ORIGINAL ARTICLE

Reservoir rock permeability prediction using SVR based on radial basis function kernel

Majid Bagheri[1] · Hadi Rezaei[1]

Abstract

Permeability is one of the most important parameters of the rock reservoir. This parameter represents the ability of fluid pass through the porous medium without changing the structure of the rock which plays a main role in the production rate of a hydrocarbon reservoir. Permeability could be obtained through laboratory using core plugs, although the measurements is highly accurate, but these analyses are costly, time-consuming, and also the core data are not available for all wells. Therefore, permeability is usually predicted using well logs by regression techniques. In this study, support vector regression (SVR) based on radial basis function is developed to estimate permeability in South Pars gas field of Iran. For this purpose, first four electrofacies were identified using a method called multi-resolution graph-based clustering (MRGC), and thereafter the method applied for each facies. To evaluate the prediction model, the correlation coefficient between the real permeability (determined from core plugs) and the estimated was calculated for each electrofacies. The values for the four electrofacies designated obtained 88.2, 78.51, 84.73, and 77.54 percent, respectively. The high amount of obtained correlations for each electrofacies demonstrates the power of the regression model for a reliable permeability prediction of the reservoir.

Keywords Permeability · Electrofacies · Support vector regression · Radial basis function · Well log · Core data

Introduction

One of the key issues in the evaluation of hydrocarbon reservoirs is the use of well data to predict petrophysical properties, such as porosity, permeability, water saturation, etc. Also, facies analysis using seismic and well data is necessary for reservoir modeling (Bagheri et al. 2013, 2015). Therefore, the main challenges of reservoir engineers are determining these components in the reservoir (Mohaghegh et al. 1996). Reservoir permeability is one of the most important properties of oil and gas fields for reservoir characterization. In fact, it is not possible to have accurate solutions to many petroleum engineering problems without having accurate permeability knowledge (Worden et al. 2018; Zhu et al. 2016).

Till now, the petroleum industry tried to acquire reliable permeability values via laboratory measurements on cores or well test interpretation. But these investigations are expensive and take time, and also the core plugs are unreachable in all wells. Therefore, permeability regularly is estimated, using empirical equations or statistical (component and non-component) regressions (Lee and Dutta-Gupta 1999) from well logs. In general, well logs provide valuable but indirect information on mineralogy, texture, sediment structure, fluid content and reservoir hydraulic properties and could be used for this purpose. Specific responds of the logs of the formation can represent electrofacies that are often correlated with the rocky facies (Bagheri and Riahi 2017).

In recent years, different methods such as artificial neural networks, hybrid intelligent, fuzzy logic, and other different machine learning techniques have been employed to overcome the common constraints in multiple regression (Taoreed and Gondal 2018; Olatunji et al. 2015; Akande et al. 2015). Artificial neural networks are gaining popularity as tools for estimating reservoir parameters such as permeability from limited, common data suites. Neural networks are non-algorithmic, analog, distributive and parallel information processing methods that have proven to be powerful

✉ Majid Bagheri
  majidbagheri@ut.ac.ir

  Hadi Rezaei
  h.rezaei72@yahoo.com

[1] Institute of Geophysics, University of Tehran, PO Box 14115-6466, Tehran, Iran

pattern recognition tools. Since they process data and learn in a parallel and distributed fashion, they are able to discover highly complex relationships between several variables that are presented to the network. As a model-free function estimator, neural networks can map input to output no matter how complex the relationship (Akande et al. 2014).

The purpose of this research was to develop a methodology using support vector machine (SVM) to predict the permeability for wells using the logs data. After clustering of well log data using multi-resolution graph-based clustering (MRGC) method and obtaining electrofacies, permeability has estimated in each facies using support vector regression (SVR) based on radial basis function. Finally, the estimated permeability is compared with the real permeability obtained from core plug to evaluate the technique. The method provides a very useful approach for predicting permeability from well log data where data are poor or relationships between inputs and desired outputs are not exactly known. Final results show the successful application of SVR as a powerful tool to predict permeability from well logs.

Electrofacies identification using well data

The first step to estimate permeability is well log clustering and identifying electrofacies. Electrofacies are a set of well log responses which specify a layer and detect it from other layers (Nashawi and Malallah 2009). Determining these electrofacies in reservoir formations is one of the common studies in describing hydrocarbon reservoir properties. The abundant use of these electrofacies and their ability to determine specific reservoir parameters, according to the type of input data, have made this method one of the most powerful tools in reservoir studies. Also, the electrofacies are used in some cases such as separation of reservoir segments from non-reservoirs, replacing rock groups in reservoir models and matching structures in the field. The significance of these data is as if they were referred to as virtual cores (Khan Mohammadi and Sherkati 2010). In this study, the method of multi-resolution graph-based clustering (MRGC) has been used for electrofacies analysis. In this method, other disadvantages of clustering methods such as previous knowledge about the number of clusters and early parameters have been eliminated. The MRGC method is very suitable for complex structural analysis and clustering the data sets in different shapes, sizes and densities (Khoshbakht and Mohammadnia 2012).

Multi-resolution graph-based clustering (MRGC) method

Introducing two parameters (NI and KRI) makes the MRGC to offer better results than other methods. The neighborhood

index (NI) is a parameter that is defined by each sample weighted x relative to all y samples. Two points near each other can be easily separated if they have a high neighborhood index (Ye and Rabiller 2000). As a result, the number of electrofacies is easily determined by the set of relations as follows:

$$
\begin{aligned}
S(x) &= \sum_{n=1}^{N-1} \exp(-m/a) \\
S_{\min} &= \text{Min}\,\{S(x_i)\} \\
S_{\max} &= \text{Max}\,\{S(x_i)\} \\
\text{NI}(x) &= \frac{S(x) - S_{\min}}{S_{\max} - S_{\min}}
\end{aligned}
\tag{1}
$$

N denotes the total number of samples, x denotes m-nearest neighbor y, and a is a smoothing parameter that is larger than zero. The value of $\text{NI}(x)$ varies between zero and one, so as with increasing $\text{NI}(x)$, sample x approaches the center of the cluster. The kernel representative index (KRI) is a parameter that combines the neighborhood index $\text{NI}(x)$, neighborhood function $M\,(x, y)$ (number of neighbors), and the distance function $D\,(x, y)$. $\text{NI}(x)$ provides the ability to identify the kernel of a cluster. By combining the two factors $M\,(x, y)$ and $D\,(x, y)$, a good balance between the size (i.e., the number of samples per cluster) and the cluster volume is created and it increases the consistency of results. If $\text{NI}(x)$ is the value of NI at the point x and y is the first neighbor of x with the condition $\text{NI}(y) > \text{NI}(x)$, the KRI at the point x is estimated using Eq. 2:

$$
\text{KRI}(x) = \text{NI}(x) \times M(x \cdot y) \times D(x \cdot y)
\tag{2}
$$

where $M\,(x, y) = m$, y is mth neighbor of x, and $D\,(x, y)$ is the distance between x and y. If the *KRI* value is sorted down and plotted, several important breakpoints can be observed. These breakpoints correspond to the number of optimum clusters in different segregations. Subsequently, using the K-nearest neighbor approach and the *NI* value, the main natural groups (absorbing groups) are formed. For this, at any point such as p from the k-nearest neighbors, a point such as q whose *NI* value is larger than p and the rest of the neighbors is chosen as the absorption point. If p absorbs all the points around it but is not absorbed by any point, then it is in the center of the cluster. But if both simultaneQuery-ously is absorbed and absorbs points around it, then it is an inner point of the cluster, and if it does not absorb any point, but is absorbed by other points, it lies on the cluster boundary. In this way, natural groups of data are called absorption groups. In the end, these absorption groups are merged into two. Provided that either or at least one of them is devoid of a nucleus previously selected at the KRI stage. In this way clusters (electrofacies) are formed. Provided that either or

at least one of them is devoid of a kernel that previously selected at the KRI stage.

Support vector regression (SVR)

Assume that a training data set is given as $\{(x_1 \cdot y_1) \cdot \cdots \cdot (x_l \cdot y_l)\} \subset \chi \times R$, where χ refers to the input data space ($\chi = R^d$). In support vector regression method introduced by Vapnik and Golowich (1997), the goal is to find the function $f(x)$ such that for each training point, the maximum deviation value ε is obtained from the real y_i value and simultaneously as uniform as possible (flat). In other words, errors less than ε are ignored in this case, but errors larger than this value will not be accepted. The function $f(x)$ is a standard linear form as follows:

$$f(x) = \langle w \cdot x_i \rangle + b \, ; w \in \chi \cdot b \in R, \qquad (3)$$

where$\langle . \rangle$ refers to an inner product in χ space. Being uniform in relation (3) means w is small. Clearly, if $w = 0$, then the simplest mode of the model is obtained. One way to achieve this is to minimize the Euclidean norm of w. We can call this issue a convex optimization problem in the form of relation (4).

$$\text{Min } \frac{1}{2} \parallel w \parallel^2$$
$$\text{Subject to } \begin{cases} y_i - \langle w \cdot x_i \rangle - b \leq \varepsilon \\ \langle w \cdot x_i \rangle + b - y_i \leq \varepsilon \end{cases} \qquad (4)$$

The assumptions in (4) state that there exists a function such as f such that it approximates all points $(x_i.y_i)$ with accuracy ε. But sometimes this may not be the case, and you need to have the capacity to make an error. For this purpose, we introduce the slack variables ξ and ξ^* for optimization in Eq. (4). Hence, we arrive at a relationship 5 that was presented by Vapnik:

$$\text{Min } \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{l} \left(\zeta_i + \zeta_i^* \right)$$
$$\text{Subject to } \begin{cases} y_i - \langle w \cdot x_i \rangle - b \leq \varepsilon + \zeta_i \\ \langle w \cdot x_i \rangle + b - y_i \leq \varepsilon + \zeta_i \\ \zeta_i \cdot \zeta_i^* \geq 0 \end{cases} \qquad (5)$$

The constant $C > 0$ determines the balance between the flatness of f and the tolerable amount of error higher than the value of ε. This is equivalent to dealing with a function called ε-insensitive loss function $|\zeta|_\varepsilon$, which is in the form of relation (6):

$$|\zeta|_\varepsilon = |y_i - f(x)| = \begin{cases} 0 & \text{if } |\zeta| \leq \varepsilon \\ |\zeta| - \varepsilon & |\zeta| - \varepsilon \text{ otherwise} \end{cases} \qquad (6)$$

In Fig. 1, only training points outside the shaded area participate as fine points in the linear model. It can be concluded that most of the optimization problems can be simply solved in the form of a dual formula. A dual formula can be considered as a way of extending support vector machines to nonlinear functions. Hence, a standard duality approach with Lagrange multipliers will help solve the problem (Fletcher 1989). The above problem relates to the linear mode of training data in input space. In regression problems with high-dimensional data (each dimension belongs to a measured quantity and the output variable is the same permeability measured from the core), the distribution of data is nonlinear, so in this case the kernel functions are used (Schölkopf and Smola 2002) and the data are transferred to a space, called Feature Space, in which case the training data will be regression similar to Fig. 1. The support vector regression method is a reliable method for estimating the function due to its unique features such as mathematics and reasoned theory, non-convergence in the local minima, and generalizability (Akande et al. 2014).

Data

The selected data belong to a gas field which is located in the Persian Gulf and is one of the largest gas fields in the world. The reservoir is heterogen, the original poroperm heterogeneities in the Upper Dalan–Kangan reservoir are inherited from their palaeo-platform depositional setting but were modified subsequently during diagenetic processes. Therefore, for precise characteri-zation of the reservoir properties in such a heterogeneous carbonate reservoir, integration of sedimentary and diagenetic features is
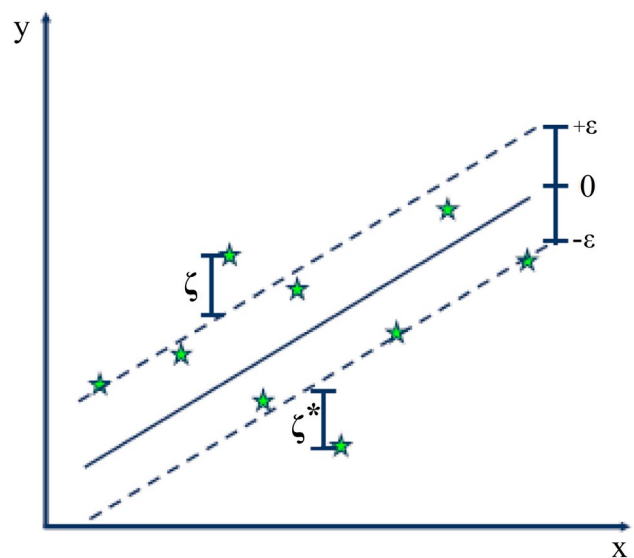

Fig. 1 A schematic view of the soft margin loss setting for a linear SVM

essential (Rahimpour-Bonab and Rahimpour-Bonab 2009). The data consist of 4 wells which include core plug and different well logs such as gamma (GR), neutron (NPHI), density (RHOB), acoustic (DT), deep laterolog (LLD), shallow laterolog (LLS), and micro spherical focused log (MSFL), photoelectric absorption factor (PEF).

Permeability estimation

The permeability estimation steps are summarized as follows:

Step 1: Selecting the proper well log data.
Step 2: Electrofacies analysis using MRGC method.
Step 3: Permeability estimation by support vector regression based on the radial basis function.

Selecting the proper well logs

For electrofacies analysis and clustering, the wells photoelectric absorption factor (PEF), acoustic (DT), neutron (NPHI), and density (RHOB) were used as more proper ones.

Electrofacies analysis using MRGC

Lithofacies characterization is the best solution for overcoming the problem of heterogeneity in determining the petrophysical properties of the reservoir rock, reservoir modeling and identifying producing zones. However, coring as the most robust method of lithofacies identification is very expensive, time-consuming and limited to a few numbers of wells.

Considering the advantages of the MRGC clustering method compared with other clustering methods, this method has been used to determine the electrofacies. MRGC is a fast method that allows the geologist or petrophysist to analyze and test different combinations of data in a short amount of time. It is also not limited by the dimensions of the data and number of the clusters. Using the Facimage module available in the Geolog software, the electrofacies of the reservoir can be acquired by importing the selected logs. The imported well logs in all four wells SP-A, SP-B, SP-C and SP-D, include DT, NPHI, RHOB and PEF logs. Table 1 shows the number of facies clustered by MRGC along with their characteristics, as a result, four electrofacies have been identified.

Figure 2 shows the cross-plot of NPHI and RHOB logs, as well as PEF and RHOB, by virtue of the four electrical facies obtained using the MRGC method.

Electrofacies is named as Table 2 and thereafter permeability will be estimated for each of them separately.

Table 3 shows the percentage of each identified facies in wells SP-A, SP-B, SP-C, and SP-D, using MRGC and Fig. 3 is a histogram of different obtained electrofacies distribution in the four wells.

Permeability estimation using SVR

The SVR as a machine learning needs a training dataset for permeability prediction. The dataset is provided using the clustered electrofacies samples in training wells SP-A, SP-B and SP-C. The samples in the well SP-D are considered unseen in the training process and used as a testing dataset to evaluate the accuracy of the SVR for estimating the permeability. The correlation coefficient (CC) criterion has been used to check the ability of the model. The CC calculates the statistical correlation between real permeability and estimated permeability in the testing well SP-D as follows:

$$CC = \frac{\sum (y_r - y'_r)(y_e - y'_e)}{\sqrt{\sum (y_r - y'_r)^2 (y_e - y'_e)^2}} \tag{7}$$

where y_r and y_e represent the actual values and the estimated values, respectively, while y'_r and y'_e are the mean of the given values. In the following, the permeability will be predicted separately for each electrofacies to make the model more exact and reliable.

Table 1 Electrofacies categorized in the studied sequence (COL = Color, PAT = Pattern)

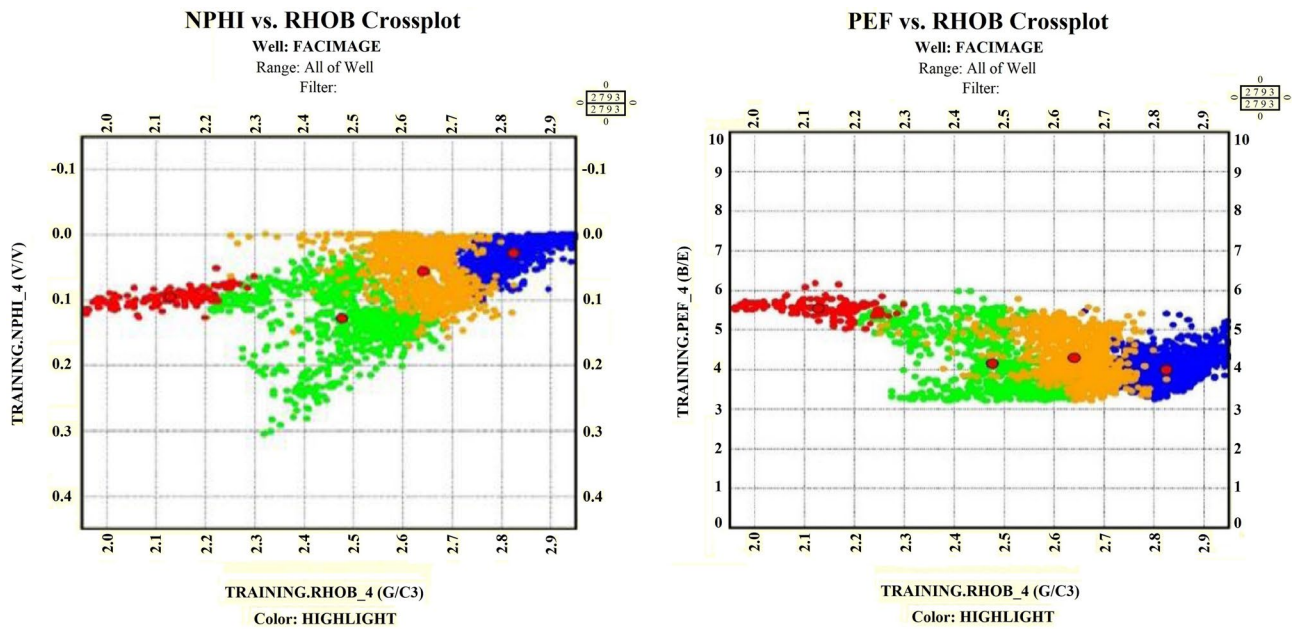| NAME	COL	PAT	WEIGHT	NPHI	DT	RHOB	PEF
FACIES_1			1033				
FACIES_2			875				
FACIES_3			742				
FACIES_4			143				

Fig. 2 The cross-plot of NPHI and RHOB logs, as well as PEF and RHOB, by virtue of the four electrical facies obtained using the MRGC method

Table 2 Naming the Electrofacies

Electrofacies color	Electrofacies number
Blue	1
Orange	2
Green	3
Red	4

Table 3 The percentage of each identified electrofacies in different wells

Well	Electrofacies 1	Electrofacies 2	Electrofacies 3	Electrofacies 4
SP-A	22.54	21.5	16.9	39.06
SP-B	17.87	27.64	23.35	31.14
SP-C	17.37	15.6	39.66	27.37
SP-D	20.88	20.12	32	27

Fig. 3 The histogram of obtained electrofacies using MRGC in the wells of SP-A, SP-B, SP-C, and SP-D

Permeability estimation of Electricfacies 1

To estimate the permeability of the group 1, the samples of facies number 1 in the three wells SP-A, SP-B and SP-C were used as learning dataset to train and create the SVR model. The radial basis kernel function is selected and optimized to build the model. To evaluate the method, the samples in the well SP-D related to electrofacies 1 is considered as an unseen dataset and correlation is calculated.
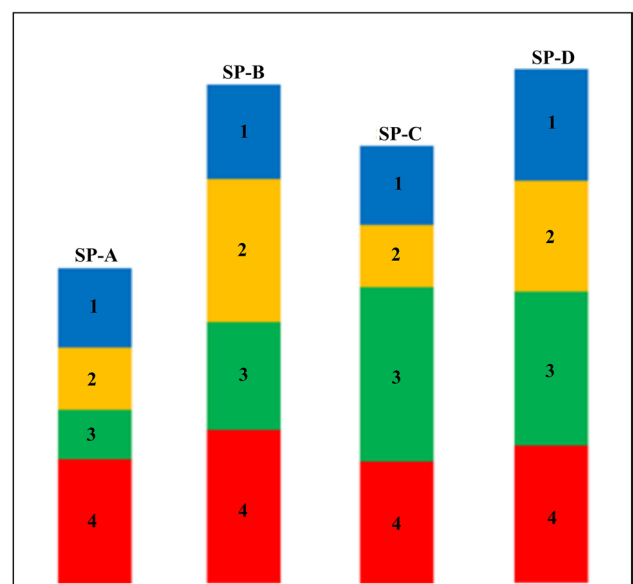
The statistical correlation between actual and estimated permeability values for this electrofacies is determined 88.2%. Figure 4, shows the scattered plot of the real permeability (P_r) versus estimated permeability (P_e) using support vector regression based on radial basis function kernel in electrofacies number 1.
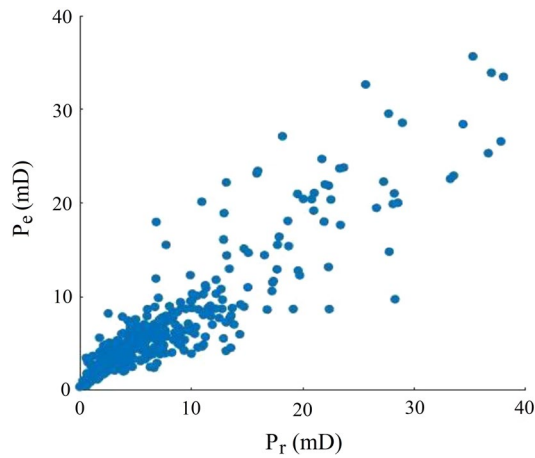
Fig. 4 Scattered plot of actual permeability values versus estimated for electrofacies number 1 in the well SP-D (P_e = estimated permeability; P_r = real permeability)

Permeability estimation of Electricfacies 2

Same as pervious section, for permeability prediction of the electrofacies group 2, learning dataset is gathered using the three wells SP-A, SP-B and SP-C to train the support vector machine. Different kinds of kernel functions of the machine are tested and radial basis is selected as the best for creating the model. To test the machine, testing data set the well SP-D is used to calculate the correlation. The correlation coefficient of real and predicted permeability is calculated as 78.51 percent for the second group. Figure 5 shows the actual permeability versus predicted permeability using SVR based on radial basis function for group 2 of electrofacies.
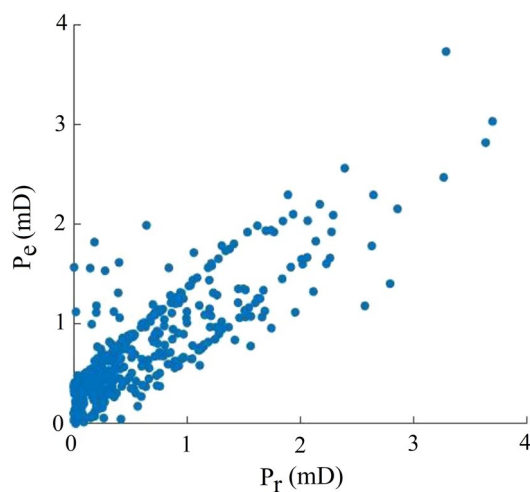
Permeability estimation of Electricfacies 3

To predict the permeability of the electrofacie group 3, the model is created using the training samples in the three wells SP-A, SP-B and SP-C using support vector regression method. To build the model radial-based kernel function is chosen as the optimized function. For investigating the machine, testing dataset which is gathered by samples in the well SP-D is used. The correlation between actual and estimated permeability for this cluster resulted in 84.73%. The real permeability versus estimated is illustrated in Fig. 6 of the electrofacies 3 in the well SP-D.

Permeability estimation of Electricfacies 4

Permeability of the cluster 4 is estimated using the support vector regression machine which is trained using learning dataset collected in wells SP-A, SP-B and SP-C. To create the machine for group 4, same as before, different kernel function is employed which radial basis function provide most powerful predictive model. To check the ability of the developed technique testing samples in well SP-D related to electrofacies 4 is used. The correlation coefficient of core and predicted permeability is calculated as the high amount 77.54%. Figure 7, represents the scattering plot of the real and estimated permeability against each other for cluster number 4. As it is obvious from the figure most of samples are overlapped which proves the high correlation.

Estimated permeability log in well SP-D

The predicted permeability for different electrofaice in the well SP-D in the sections before is merged together to create
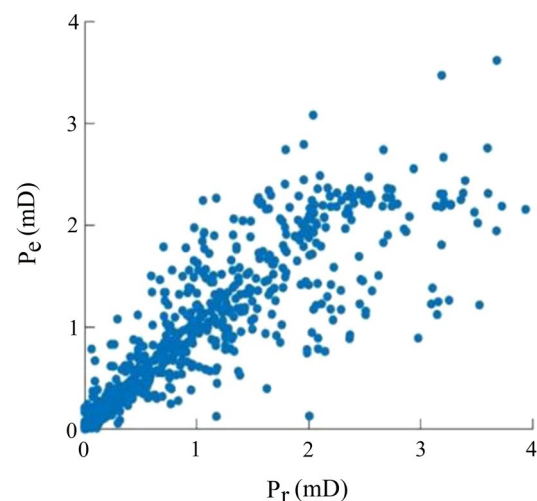

Fig. 5 The actual permeability versus predicted permeability using SVR based on radial basis function for group 2

Fig. 6 The real permeability versus estimated permeability plot of the electrofacies 3 in the well SP-D
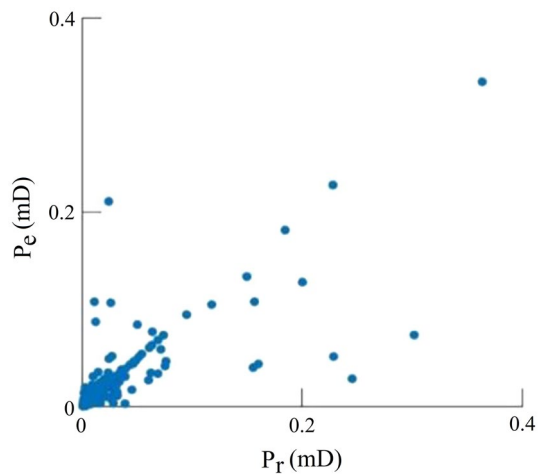
Fig. 7 The scattering plot of the real and estimated permeability against each other for cluster number 4
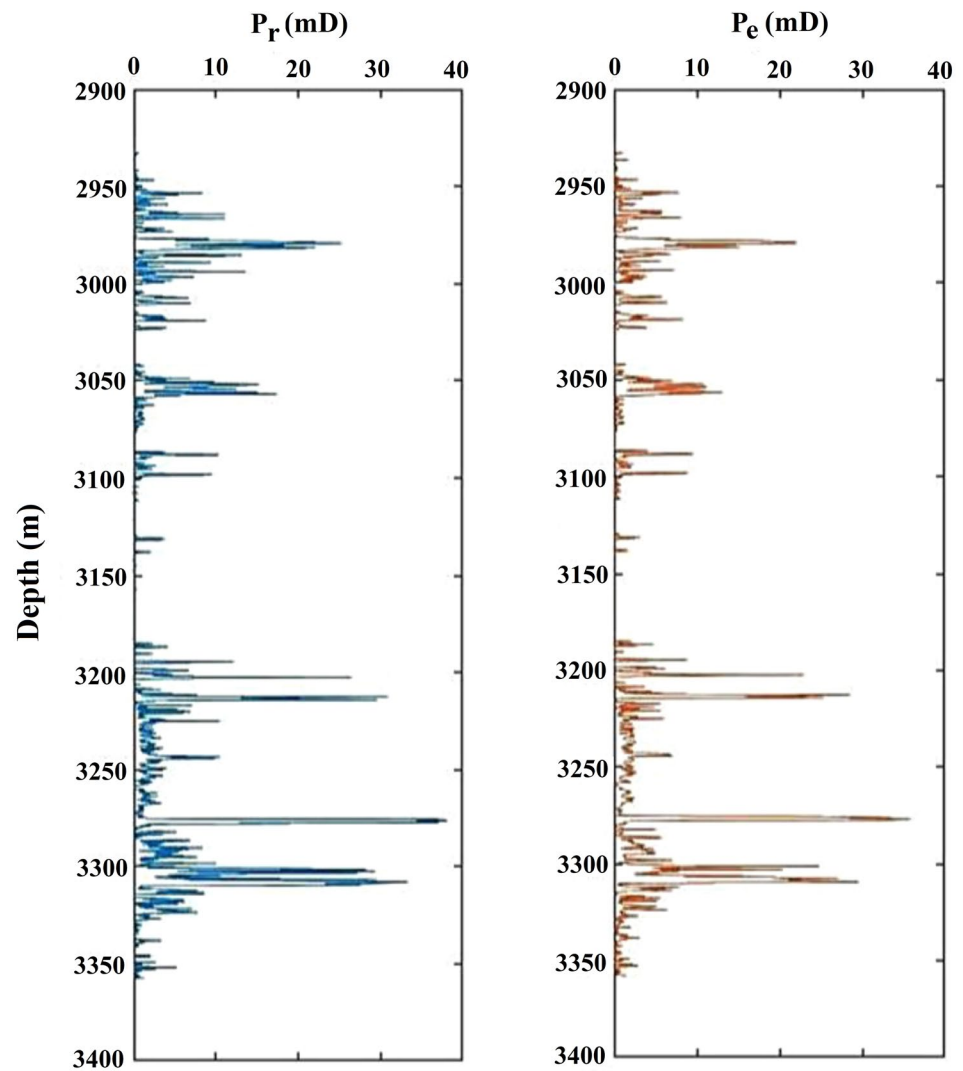
a permeability log. Figure 8, shows the real (P_r) and estimated (P_e) permeability log in the well SP-D.

To compare the results better, a part of the logs is selected (depth from 3180 to 3270 m) and the graphs are plotted together in Fig. 9. From the figure, the similarity of the permeability which is estimated using support vector regression models and real permeability log is obvious. The high matching of the plots in Fig. 9 confirms the ability of created model for permeability prediction using SVR based on radial basis kernel function algorithm.

Conclusion

Study of reservoir properties such as porosity, water saturation, permeability, and volume of shale is really helpful to reduce the risk of producing wells drilling. Permeability is one of the most important parts of reservoir studies in which

Fig. 8 The real (left) and the estimated (right) permeability logs in the well SP-D
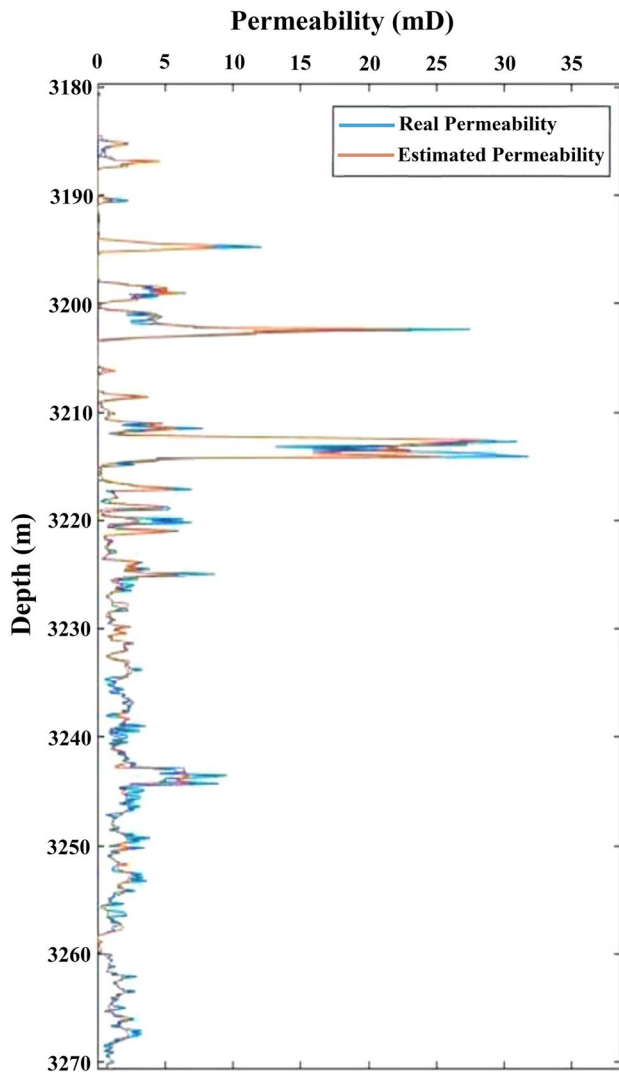
Fig. 9 More accurate comparison of Real (blue) and estimated (red) permeability logs in an arbitrary range in well SP-D. The high matching of the plots in Fig. 9 confirms the ability of created model for permeability prediction using SVR based on radial basis kernel function algorithm

were unseen to test and evaluate the method. The radial basis kernel function was developed to create more powerful support vector machine. The trained model was used to estimate the permeability in well SP-D for different electrofacies, separately. The correlation coefficient criteria, calculated between real and estimated permeability were used to investigate the accuracy and efficiency of the created models.

The correlation coefficient for the four facies in the well SP-D determined 88.2, 78.51, 84.73, and 77.54, respectively. The high amounts of correlation coefficient in the hidden well SP-D refer to the prediction strength of SVR regression model based on radial basis function kernel regarding the heterogeneity of the South Pars gas field reservoir. According to the final results, it could be concluded that SVR based on radial basis function is a powerful algorithm which predicts permeability using well logs reliably.

References

Akande KO, Owolabi TO, Twaha S, Olatunji SO (2014) Performance comparison of SVM and ANN in predicting compressive strength of concrete. IOSR-JCE 16(5):88–94

Akande KO, Owolabi TO, Olatunji SO (2015) Investigating the effect of correlation-based feature selection on the performance of support vector machines in reservoir characterization. J Nat Gas Sci Eng 22:515–522

Bagheri M, Riahi MA (2015) Seismic facies analysis from well logs based on supervised classification scheme with different machine learning techniques. Arab J Geosci 8:7153

Bagheri M, Riahi MA (2017) Modeling the facies of reservoir using seismic data with missing attributes by dissimilarity-based classification. J Earth Sci 28:703

Bagheri M, Riahi MA, Hashemi H (2013) Reservoir lithofacies analysis using 3D seismic in dissimilarity space. J Geophys Eng 10(3):9

Fletcher R (1989) Practical methods of optimization. Wiley, New York

Khan Mohammadi M, Sherkati SH (2010) Fracturing analysis in south pars gas field. Explor Prod Mon 77:43–49

Khoshbakht F, Mohammadnia M (2012) Assessment of clustering methods for predicting permeability in a heterogeneous carbonate-reservoir. J Pet Sci Technol 2(2):50–57

Lee SH, Dutta-Gupta A (1999) Electrofacies characterization and permeability predictions in carbonate reservoirs: role of multivariate analysis and nonparametric regression. In: SPE annual technical conference and exhibition, Houston, Texas (SPE 56658)

Mohaghegh S, Arefi R, Ameri S (1996) Virtual measurement of heterogeneous formation permeability using geophysical well log responses. Log Analyst (Society of Professional Well Log Analysts) 37:32–39

Nashawi IS, Malallah A (2009) Improved electrofacies characterization and permeability predictions in sandstone reservoirs using a data mining and expert system approach. Petrophysics 50(03)

Olatunji SO, Selamat A, Azeez ARA (2015) Modeling permeability and PVT properties of oil and gas reservoir using hybrid model based on type-2 fuzzy logic systems. Neurocomputing, V 157:125–142

various methods with different levels of efficiency have been introduced till now. In this study the support vector regression method is employed for permeability prediction using well logs in South-Pars gas field. SVR is a reliable method for estimating the function due to its unique features such as mathematics and reasoned theory, non-convergence in the local minima and generalizability.

To estimate the permeability using well logs, first by the multi-resolution graph-based clustering method, the number of four electrofacies in the reservoir sequence was obtained in the wells SP-A, SP-B, SP-C, and SP-D. The samples in the wells SP-A, SP-B, and SP-C were considered as learning dataset to train the SVR and the samples in the well SP-D

Rahimpour-Bonab B, Rahimpour-Bonab H (2009) Effects of depositional and diagenetic characteristics on carbonate reservoir quality: a case study from the South Pars gas field in the Persian Gulf. Pet Geosci 15:325–344

Schölkopf B, Smola AJ (2002) Learning with Kernels. MIT Press, Cambridge

Taoreed O, Gondal OwolabiMohammed A (2018) Development of hybrid extreme learning machine based chemo-metrics for precise quantitative analysis of LIBS spectra using internal reference pre-processing method. Analytica Chimica Acta, V 1030:33–41

Vapnik VS, Golowich Smola A (1997) Support vector method for function approximation, regression estimation, and signal processing, neural information processing systems, vol 9. MIT Press, Cambridge

Worden RH, Armitage PJ, Butcher AR, Churchill JM, Csoma AE, Hollis C, Lander RH, Omma JE (2018) Petroleum reservoir quality prediction: overview and contrasting approaches from sandstone and carbonate communities. Geol Soc Spec Pub 435(1):1–31

Ye SJ, Rabiller P (2000) A new tool for electrofacies analysis: multi-resolution graph-based clustering: SPWLA 41st Annual Logging Symposium, Dallas, Texas, USA

Zhu L, Zhang C, Wei Y, Chao-MO Zhang (2016) Permeability prediction of the tight sandstone reservoirs using hybrid intelligent algorithm and nuclear magnetic resonance logging data. Arab J Sci Eng 42:25. https://doi.org/10.1007/s13369-016-2365-2