Applied Single Correlation Analysis and Multiple Regression Analysis for Rapid Prediction of Reservoir Sensitivity

Qu Lei

The First oil production plant of Daqing oilfield, Heilongjiang, Daqing, 163318, China

qulei0527@163.com

Keywords: reservoir protection, sensitivity analysis, single correlation analysis, multiple regression, core analysis

Abstract. Sensitivity study, the premise of the analysis of reservoir damage mechanism, is significant to optimize all the links of drilling and development process, as well as to develop the systematic oil and gas reservoir protection technology scheme. To study a variety of methods developed in recent years for the reservoir sensitivity prediction, it is found that using the combination of single correlation analysis and multiple regression analysis is a new, ideal, and rapid method for the prediction of reservoir sensitivity. Extract the relevant information of sensitivity on the basis of conventional core analysis and sensitivity analysis of minerals, and apply the single correlation analysis and multiple regression analysis to predict the reservoir sensitivity, the accuracy rate can reach 85%, which basically meets the needs of reservoir sensitivity prediction. Compared with the single prediction method, the prediction results obtaining from the combination of the two methods are obviously improved. It is simple, highly applicable, and has a clear physical meaning.

The most common method of reservoir sensitivity prediction is sensitivity experiment. Conventional reservoir sensitivity experiment method requires a large number of cores, and the test is a long term procedure, while, during the process of drilling and cutting cores as well as pretreatment for the experiment, different degrees of damage may have been caused to cores. Therefore, conventional evaluation method needs plenty of manpower and material resources, it cannot make sure that the predicting outcome of each time is exactly correct however. Various methods for the fast prediction of reservoir sensitivity have been developed in recent years, such as Elman neural network, BP neural network, grey theory appraising method, multiple discriminate analysis, fuzzy mathematical method and so on, while, there exists defect in each method[1].

Elman network can be regarded as forward neural network which possesses partial memory unit and feedback connection. Its major structure is feedback connection, which is used to memorize the output value of the previous moment, and its convergence speed is slow due to the fixed connection weight. It usually adopts BP neural network model in reservoir sensitivity prediction[2-3]. However, there exists disadvantages like low convergence speed and failure in global optimization in the conventional BT algorithm towards BP algorithm. Neural network method changes the characteristics of all problems into figures, inference into numerical calculation, while neural network fails to carry out work when the data is insufficient. Besides, it requires corresponding complicated computer program when taking use of neural network method, thus the widespread use of this method has been limited.

Grey theory is a new approach aims at studying the uncertainty of less data and poor information, which owns the advantage of being required with less data and small amount of calculation, thus it is extensively used in some certain industries[4-5]. Within the grey model, the plane nipped between the upper and lower bounds of the forecasted future value is called gray plane, which is unfold in a shape of trumpet, that is to say, the farther in the future time, the larger grey interval of prediction value, in other words, the bigger the grey scale, the smaller the practical significance of the prediction value.

Multiple discriminate analysis first divides sensitivity and the known samples of fabric parameter into several groups based on the size of sensitivity, respectively establishing the discriminant functional equation of each group of sample[6]. It is impossible to grouping in detail, thus the predicting outcome is only an interval, but not a specific numerical value.

Based on the current method of fast prediction for reservoir sensitivity, this article combines with the previous research results, synthesizes the applied single correlation analysis and multiple regressions analysis to forecast reservoir sensitivity, getting over the disadvantages of other methods in forecasting reservoir sensitivity, supplementing the existed defects.

1. SINGLE CORRECTION ANALYSIS

When one or several interrelated variables are taken certain values, although the precise value of its another corresponding variable cannot be confirmed, it still changes within a certain rage in accordance with certain rules. The dependency relation objectively exists among undemanding and uncertain quantities of such phenomenon is known as correlativity[7].

Back in 1980, British statistician Pearson had put forward a calculation formula of measuring the linear relationship between two variables, see formula (1), r usually acts as product moment correlation coefficient. The value range of correlation coefficient r is from -1 to 1; the size of |r| reveals the strong and weak point of the linear relationship between variable x and y[8-9]. The correlation coefficient value significance and correlative degree evaluation standard see table 1 and table 2.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

Table 1 Significance of correlation coefficient value and other values

Value of |r|	0	1			0<|r|≦1
		-1	1	-1≦r<0	0<|r|≦1
Significance	No linear correlation	Perfect negative correlation	Perfect positive correlation	Negative correlation	Positive correlation

Table 2 Correlative degree evaluation standard

Value of |r|	0~0.3	0.3~0.5	0.5~0.8	0.8~1.0
Degree of correlation	Weak correlation	Low correlation	Significant correlation	High correlation

2. MULTIPLE REGRESSION ANALYSIS

Multiple regression analysis measures the common relation of the changes in the quantity of 2 or more than 2 variables which has correlativity on the basis of the current correlation analysis, it focuses on studying the interrelation among variables, as well as establishing empirical formula among variables, so as to achieve the goal of predicting or controlling[10].

2.1 Multiple linear regression model

Multiple regression means regressing the most appropriate function from large amounts of data: $y = f(x_1, x_2, \cdots x_p)$. The selectable range of function f is broad, thus it is hard to make the choice, generally, we suppose f is linear function

$$f(x_1, x_2, \cdots x_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \quad (2)$$

Thereinto, β_0, β_1, ... β_p are all unknown coefficient, then the formula $y=\beta_0+\beta_1 x_1+...+\beta_p x_p+e$ is called multiple linear regression model[11]. Without loss of generality, the total can be set as $\beta_0 = 0$, because the independent variable $x_0=1$ can be brought formally, thus we can conclude that

$$y = \sum_{i=0}^{p} \beta_i x_i + e \qquad (3)$$

Thereinto: y_i is the observed value (i=1, 2,..., n), n>p) of y, when $x_1=x_{i1}$, $x_2=x_{i2}$, ..., $x_p=x_{ip}$.

2.2 Multiple linear regression coefficient

If $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = (x_{ij})_{n\times p}, e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}$

Then the model can be: $Y = X\beta + e$ (4)

Thereinto, e acts as n-dimention random vector, X is n×p constant matrix, and β is p-dimention unknown parameter vector. Evaluating regression coefficient is to work out β to make $Q(\beta) = \sum_{i=1}^{n} [y_i - (\beta_1 x_{i1} + \beta_3 x_{i2} + \cdots + \beta_p x_{ip})]^2$ reach the minimum. It can be concluded from derivation and proof that when X is under full value, the regression coefficient that meets the requirement is as follows[11]: $\beta = (X'X)^{-1} X'Y$ (5)

3. THE COMBINED FORECASTING RESERVOIR SENSITIVITY

Correlation analysis means to the intimate level of interdependent relations among phenomena by taking use of index. Regression analysis refers to the statistic analysis method that choosing an appropriate teaching model towards the phenomenon that owns correlativity on the basis of the specific forms of the correlativity, to approximately express the average changing relationship among variables.

Correlation analysis is required to depend on aggression analysis to indicate the specific form related with the number of phenomena, while aggression analysis needs to depend correlation analysis to show correlative degree of the changes towards the the number of phenomena. Only when there exists high correlation between variables, can regression analysis for seeking its relevant specific form be meaningful. Correlation analysis is the foundation and precondition of regression analysis, and regression analysis is continuation and deepening of correlation analysis.

Single correlation coefficient method determines the correlative degree between fabric parameter and sensitivity index, representing the physical significance of each fabric parameter, but it cannot solve the situation of complex influence factors (more factors or synergistic effect). While multiple linear regression analysis method is a better way to solve the problem of joint influence caused by factors, it cannot distinguish the contribution value of each fabric parameter on sensitivity index however. Besides, there exists some more influence factors in reservoir sensitivity, thus it is hard to confirm the influence degree of a certain factor on the study. At this moment, the application of linear regression to studying problems becomes especially complex, and the results obtained are inaccuracy in most cases. While this problem can be solved by combining the two methods. First, work out the single correlation coefficient between various sensitivities and the influence factors of the area by taking use of single correlation coefficient method, and find out the main influence factor of this area's sensitivity, and then, determine regression formula with linear regression method, namely the sensitivity assessment model that is suitable for the area. As a result, various sensitivity degree can be forecasted through this model[12].

4. FIELD APPLICATION

According to the viewpoint of materialist dialectics, the external manifestation of substance is a reflection of its internal property, besides, there exists a certain relation among the properties of substance. If we regard reservoir sensitivity as the external manifestation of the internal properties of reservoir, then reservoir sensitivity must be related with the mineral property and pore property of reservoir rocks as well as reservoir fluid property[13]. Therefore, it needs to take use of single correlation analysis method to study the relation between the internal property of reservoir and sensitivity first, finding out the major internal factor that influences reservoir sensitivity.

4.1 Determination towards the influence factor of sensitivity

4.1.1 Influence factor: there exist various influence factors for sensibility damage, and their interaction is complicated. It can be concluded from the analysis on sensitivity characteristics of Hailaer Oilfield and experimental data that reservoir sensitivity is closely related with reservoir physical performance, rock structure, pore structure, mineral composition of rocks, as well as formation fluid. The determined influence factors of various sensitivity damages by searching large number of relevant literature materials see table 3.

The content of clay minerals is an important index for studying reservoir damage, but influence caused by the content of clay minerals on sensitivity does not all depend on the amount of clay content, but has a lot to do with the varieties and attitude of clay. In addition, it is also related with factors like cementation type, particle classification, types of formation water and etc. However, all the factors cannot be quantified, these factors are not taken into consideration in multiple linear regression.

Table 3 sensitivity influence factors

Sensitivity types	Influence factors
Water sensitivity	Contents of montmorillonite, illite, kaolinite, illite smectite mixed layer, green smectite mixed layer, as well as degradative chlorite, formation water salinity, formation water pH, gasometry permeability, median grain diameter, carbonate content and shale content.
Salt sensitivity	Contents of kaolinite, illite, chlorite, montmorillonite, illite smectite mixed layer, hydrated mirrorstone, as well as formation water salinity, formation water pH, gasometry permeability, median grain diameter
Alkali sensitivity	Contents of kaolinite, illite, chlorite, montmorillonite, (potassium sodium) feldspar, microcrystalline quartz, and opal, as well as various clay minerals, formation water salinity, formation water pH, gasometry permeability, median grain diameter, gypsum content and calcium ion content
Hydrochloric acid sensitivity	Contents of chlorite, ferrocalcite, ankerite, hematite, hydrobiotite, green smectite mixed layer as well as pyrite, glauconite, gasometry permeability
Hydrofluoric acid sensitivity	Contents of calcite, limestone, dolomite, anorthite, zeolite, as well as zeolite, various clay minerals, gasometry permeability
Velocity sensitivity	Contents of kaolinite, illite, microcrystal muscovite, microcrystalline quartz, as well as feldspar, formation water pH, gasometry permeability, median grain diameter, shale content and carbonate content.

4.1.2 Evaluate single correlation coefficient. We have sorted materials like sensitivity of Hailaer Oilfield, clay minerals, formation water analysis, physical performance, slice and stuff by levels, and have calculated the single correlation coefficient between reservoir sensitivity and its influence factors in accordance with formula (1), the result see table 4. The value of correlation coefficient in the table reveals the of correlative degree of the major sensitivity influence factors with various sensitivities, namely it reflects the intensity level of influence on various sensitivities.

Table 4 The single correlation coefficient for the main influence of various sensitivities

Influence factors	Water sensitivity	Velocity sensitivity	Salt sensitivity	Acid sensitivity	Alkali sensitivity
Montmorillonite	0.7338	0.7653	0.6601	0.5213	0.7042
Illite	0.5579	0.4136	0.3928	0.6532	0.3181
Kaolinite	0.0945	0.5106	0.3989	0.0728	0.2354
Chlorite	0.3813	0.4031	0.5615	0.7781	0.6605
Illite smectite mixed layer	0.6846	0.8061	0.6184	0.6489	0.8089
Total amount of clay	0.7001	0.4878	0.5082	0.7163	0.7408
Quartz skeleton	-0.1234	-0.3602	-0.2023	-0.3209	-0.6352
Feldspar framework	-0.4961	-0.0917	-0.0167	-0.0134	-0.4457
Carbonate	0.5012	0.2583	0.4047	0.0761	0.6029
Physical performance index	-0.2931	-0.2236	-0.4298	-0.6104	-0.0757
Gasometry permeability	0.6113	0.7045	0.7599	0.5784	0.5682
Formation water salinity	0.3251	0.1523	0.1327	0.5354	0.2158
Formation water pH	0.2103	0.1324	0.0154	0.3212	0.7861
Shale content	0.5873	0.6224	0.2258	0.4431	0.3985
median grain diameter	0.3465	0.5246	0.4912	0.6348	0.5837

Table 4 reveals the correlation value between fabric parameter and sensitivity. Select out factors that have a strong relationship with sensitivity, namely finding out influence factors, of which the correlation coefficient is >0.5, conducting linear regression predicting sensitivity.

4.2 Linear regression predicting sensitivity

Based on the water sensitivity prediction of Hailaer area, carrying out sensitivity prediction analysis. It can be seen from table 4 that there are seven factors significantly or highly correlated with (correlation coefficient is >0.5) the water sensitivity prediction of Hailaer area, they are montmorillonite content , illite content, illite smectite mixed layer content, total amount of clay, carbonate content, gasometry permeability and shale content. Establish regression model to predict water sensitivity by taking the seven factors as dependent variable.

4.2.1 Collecting data. Consulting relevant literature and collect information that linear regression requires based on the determined main influence factors. Thereinto, gasometry permeability is the rock permeability tested in the air, and the content of other factors means absolutely hundred percent of the content, which is the percentage content for measuring the weight of clay minerals in pores. Because there exists less of measured data of carbonate content and shale content, we have referred to other corresponding data of co-well or data that is close to the same position of the well when searching the information of the two factors, besides, we have obtained carbonate content and shale content at the same position of after calculation, conducting linear regression by taking use of the collected factors and data of water sensitivity, establishing water sensitivity prediction model.

4.2.2 Establish sensitivity prediction model. Based on the concept of multiple linear regression, establish the following model:

$$\text{Water sensitivity index}=f\,(x_1, x_2, x_3, x_4, x_5, x_6, x_7) \qquad (6)$$

In the formula, independent variable x_1, x_2, x_3, x_4, x_5, x_6, x_7 respectively are montmorillonite content, illite content, illite smectite mixed layer content, total amount of clay, carbonate content, gasometry permeability and shale content, and the observed value is water sensitivity index. Independent variable X and observed value Y matrix can be obtained from multiple linear regression model. Therefore, the calculated coefficient of the regression model through regression coefficient formula is as follows: $\beta = (X'X)^{-1}X'Y = (-0.0286, 0.0155, -0.0186, 0.0128, -0.00103, 0.0218, 0.1473)'$

Substitute regression coefficient into formula (6), we can obtain the water sensitivity prediction model of Hailaer area:

$$D_{water} = -0.0286x_1 + 0.0155x_2 - 0.0186x_3 + 0.0128x_4 - 0.00103x_5 + 0.0218x_6 + 0.1473x_7 \quad (7)$$

4.2.3 Analysis of water sensitivity prediction result. We have conducted water sensitivity analysis on the known sample of reservoir sensitivity of Hailaer area by taking use of the obtained water sensitivity prediction model, the comparison of the measured result and predicting result of all sensitivities see table 5.

It can be seen that 15 samples in Hailaer area have been forecasted comprehensively applying single correlation analysis and multiple regression analysis, 13 samples are under correct prediction thereinto, 2 samples are under prediction error, the coincidence rate is approximately 87%.

Besides, we have predicted the salt sensitivity, alkali sensitivity, velocity sensitivity and acid sensitivity with the same method, and the predicting result is as follows: medium to strong salt sensitivity, moderate alkali sensitivity, moderate velocity sensitivity, no acid sensitivity, the accuracy respectively are 87%, 92%, 80% and 80%, and the average accuracy reaches 85%, which is higher than the accuracy of independent application of single correlation coefficient or multiple regression analysis, and has satisfied the requirement of field application. It is approved to be an ideal new method of predicting reservoir sensitivity.

Table 5 Analysis result of sensitivity prediction

Rock sample No.	Measured/predicted water sensitivity index	Measured/predicted salt sensitivity index	Measured/predicted alkali sensitivity index	Measured/predicted velocity sensitivity index	Measured/predicted acid sensitivity index
1	0.26/0.23	0.64/0.56	0.36/0.56	0.46/0.42	0.14/0.22
2	0.26/.017	0.57/0.67	0.64/0.53	0.50/0.47	0.10/0.17
3	0.25/0.25	0.55/0.37	0.58/0.61	0.45/0.27	0.05/0.07
4	0.31/0.27	0.62/0.54	0.43/0.52	0.51/0.47	0.11/0.06
5	0.21/0.25	0.62/0.45	0.39/0.46	0.54/0.45	0.12/0.16
6	0.23/0.21	0.58/0.52	0.56/0.45	0.43/0.50	0.08/0.23
7	0.22/0.16	0.64/0.56	0.51/0.48	0.53/0.46	0.07/0.06
8	0.2/0.19	0.70/0.64	0.61/0.53	0.57/0.73	0.05/0.12
9	0.26/0.16	0.56/0.53	0.42/0.51	0.62/0.56	0.11/0.15
10	0.21/0.2	0.57/0.65	0.37/0.45	0.51/0.62	0.08/0.06
11	0.16/0.18	0.58/0.52	0.38/0.27	0.48/0.58	0.14/0.08
12	0.32/0.25	0.55/0.66	0.45/0.60	0.45/0.56	0.12/0.11
13	0.39/0.32	0.64/0.72	0.56/0.37	0.43/0.22	0.09/0.26
14	0.11/0.2	0.52/0.60	0.46/0.56	0.62/0.54	0.12/0.14
15	0.37/0.34	0.57/0.66	0.55/0.62	0.47/0.36	0.07/0.08
Accuracy	87%	87%	92%	80%	80%

5. CONCLUSION

(1) There exist disadvantages in both single correlation coefficient method and multiple regression analysis when forecasting sensitivity, but they can complement each other, conducting sensitivity prediction by combining the two methods contribute to get over the defects, so as to improve the accuracy of sensitivity prediction.

(2) Analyze the influence factors of reservoir sensitivity in Hailaer area, and the result shows that the main influence factors contain the following ten parameters: montmorillonite, illite, chlorite, illite smectite mixed layer, quartz, volume fraction of feldspar, as well as the total amount of clay, shale content, physical index, and gasometry permeability.

(3) It requires large amounts of data when predicting reservoir sensitivity by comprehensively applying single correlation coefficient method and multiple regression analysis, comprehensiveness and representativeness of data collection are the key points of working put the scientific predicting result.

(4) The accuracy of prediction methods in this article is an average of 85%, with the highest accuracy of 92%, which indicates that the reliability of prediction methods is higher, and it reveals the potential degree of reservoir sensitivity quantitatively, providing reliable basis for enacting technical measures to protect oil-gas layer.

References:

[1] Zhou Fengde, Yao guangqing, Wang Guochang. *Application of Elman Network in Low Reservoir Sensitivity Prediction* [J]. Geological Science and Technology Information, 91-94

[2] Ye Zhengrong, Fan Shizhong, Yang Puhua. *Comprehensive Application of BP Network And Optimization of Algorithm for Establishing Reservoir Sensitivity Prediction Model* [J]. Drilling Fluid And Completion fluid, 2000, 17 (3): 9-13.

[3] Gan Lu, Wang Yi, Wang Bo. *Application of BP Neutral in Fault Diagnosis of Electric Submersible Pump Well* [J]. Oil Drilling and Production Technology, 2011, 33 (2): 124-127.

[4] Li Junjian, Jiang Hanqiao, Chen Minfeng and etc. *Relevance and Analysis of Adaptability Towards the Influence Factors of Polymer flooding Effect* [J]. Oil Drilling and Production Technology, 2008, 30 (2): 86-90.

[5] Zhang Xuanqi. *Grey Evalution on Reservoir Sensitivity* [J]. Petroleum Geology and Development of Daqing, 2004, 23 (6): 60-62.

[6] Wang Bin, Fan Shizhong. *Development of Expert System for Reservoir Sensitivity* [J]. Drilling Fluid And Completion fluid, 2002, 19 (6): 22-24.

[7] Pu Chunsheng, Pei Runyou, Wu Feipeng and etc. *Grey Relational Evaluation Model for Hydraulic Fracturing Performance of Oil Field in Hujian Mountain* [J]. Oil Drilling and Production Technology, 2010, 32 (4): 54-56.

[8] Zheng Deru. *Regression and Correlation Analysis* [M]. Shanghai: Shanghai People's Publishing House, 1984.

[9] BENDAT J S, PIERSOL A G. *Engineering Application of Correlation and Spectral Analysis* [M]. Ling Fugen translates. Beijing: National Defense Industry Press, 1983.

[10] He Xiaoqun. *Practical Regression Analysis* [M] Beijing: Higher Education Press, 2008.

[11] Chen Jiading, Sun Shanze, Li Dongfeng and etc. *Handout of Mathematical Statistics* [M]. Beijing: Higher Education Press, 2007.

[12] Sun Jianmeng, Li Zhaocheng, Tan Weiyi. *Quick Prediction of Reservoir Sensitivity with Single Correlation Analysis* [J]. Drilling Fluid And Completion fluid, 1996, 16(1): 1-5.

[13] Sun Shuqiang. *Protection Technology of Oil-gas Layer* [M]. Dongying, Shangdong: Petroleum University Publishing House, 2005.