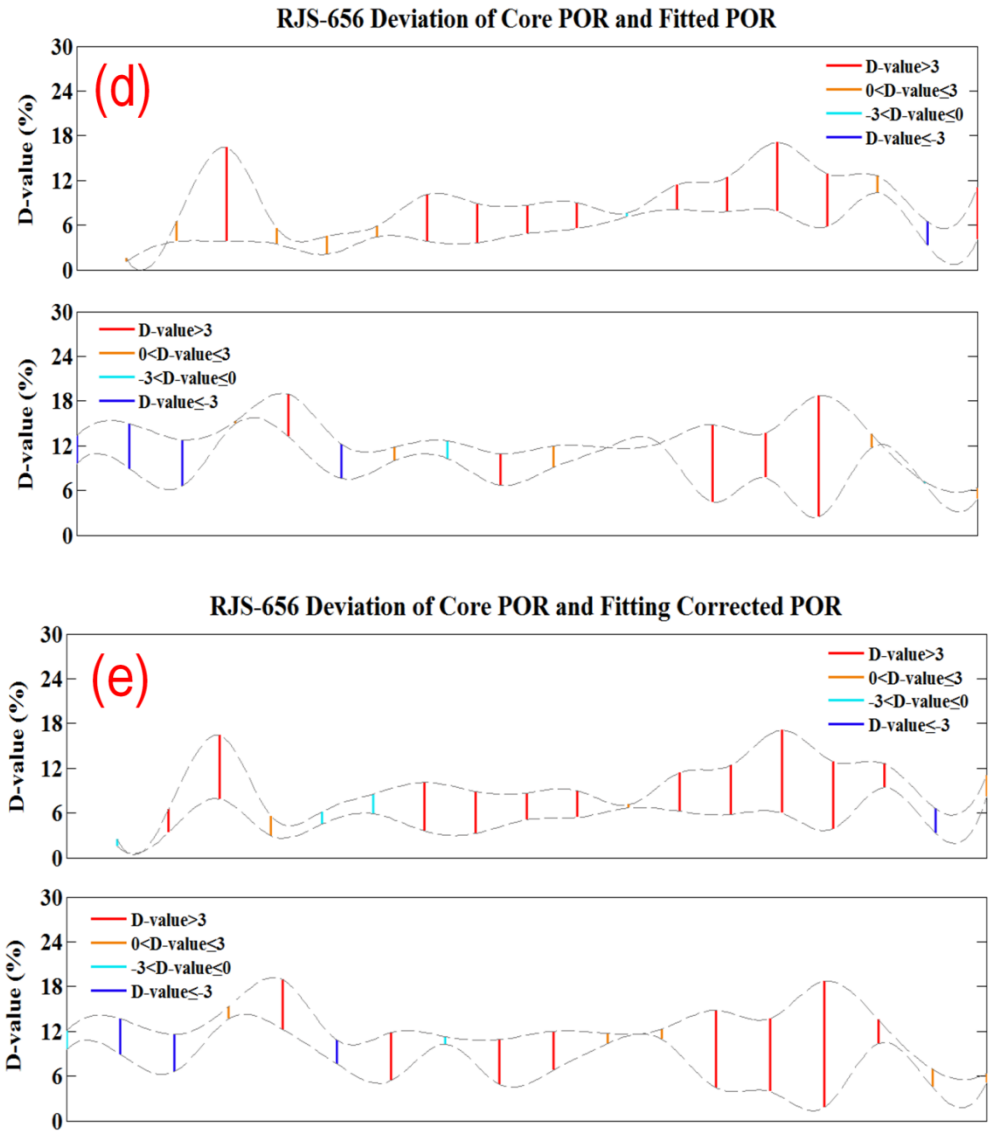
(a) Distribution of core porosity; (b) Distribution of fitted porosity; (c) Distribution of fitting corrected porosity; (d) Deviation of core and fitted porosity; (e) Deviation of core and fitting corrected porosity

Fig. 9. Comprehensive processed results of RJS-726 well.


RJS-656 Core POR
(a)
POR>24%
18%<POR≤24%
12%<POR≤18%
6%<POR≤12%
POR≤6%
POR Value (%)
30
24
18
12
6
0

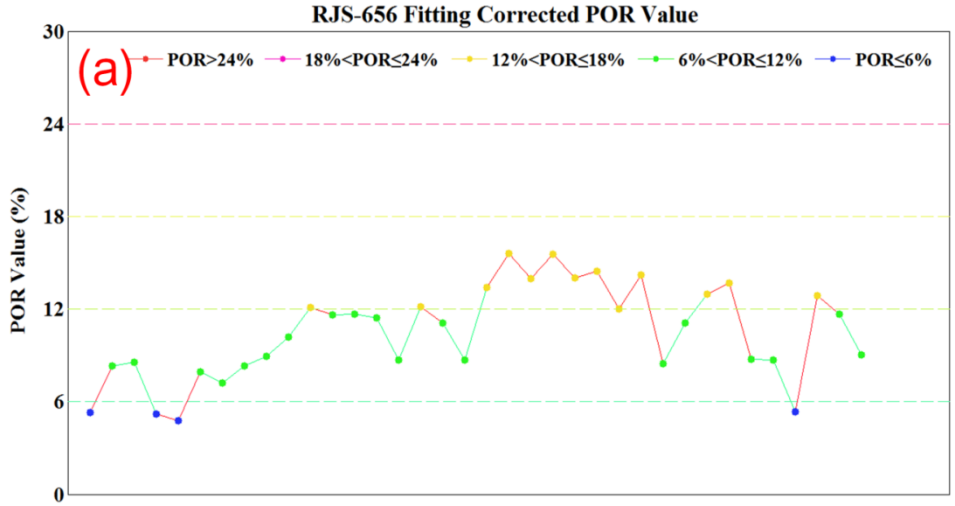

RJS-656 Fitted POR
(b)
POR>24%
18%<POR≤24%
12%<POR≤18%
6%<POR≤12%
POR≤6%
POR Value (%)
30
24
18
12
6
0

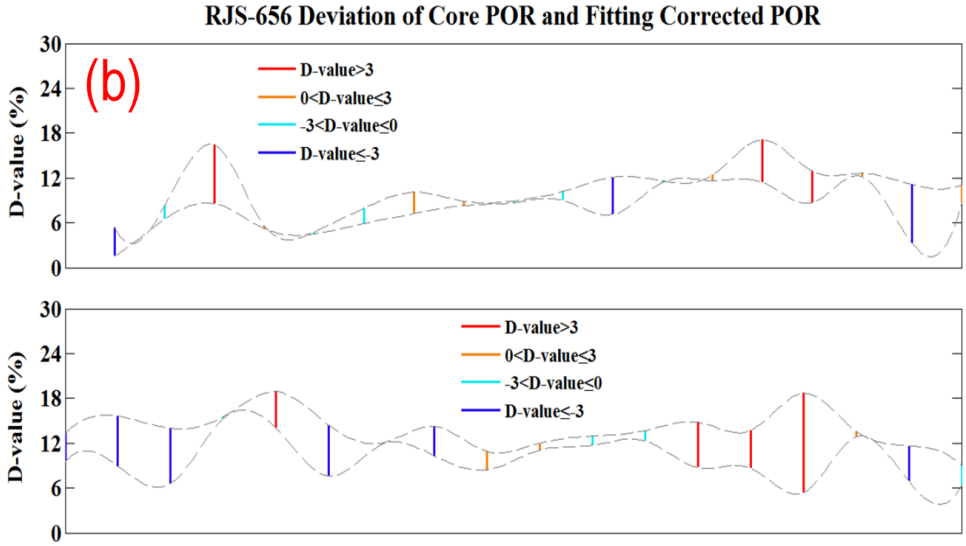

RJS-656 Fitting Corrected POR
(c)
POR>24%
18%<POR≤24%
12%<POR≤18%
6%<POR≤12%
POR≤6%
POR Value (%)
30
24
18
12
6
0



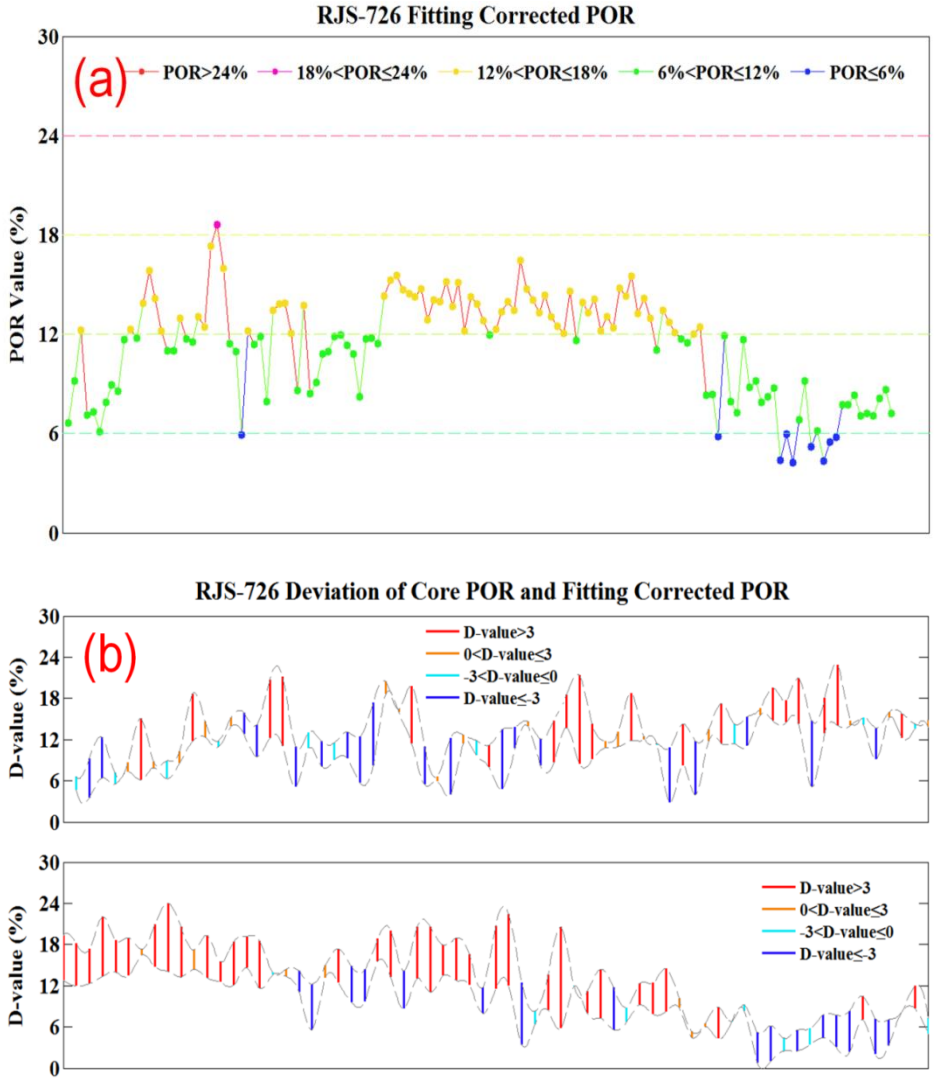
(a) Distribution of core porosity; (b) Distribution of fitted porosity; (c) Distribution of fitting corrected porosity; (d) Deviation of core and fitted porosity; (e) Deviation of core and fitting corrected porosity

Fig.10. Comprehensive processed results of RJS-656 well.





(a) Distribution of fitting corrected porosity; (b) Deviation of core and fitting corrected porosity
Fig.11. Processed results of RJS-656 well.



(a) Distribution of Fitting corrected porosity; (b) Deviation of Core Porosity and Fitting Corrected Porosity

Fig.12. Processed results of RJS-726 well.

IV. CONCLUSION AND RECOMMENDATION

Based on the comprehensive analysis of three wells located in one middle structural high of IARA oil field, the advantages of the proposed combined method can be summarized as below:

(1) The proposed method has the capability of selecting out the significant logging curves for the porosity through N-way analysis of variance. Before operating the calculation, each logging curve should be divided into several levels in accordance with the applied gradation standard. The principle of designing gradation is that both theoretical and calculated freedom degree must be equivalent. Based on the logical relationship between p values and defined confidence level, the significant curves can be found out, and then the raw data is filtered, accordingly.

(2) According to results of multiple comparisons, the fitted values obtained from the fitting model can be accurately corrected by the creative algorithm of fitting correction so as to increase the accuracy of the fitted values. In this way, the corrected porosity will be generally getting much closer to the core data with respect to the fitted ones, which has been proved in the first verification of porosity prediction of RJS-682 well.

(3) Improving robustness of the fitted coefficients can significantly enhance the predictive capability of the proposed combined method. Compared to the fitted results of RJS-656 well in its first process, the accuracy of corrected values obtained from the compounding logging data is substantially raised by nearly 30%. Moreover, the second verification of RJS-726 well also receives the same good effect with accuracy improved by almost 20%.

(4) In contrast with the physical models, the proposed combined method can predict the porosity in a cost-efficient way without aiding any geological parameters, especially in those circumstances of lack of the available logging data.

This study provides a cost-efficient tool for petroleum researchers and engineers to acquire the porosity of carbonate formations. Accordingly, the results are reliable enough to set as the reference data for the related geological basic works. For the next step, more practical methods regarding predicting permeability and pore pressure with limited data will be probed and studied in order to simplify the processing procedures of reservoir evaluation.

ACKNOWLEDGE

This work is supported by the China University of Petroleum-Beijing. We would like to thank the reviewers and the editors for their constructive reviews and suggestions of this paper, all of which result in its material improvement.

REFERENCE

[1]. Asoodeh, M., Bagheripour, P., 2013a. Core porosity estimation through different training approaches for neural network: back-propagation learning vs. genetic algorithm. Int. J. Comput. Appl. 63 (5), 11-15 (0975-8887).

[2]. Bassiouni, Z., 1994. "Theory, Measurement, and Interpretation of Well Logs", First Printing, Henry L. Doherty Memorial Fund of AIME, SPE, Richardson, TX, 372.

[3]. Berger, J.O., Delampady, M., 1987. Testing precise hypotheses. Statistical Science, 2(3), 317-335.

[4]. Berger, J.O., Sellke, T., 1987. Testing a point null hypothesis: The irreconcilability of P value and evidence. Journal of American Statistical Association, 82(397), 112-122.

[5]. Boadu, F.K., 2001. Predicting oil saturation from velocities using petrophysical models and artificial neural networks. J. Pet. Sci. Eng, 30 (3), 143-154.

[6]. Byrnes, A.P., Wilson, M.D., 1991. Aspects of porosity prediction using multivariate linear regression. AAPG Bulletin, 75(3), 7-10.

[7]. Delampady, M., Berger, J., 1987. Lower bounds on posterior probabilities for multinomial and χ^2 test. Technical report 86-37, Dept. Statistics, Purdue Univ.

[8]. Dempster, A.P., 1971. Model searching and estimation in the logic of inference. In Foundations of statistics inference (V.P.Godambe and D.A.Sprott, eds.). Holt, Rinehart, and Winston, Toronto.

[9]. Dickey, J.M., 1971. The weighted likelihood ratio, linear hypotheses on normal location parameters. Ann. Math. Statist. 42 204-223.

[10]. Down, P.A., 1992. A review of recent development in Geostatistics. Computer & Geoscience, 17(10), 1481-1500.

[11]. Ehrenberg, S.N., 1993. Preservation of anomalously high porosity in deeply buried sandstones by grain-coating chlorite: examples from the Norwegian continental shelf. AAPG Bulletin, 77, 1260-1286.

[12]. Emilson, P.L., ALexander, C.V., 2011. 3D porosity prediction from seismic inversion and neural networks. Computers & Geosciences, 37(8), 1174-1180.

[13]. Fainstein, R., Jamieson, G., Hannan, A., Eiles, N., Krueger, A., Schelander, D., 2001. Offshore Brazil Santos Basin exploration potential from recently acquired seismic data. 7th International Congress of the Brazilian Geophysical Society, 52-55.

[14]. Fisher, R.A., 1936. The Use of Multiple Measurements in Taxonomic Problem. Annual of Eugenics, 7, 179-188.

[15]. Gupta, A., Civan, F., 1994. An improved model for laboratory measurement of matrix fracture transfer function parameters in immiscible displacement. Paper SPE 28929 presented at the SPE 69th Annual Technical Conference and Exhibition held in New Orleans, LA USA, Sept, 25-28.

[16]. Hamilton, E.L., Bachman, R.T., Berger, W.H., 1982. Acoustic and related properties of calcareous deep-sea sediments. Journal of Sedimentary Petrology, 52(3), 733-753.

[17]. Han, D.A., Nur, A., Morgan, D., 1986. Effect of porosity and clay content on wave velocity in sandstones. Geophysics, 51(11), 2093-2017.

[18]. He, J., Rui, Z., Ling, K., 2016. A new method to determine Biot's coefficients of Bakken samples. Journal of Natural Gas Science and Engineering, 35, 259-264.

[19]. Helle, H.B., Bhatt, A., Ursin, B., 2001. Porosity and permeability prediction from wireline logs using artificial neural networks: a North Sea case study. Geophys. Prospect, 49, 431-444.
[20]. Jamialahmadi, M., Javadpour, F.G., 2000. Relationship of permeability, porosity and depth using an artificial neural network. J. Pet. Sci. Eng, 26, 235-239.
[21]. Jurgen, A., Zhi, Y.G., Marianela, S., 2014. Salt-Structural styles and kinematic evolution of the Jequitinhonha deepwater fold belt, central Brazil passive margin. Marine and Petroleum Geology, 37, 101-120.
[22]. Kahraman, S., Yeken, T., 2008. Determination of physical properties of carbonate rocks from P-wave velocity. Bulletin of Engineer Geology and the Environment, 67(2), 277-281.
[23]. Keys, R.G., Xu, S.Y., 2002. An approximation for the Xu-White velocity model. Geophysics, 67(5), 1406-1414.
[24]. Lindley, D.V., 1957. A statistical paradox. Biometrika, 44(1-2), 187-192.
[25]. Moczydlower, B., Salomao, M.C., Romeu, R.K., 2012. Development of the Brazilian Pre-Salt Fields-When to Pay Information and When to Pay Flexibility. SPE 152860 presented at the 2012 SPE Latin American and Caribbean Petroleum Engineering Conference held in Mexico City, Mexico, 16-18 April.
[26]. Montaron, B., Tapponier, P., 2010. A quantitative model for salt deposition in actively spreading basins. American Association of Petroleum Geologists, Search and Discovery, Article 30117.
[27]. Mohriak, W.U., 2001. Salt tectonics, volcanic centers, fracture zones and their relationship with the origin and evolution of the South Atlantic Ocean: geophysical evidence in the Brazilian and West African margins. 7th International Congress of the Brazilian Geophysical Society, Salvador, Expanded Abstracts, 1594.
[28]. Mohriak, W.U., 2005. Salt tectonics in Atlantic-type sedimentary basins: Brazilian and West African perspectives applied to the North Atlantic Margin. In: Post, P.J., Rosen, N.C., Olson, D.L., Palmes, S.L., Lyons, K.T., Newton, G.B., (Eds.). GCSSEPM 25th Annual Bob F. Perkins Research Conference, Petroleum Systems of Divergent Continental Margin Basins, 375-413.
[29]. Mohriak, W.U., Nemcok, M., Enciso, G., 2008a. South Atlantic divergent margin evolution: rift-border uplift and salt tectonics in the basins of SE Brazil. In: Pankhurst, R.J., Trouw, R.A.J., Brito Neves, B.B., de Wit, M.J., (Eds.). West Gondwana Pre-Cenozoic Correlations Across the South Atlantic Region. Geological Society, London, Special Publications, 294, 365-398.
[30]. Moraes, M.A.S., De, Ros, L.F., 1990. Infiltrated clays in fluvial Jurassic sandstones of Rectncavo basin, northeastern Brazil. Journal of Sedimentary Petrology, 60, 809-819.
[31]. Nahser, M.A., Wang, Y.H., 2011. Porosity prediction using the group method of data handling. Geophysics, 76(5), O15-O22.
[32]. Ou, C., Rui, R., Li, C., Yong, H., 2016. Multi-index and two-level evaluation of shale gas reserve quality. Journal of Natural Gas Science and Engineering, 35, 1139-1145.
[33]. Quirk, D.G., Hirsch, K.K., Hsu, D., Von Nicolai, C., Ings, S. J., Lassen, B., Schoedt, N.H., 2012. Salt tectonics on passive margins: examples from Santos, Campos and Kwanza basins. In: Alsop, G.I., Archer, S.G., Hartley, A.J., Grant, N.T., Hodgkinson, R., (Eds.). Salt Tectonics, Sediments and Prospectivity. Geological Society, London, Special Publications, 363, 207-244.
[34]. Raymer, L.L., Hunt, E.R., Gardner, J.S., 1980. An improved sonic transit time-to-porosity transforms. Society of Petrophysicist and Well-Log Analysts, 21, 1-13.
[35]. Rui, Z., Metz, P. A., Reynolds, D., Chen, G., Zhou, X., 2011a. Historical pipeline construction cost analysis. International journal of oil, gas and Coal Technology, 4(3), 244-263.
[36]. Rui, Z., Metz, P. A., Reynolds, D., Chen, G., Zhou, X., 2011b. Regression models estimate pipeline construction costs. Oil & Gas Journal, 109(14), 120-127.
[37]. Rui, Z., Metz, P. A., Chen, G., 2012a. An analysis of inaccuracy in pipeline construction cost estimation. International journal of oil, gas and Coal Technology, 5(1), 29-46.
[38]. Rui, Z., Metz, P. A., Chen, G., Zhou, X., Wang, X., 2012b. Regressions allow development of compressor cost estimation models. Oil & Gas Journal, 110(1a), 110-115.
[39]. Rui, Z., 2012c. Study confirms feasibility of in-state Alaska gas pipeline. Oil & Gas Journal, 110(10), 112-120.
[40]. Rui, Z., 2012d. Study finds in-state gas pipeline feasibility. Oil & Gas Journal, 110(11), 124-130.
[41]. Rui, Z., Metz, P.A., Chen, G., Zhou, X., 2013a. Inaccuracy in pipeline-compressor-station construction cost estimation. Oil and Gas Facilities, 2(5), 71-79.
[42]. Rui, Z., Wang, X., 2013b. A comprehensive analysis of natural gas distribution pipeline incidents. International journal of oil, gas and Coal Technology, 6(5), 528-548.
[43]. Ren, Z., Wu., X., Liu, D., Rui, R., Guo, Wei., Chen, Z., 2016. Semi-analytical model of the transient pressure behavior of complex fracture networks in tight oil reservoirs. Journal of Natural Gas Science and Engineering, 35(a), 497-508.
[44]. Sun, J., Gamboa, E., Schechter, D., Rui, Z., 2016. An intergrated workflow for characterization and simulation of complex fracture networks utilizing microseismic and horizontal core data. Journal of Natural Gas Science and Engineering, 34, 1347-1360.
[45]. Wyllie, M.R.J., Gregory, A.R., Gardner, L.W., 1956. Elastic wave velocity in heterogeneous and porous media. Geophysics, 21(1), 41-70.
[46]. Xu, S.Y., White, R.E., 1995. A new velocity model for clay-sand mixtures. Geophysical prospecting, 43(1), 91-118.
[47]. Yi, K.N., Cheng X.J., 2013. A Frequently Mistaken Concept in "Hypothesis Test": A Dispute Triggered by China Undergraduate Mathematical Contest in Modeling. Journal of Chongqing University of Technology: Natural Science, 27(4), 106-109.
[48]. Zhao, X., Rui, Z., Liao, X., Zhang, R. 2015a. The qualitative and quantitative fracture evaluation methodology in shale gas reservoir. Journal of Natural Gas Science and Engineering, 27, 486-495.

[49]. Zhao, X., Rui, Z., Liao, X., Zhang, R. 2015b. A simulation method for modified isochronal well testing to determine shale gas well productivity. Journal of Natural Gas Science and Engineering, 27(2), 479-485.
[50]. Zhao, X., Rui, Z., Liao, X., 2016. Case studies on the CO_2 storage and EOR in heterogeneous, highly water-saturated, and extra-low permeability Chinese reservoirs. Journal of Natural Gas Science and Engineering, 29, 275-283.
[51]. Zhao, Y., Rui, Z., 2014. Pipeline compressor station construction cost analysis. International Journal of Oil, Gas and Coal Technology, 8(1), 41-61.

Yufeng Gua, . "Quick porosity prediction for carbonate reservoirs using modified N-Way analysis of variance." International Refereed Journal of Engineering and Science (IRJES), vol. 07, no. 01, 2018, pp. 73–95.