ELSEVIER
CrossMark
Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 102 (2016) 434 – 440
Procedia
Computer Science

12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria

Prediction of multivariable properties of reservoir rocks by using fuzzy clustering

R.Y. Aliyarov [a]*, R.A. Ramazanov[b]

[a]*Azerbaijan State Oil and Industry University, Azadliq 20, Baku, Azerbaijan*
[b]*Baker Hughes, 3 Salyan highway 15th km, AZ1063, Baku, Azerbaijan*

Abstract

Forecasting of geological parameters is very important for decision making on investment to exploration of new hydrocarbon structures and fields. On the one hand, the complexity of this problem is originated from the nonlinearity and uncertainty of behavior of an ensemble of interrelated parameters changing with respect to the depth. This phenomenon is considered analogously to time series, where the depth plays the role of time. On the other hand, the available data are irregular over the depth, as represent different geological bodies with distinct properties. These features mandate necessity to consider multivariable time series of geological parameters with irregular intervals. In this paper, we consider multilag forecasting of five geological parameters over the depth. As a model of forecasting, fuzzy c-means based fuzzy if-then rules are used and this allows better capture of high complexity of the considered phenomena than the classical precise forecasting model. The experimental data show validity of the suggested approach.

Peer-review under responsibility of the Organizing Committee of ICAFS 2016

*Keywords:*reservoir rocks; porosity; compaction;South Caspian sedimentary basin; forecast; multivariate time series; fuzzy C-means clustering

1. Introduction

The prediction of reservoir rocks properties is an important part of preparation prior to make decision on start of an exploration works at new structures, considering high cost associated with the operations.

* R.Y. Aliyarov, **R.A. Ramazanov. Tel.: +994702292969*, +994557872392**
E-mail address: r.aliyarov@asoiu.edu.az*, rizvan.ramazanov@bakerhughes.com**

Peer-review under responsibility of the Organizing Committee of ICAFS 2016
doi:10.1016/j.procs.2016.09.423

The problem is quite actual for South Caspian sedimentary basin (SCB), considering the history of its development. Most of discovered oil and gas fields involved in production are located on the periphery of the SCB while a new perspective structures are found in weakly studied deep water part of the depression, relatively far away from the periphery of the basin.

The traditional approach for reservoir rocks parameters prediction is based on construction of relationship between the parameters and depth or effective stress (more often depth) and on establishing the approximation functions derived from correlation-regression analysis[1]. The satisfactory results obtained from this approach were considered acceptable assuming that the predicted parameter is only depends on depth or effective stress.

Another approach relying on multidimensional correlation-regression analysis could not be considered as confident for wider use because of requirement of known arguments in multidimensional equation.

It is well known that the reservoir rock properties developed under number of geological factors which role and magnitude varied in time and spatial during post sedimentation period. Therefore the prediction method should be designed to account the variation of geological parameters influenced to the predicted rock characteristics.

Let us discuss the state-of-the-art solution of the considered problem. The experimental studies on the rock samples from the SCB fields were performed and described by Buryakovsky and others[1]. The study based on the analysis of the core samples showed that the rock properties vary with increase of overburden stress representing the rock occurrence at different depths. The results indicated that the porosity decrease of 7% could happen if the overburden pressure increased to 100 MPa.

Buryakovsky and others[2] have paid a special attention to define the mathematical model of the compaction of sandstone and shale formations in SCB. The authors used statistical approach based on the empirical data and inference of interconnections through generalization, analysis, and comparison of the features of geologic systems at certain discrete moments of the geologic time. Approximation of the discrete data by a continuous function obtains an empirical equation for a parameter of the geologic object under study as a function of time. They constructed the model of the shale porosity in relation with depth of burial, formation geological age and lithology. The relationship between porosity of sand formations and burial depth was generated using correlation-regression analysis method. The obtained relationship was used to predict the petrophysical parameters including porosity up to 9000 m depth.

The paper[3] is dedicated to the prediction of main petrophysical parameters as porosity and permeability based on the real log data and utilizing the artificial intelligent techniques: Fuzzy Logic, Support Vector Machine and Functional Networks. The obtained prediction showed higher confidence in results for hybrid model rather than of individual technique.

The combination of type -2 fuzzy logic system and sensitivity based linear learning method (SBLLM) was used in work[4] as hybrid approach for prediction of the oil and gas reservoir permeability and PVT parameters in Middle East fields. The parameters in oil and gas industry are discrete and variable depending on number of parameters which in turn creates uncertainties. The uncertainty handling capability of type-2 fuzzy logic system was combined with generalization potential of SBLLM made the model more robust. The conclusion was made that the hybrid model has advantage over the separate usage of the methods; hence the prediction results has better correlation with real data and the machine learning and computation time is more convenient for implementation of the method in the industry.

The applications of fuzzy logic in petrophysics and the basic concepts behind the litho-facies, permeability and thin bed resolutions computation using fuzzy logic methods considered in paper[5]. The results of fuzzy logic prediction methods deployed at North Sea oil and gas fields data helped to conclude that the method can be used as a simple tool for confirming knowing correlations and as powerful predictor in uncored wells.

S. Cuddy[6] considers the fuzzy logic tool utilization for litho-facies and permeability prediction in North Sea oil and gas fields and comparison of fuzzy logic with other models. The litho-facies typing is used for well correlation and as input for building a 3D model of the field.

Arash Mirzabozorg and others[7] incorporated the reservoir engineering knowledge into history matching and optimization framework, by coupling a rule based fuzzy system with population based sampling method. The method was used for investigation of future performance forecast of the Teal South reservoir model.

The literature review shows that the prediction of geologic features is predominantly based on mathematical models utilizing traditional analytical and statistical approaches with a few attempts to use computational intelligence techniques. Due to the complexity and uncertainty related to rock properties' behavior, the use of fuzzy sets theory[8-10] would provide a more adequate basis for forecasting. In this paper we consider Fuzzy C-means clustering based prediction of multivariable properties of reservoir rocks.

The paper is structured as follows. In Section 2 we provide some prerequisite material including multivariate time series, fuzzy C-means clustering etc. which is used in the study. In Section 3 we formulate the statement of the problem of multivariate time series prediction and the fuzzy C-means clustering based solution method. In Section 4 we consider an application of the proposed study. Section 5 is conclusion.

2. Preliminaries

Definition 1. Multivariate time series. Consider n time series variables $\{y_{1t}\},...,\{y_{nt}\}$. A multivariate time series is the (n×1) vector time series Y_t where the ith row of Y_t is $\{y_{it}\}$. That is, for any time $Y_t = (y_{1t},..., y_{nt})'$.

Definition 2. Fuzzy C-means clustering. The problem of fuzzy C-means clustering consists in partition of the considered time series $X=\{x_1,x_2,\ldots,x_n\}$ into c fuzzy clusters such that the following criterion is minimized:

$$\boldsymbol{J_m = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^{m} \left\| x_i - v_j \right\|^2 \rightarrow min} \qquad (1)$$

subject to

$$\boldsymbol{0 < \sum_{i=1}^{n} u_{ij} < n \ \ (j = \overline{1,c}) \& \sum_{i=1}^{c} u_{ij} = 1 \ \ (i = \overline{1,n})} \qquad (2)$$

where $\left\| \cdot \right\|^2$ is the Euclid distance, c is the number of clusters given in advance, m is the value of fuzzifier which defines curvature of membership functions of obtained clusters, are coordinates of centers of clusters to be found.

Definition 3. Fuzzy inference system (FIS). Fuzzy inference system is a system that maps fuzzy inputs to fuzzy outputs by using fuzzy IF-THEN rules:

$$\begin{aligned} &IF\ y_1\ is\ A_{11}\ and\ y_2\ is\ A_{12}\ and,...,and\ y_n\ is\ A_{1n} \\ &THEN\ z_1\ is\ B_{11}\ and\ z_2\ is\ B_{12}\ and,...,and\ z_m\ is\ B_{1m} \end{aligned}$$

or

$$\begin{aligned} &IF\ y_1\ is\ A_{21}\ and\ y_2\ is\ A_{22}\ and,...,and\ y_n\ is\ A_{2n} \\ &THEN\ z_1\ is\ B_{21}\ and\ z_2\ is\ B_{22}\ and,...,and\ z_m\ is\ B_{2m} \end{aligned}$$

or
...
or

where $A_{k1}, A_{k2},..., A_{kn}, B_{k1}, B_{k2},..., B_{km}$ are fuzzy sets.

Given current values of inputs $y_1,..., y_n$, the FIS computes corresponding values of outputs $y_1,..., y_n$ as follows.

Step 1. Fuzzification. The membership degrees $\mu_{A_{k1}}(y_1),...,\mu_{A_{k1}}(y_n)$ for every rule $k = 1,..., K$ are determined.

Step 2. The rule activation degree α_k is determined for every rule $k = 1,..., K$ by using the operator “and”. This operator is the fuzzy conjunction[11-12] $T(\mu_{A_{k1}}(y_1),...,\mu_{A_{kn}}(y_n))$. In Mamdani-type FIS $T = \min$.

Step 3. The membership functions of the outputs are determined:

$\mu_{B'_{k1}}(z_1) = \min(\alpha_k, \mu_{B_{k1}}(z_1)),...,\ \mu_{B'_{km}}(z_m) = \min(\alpha_k, \mu_{B_{km}}(z_m))$

Step 4. The aggregation of the fuzzy outputs are determined by using the “or” operator S :

$\mu_{B_1}(z_1) = S(\mu_{B'_{11}}(z_1),...,\mu_{B'_{1m}}(z_1))$,

. . $\mu_{B_m}(z_m) = S(\mu_{B'_{1m}}(z_m),...,\mu_{B'_{Km}}(z_m))$

Operator S is the fuzzy disjunction[12-13]. In Mamdani-type FIS $S = \max$.

Step 5.Compute z_l by using defuzzification of $B_l, l = 1,..., L$:

$$z_l = \frac{\int_{Z_l} \mu_{B_l}(z_l) z_l}{\int_{Z_l} \mu_{B_l}(z_l)},$$

where Z_l is the universal set for z_l.

Definition 4. Linear interpolation of irregular time series. Let time series $(t_1, y_1),...,(t_n, y_n)$ be given. If the condition $t_{i+1} - t_i \neq const$, $(t_1, y_1),...,(t_n, y_n)$ is referred as to irregular time series. Linear interpolation of irregular time series to regular time series is as follows. Denote $t_1^* = t_1, t_n^* = t_n$. $[t_1^*, t_2^*]$ interval is divided by equally spaced grid points $t_1^*, t_2^*,..., t_{m-1}^*, t_m^*$, i.e. the spacing $\Delta t_i = t_{i+1} - t_i$ is the constant interval. The values $y_i^*, i = 2,..., m-1$ that correspond to equally spaced grid points $t \in \{t_i^*\}, i = 1,..., m$ are found by using linear interpolation as follows:

$$y_i^* = \frac{t_{i+1} - t}{t_{i+1} - t_i} y_i + \frac{t - t_i}{t_{i+1} - t_i} y_{i+1}.$$

3. Statement of problem and the method of solution

3.1. Geological parameters behavior

The sedimentary basin develops under variation of certain geological conditions including the source of classic material, the sediments transportation speed, the depth of basin, hydrodynamic and geochemical environment during the sedimentation and number of local and regional processes occurred after the sedimentation. All these aspects led in heterogeneity, the complexity of substantial and mineralogical composition of the rocks and maze of pore space. The complexity of rocks causes the multidimensionality of information base as well as its heterogeneity and fuzziness.

The distinctive feature of SCB is significant depth difference of reservoir rocks occurrence at the hydrocarbon fields located on periphery and submerged parts of the basin. The depth difference is more than 6,000 m at the discovered hydrocarbon fields and increases toward the central part of SCB. The reservoir rock properties should be influenced by such big depth difference in occurrence of identical stratigraphic horizons and therefore it is critical to predict the rocks properties at depths of prospective structures.

The main parameter considered for forecasting with the depth in current paper is porosity in association with carbonate, sand, silt and shale content in the rocks. Porosity decreases with increase of the depth due to compaction of the reservoir rocks. The compaction is complex process and its intensity depends on the rock properties, mineralogical composition and effective stress.

3.2. Formal statement of problem of multivariate time series forecasting

Let the multivariate time series of rock properties $\{\mathrm{Y}_t\}$ be given. Consider the problem of multilag forecasting of $\{\mathrm{Y}_t\}$, i.e. determination of $\{\mathrm{Y}_{t+1}\},...,\{\mathrm{Y}_{t+n}\}$. We assume that the next value of the vector time series $\{\mathrm{Y}_{t+i}\}, i = 1,..., n$ depends on the previous values $\{\mathrm{Y}_{t+i}\}, i = 0,..., n-1$. Formally, there exists such a function f that $\{\mathrm{Y}_{t+i}\} = f\{\mathrm{Y}_{t+i-1}, \mathrm{Y}_{t+i-2},..., \mathrm{Y}_{t+i-m+1}\}$. Then the problem of forecasting consists in determination of an approximated f and computation of an approximated $\{\mathrm{Y}_{t+i}\}$ denoted $\{\hat{\mathrm{Y}}_{t+i}\} = f\{\mathrm{Y}_{t+i-1}, \mathrm{Y}_{t+i-2},..., \mathrm{Y}_{t+i-m+1}\}$. Consider a multilag forecasting problem:

Find such f and m that

$$
\begin{aligned}
&\{\hat{Y}_{t+1}\} = f\{Y_t, Y_{t-1}, ..., Y_{t-m}\},\\
&\{\hat{Y}_{t+2}\} = f\{\hat{Y}_{t+1}, Y_t, ..., Y_{t-m+1}\},\\
&...\\
&\{\hat{Y}_{t+n}\} = f\{\hat{Y}_{t+n-1}, \hat{Y}_{t+n-2}, ..., Y_t\},
\end{aligned}
$$

and

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\{Y_{t+i}\} - \{\hat{Y}_{t+i}\}\right)^2}{n}} \to \min$$

3.3. Solution of the problem

Solution of the problem consists of the following steps:

Step 1. Given data $\{Y_{t+j}\},\ j = 1,...,l$, construct the interpolated data $\{Y_{t+j}\},\ j = 1,..., N*l$.

Step 2. Conduct Fuzzy C-means clustering (Definition 2) of the interpolated data $\{Y_{t+j}\},\ j = 1,..., N*l$ to obtain fuzzy IF-THEN rules:

Rule $: k,\ k = 1,..., K$

$$
\begin{aligned}
&IF\ y_{1t-m}\ is\ A_{k1}^{1}\ and\ y_{1t-m+1}\ is\ A_{k2}^{1}\ and, ..., and\ y_{1t}\ is\ A_{km}^{1}\\
&and\ y_{2t-m}\ is\ A_{k1}^{2}\ and\ y_{2t-m+1}\ is\ A_{k2}^{2}\ and, ..., and\ y_{2t}\ is\ A_{km}^{2}\\
&...\\
&and\ y_{nt-m}\ is\ A_{k1}^{n}\ and\ y_{nt-m+1}\ is\ A_{k2}^{n}\ and, ..., and\ y_{nt}\ is\ A_{nm}^{n}\\
&THEN\\
&y_{1t+1}\ is\ B_{n1}^{1}\ and\ y_{2t+1}\ is\ B_{n1}^{2}\ and, ..., and\ y_{nt+1}\ is\ B_{k1}^{n}
\end{aligned}
$$

Step 3. Conduct testing of the constructed IF-THEN rules.

Step 4. Conduct multilag forecasting

4. Experimental investigation

The examined data set consist of experimentally derived parameters from core samples of the “Fasila” suit of Productive Series middle Pliocene age formations in SCB. The core sample data set characterized by sand, silt, carbonate and shale components of each sample. The sand, silt and shale components in the data set were distinguished based on the predefined grain size ranges and represented by the percentage of measured weight of solid part of each rock sample.

At Step 1 we conduct linear interpolation of the data in Table 1.

At Step 2 we conduct Fuzzy C-means clustering of the data given in Table 1 to construct Mamdani type of fuzzy IF-THEN rules. We have found that the optimal number of fuzzy clusters is 10. Some of the constructed IF-THEN rules are given below:

IF y_{1t-1} *is Low average and* y_{2t-1} *is Average and* y_{3t-1} *is Low average and* y_{4t-1} *is Low average and*
and y_{5t-1} *is Average*
and y_{1t} *is Low average and* y_{2t} *is Average and* y_{3t} *is Low average and* y_{4t} *is Low average*
and y_{5t} *is Average*
THEN
y_{1t} *is Low average and* y_{2t} *is Low average and* y_{3t} *is Low average and* y_{4t} *is Average and*
and y_{5t} *is Average*

IF y_{1t-1} *is Average and* y_{2t-1} *is Low and* y_{3t-1} *is Average and* y_{4t-1} *is Average and*
and y_{5t-1} *is Average*
and y_{1t} *is Low average and* y_{2t} *is High Average and* y_{3t} *is Low average and* y_{4t} *is Low*
and y_{5t} *is High*
THEN
y_{1t} *is Low and* y_{2t} *is High average and* y_{3t} *is Low and* y_{4t} *is Average and*
and y_{5t} *is High*

Table 1. An example of data set used for study

Depth	Carbonate %	Sand %	Silt %	Shale%	Porosity %
170.1	5.0	35.0	34.5	30.5	22.7
327.5	4.6	59.9	18.0	22.1	30.0
342.5	9.0	16.1	62.7	21.2	26.2
361.5	12.0	1.3	64.7	34.0	21.6
414.0	14.0	26.6	38.8	34.6	21.4
…					
5693.0	7.1	0.7	73.5	25.8	14.0
5693.5	6.0	30.0	53.1	16.9	10.0
5853.0	0.0	11.9	42.6	45.5	18.3

The training RMSEs for the fuzzy IF-THEN rules are as follows:
$RMSE(y_1) = 2.6, RMSE(y_2) = 15, RMSE(y_3) = 11.2, RMSE(y_4) = 7.1,\ RMSE(y_5) = 4.$

At Step 3 we conduct testing of the fuzzy IF-THEN rules. The testing RMSEs are as follows:
$RMSE(y_1) = 3.8, RMSE(y_2) = 18.8, RMSE(y_3) = 13.5, RMSE(y_4) = 10.9,\ RMSE(y_5) = 6.9.$

Finally, at Step 4 we conduct multi lag forecasting of the considered five variables from 5,900 till 6,200 meters depth with the step of 35 meters. The obtained results are shown in Table 2.

Table 2. The results of the forecast

	Forecasted values				
Depth	Carbonate %	Sand %	Silt %	Shale%	Porosity %
5887.652	7.39	45.74	31.19	26.71	21.95
5922.304	6.88	46.00	29.18	26.73	22.85
5956.955	7.73	40.14	34.73	26.00	21.39
5991.607	7.95	38.33	36.74	24.19	19.63

6026.259	8.88	36.95	39.86	22.84	18.79
6060.911	9.01	35.46	40.14	23.08	18.74
6095.563	10.06	33.33	40.57	24.63	17.36
6130.215	10.70	31.05	41.45	25.85	16.41
6164.866	11.19	29.98	42.81	26.00	15.96
6199.518	11.50	28.59	43.49	26.89	15.45

The prediction results of examined parameters show presence of some trends for each parameter. Generally there should not be necessarily the trends between all considered parameters and the depth except the porosity. However the depth of studied core samples is also indirectly represents their areal position as the considered formation burial depth varies over the region and this in turn creates some trends in parameters. The forecast results are reasonable and demonstrate the porosity reduction with the depths with the consideration of variation of the other parameters.

5. Conclusion

We proposed a multivariable multilag model for forecasting of an ensemble of important geological parameters as a function of the depth which can help in estimating of new the structures' potential. Interpolation based processing of irregular geological data and fuzzy clustering based forecasting of multivariable time series is able to provide a suitable accuracy of the prediction. The fuzzy based approach used in this study allowed to consider the variation of different parameters during the forecasting and demonstrated an advantage over classic forecast methods. Experimental investigation shows validity of the proposed forecasting methodology.

References

1. Buryakovsky LA, Dzhafarov IS, Dzhevanshir RD. *Prediction of Petrophysical Properties of Oil and Gas Reservoirs and Seals*. Moscow: Nedra Publishing House ;1982.
2. Buryakovsky LA, Chilingar GV, Aminzadeh F, *Petroleum Geology of the South Caspian Basin*. Gulf Professional Publishing;2001
3. Anifowose F, Abdulraheem A. *Prediction of porosity and permeability of oil and gas reservoirs using hybrid computational intelligence models*.In *Proc. North Africa Technical Conference and Exhibition, 14-17 February, Cairo, Egypt*, 2010;SPE 126649.
4. Olatunji SO, Selamat A, Abdurraheem AA. *Harnessing the Power of type-2 fuzzy logic system in the Prediction of reservoir properties. SPE Saudi Arabia Section Annual Technical Symposium and Exhibition, 21-23 April, Al-Khobar, Saudi Arabia* 2015;SPE-178005-MS.
5. Cuddy S. *The Application of the mathematics of fuzzy logic to petrophysics*. 38th Annual Logging Symposium SPWLA 1997.
6. Cuddy SJ. *Litho-Facies and permeability prediction from electrical logs using fuzzy logic.* 2000;SPE-65411-PA.
7. Mirzabozorg A, Nghiem L, Zhangxin C, Chaodong Y, Heng L. *How Does the Incorporation of engineering knowledge using fuzzy logic during history matching impact reservoir performance prediction?.* In *Proc. Heavy Oil Conference-Canada, Calgary, Alberta, Canada 10-12 June,* 2014.
8. Aliev RA. *Fundamentals of the fuzzy logic-based generalized theory of decisions*. New York, Berlin: Springer; 2013.
9. Aliev RA, Pedrycz W, Huseynov OH. Decision theory with imprecise probabilities. *Int J Inf Tech Decis* 2012; **11(02)**: 271-306..
10. Aliev RA, Mamedova GA, Aliev RR. *Fuzzy Sets Theory and its application* . Tabriz: Tabriz University; 1993.
11. Aliev RA, Alizadeh AV, Guirimov BG. Unprecisiated information-based approach to decision making with imperfect information. In *Proc. 9th Int. Conf. on Application of Fuzzy Systems and Soft Computing ICAFS*-2010, Prague, 2010, p. 387-397.
12. Aliev RA, Tserkovny A. Systemic approach to fuzzy logic formalization for approximate reasoning. *Information Sciences* 2011;**181**:1045-1059. .
13. Aliev RA. Fuzzy expert systems. In: Aminzadeh F, Jamshidi M, editors. *Soft computing*: Fuzzy Logic, Neural Networks *and Distributed Artificial Intelligence.* New Jersey: Prentice-Hall, Inc. 1994; p. 99-108.