University of Tehran Press

Journal of the Earth and Space Physics

Homepage: https://jesphys.ut.ac.ir

Vol. 48, No. 4, Winter 2023, P. 77-86
(Research Article)

Prediction of Water Saturation by FSVM using Well Logs in a Gas Field

Moosavi, N.[1] | Bagheri, M.[2] ✉ | Nabi-Bidhendi, M.[3] | Heidari, R.[4]

1. Department of Earth Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran. E-mail: nastaran.moosavi@gmail.com
2. **Corresponding Author**, Department of Earth Physics, Institute of Geophysics, University of Tehran, Tehran, Iran. E-mail: majidbagheri@ut.ac.ir
3. Department of Earth Physics, Institute of Geophysics, University of Tehran, Tehran, Iran. E-mail: mnbhendi@ut.ac.ir
4. Department of Earth Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran. E-mail: r.heidari61@gmail.com

(Received: 11 Dec 2021, Revised: 27 Dec 2021, Accepted: 19 April 2022, Published online: 5 March 2023)

Abstract

Water saturation is one of the key petrophysical parameters that mainly affects the accuracy of initial oil estimation related to a hydrocarbon reservoir. Approximation of this parameter is inevitable since it has a high effect on economic development of hydrocarbon reservoirs. In this paper, we approximate a function, using two wells with two core data sets belonging to each well, to predict water saturation by means of Support Vector Machine (SVM) algorithm in one of the gas reservoirs in the Persian Gulf. Due to the inevitability of noise and outliers in the measured data, SVM is modified to Fuzzy SVM (FSVM). For this purpose, a membership function is applied on the points, so each data point receives a membership degree. In this case, each input point is able to contribute to the learning of decision function. In other words, FSVM is able to enhance SVM by devoting less value to noise and outliers, as a result, better models compared to SVM can be produced. In this study, application of SVM for regression purpose (Support Vector Regression) is carried out on eight logs of DT, GR, RHOB, NPHI, LLD, LLS, MSFL, PEF as input with relevant core data belonging to a gas zone. Then, we determine the coefficients based on the comparison between predicted water saturation (using both SVR and fuzzy SVR algorithm) and core data. Our results show that the predicted water saturation from fuzzy SVR and SVR are 95% and 71%, respectively (higher for fuzzy SVR than SVR).

Keywords: Water Saturation, Hydrocarbon Reservoirs, SVM, FSVM, Well Logs, Core Data.

1. Introduction

Water saturation is the fraction of water in a given pore space. It is expressed in volume per volume, percent or saturation units. This is one of the most applicable petrophysical parameters to evaluate petroleum reservoirs, which directly affects success of drilling operations, complementary and production of oil and gas wells. Therefore, an accurate estimation of this parameter is necessary in exploitation of oil and gas reservoirs. There are two main methods to investigate reservoir parameters; first, core data analysis as a direct method, and second, using well logs as an indirect method.

Core analysis refers to the laboratory study of the core samples retrieved from the cores obtained from drilling operations. It is necessary to analyze the reservoir rock and formation to understand the properties such as porosity, permeability, wettability, and fluid saturation. One of the basic methods to measure water saturation, which was first introduced by the American chemists Ernest Woodward Dean (1888-1959) and David Dewey Stark (1839-1979) in 1920, is to use distillation extraction in a core sample. In this method, water is vaporized using boiling solvent. Then, water is condensed and gathered in a calibrated container. Boiling solvent is also condensed and its volume is reduced from the water. With this method, the volume of water can be measured from the core sample. X-Ray computerized Tomography (CT) and Magnetic Resonance Imaging (MRI) are other alternatives to obtain fluid saturations (Tonstad et al., 1990).

It must be mentioned that after drilling

Cite this article: Moosavi, N., Bagheri, M., Nabi-Bidhendi, M., & Heidari, R. (2023). Prediction of Water Saturation by FSVM using Well Logs in a Gas Field. *Journal of the Earth and Space Physics*, 48(4), 77-86. DOI: http//doi.org/10.22059/jesphys.2022.334938.1007389

Publisher: University of Tehran Press.
DOI: http//doi.org/10.22059/jesphys.2022.334938.1007389

Print ISSN: 2538-371X
Online ISSN: 2538-3906

and bringing up the core to the surface, preservation is necessary to keep the core in the same situation as existed. It means that when the core is cut and is taken to the surface, because of unavoidable changes in pressure, temperature and other conditions inside the earth, the fluid content in the rock alters. If good preservation of the core at the well site does not carried out, then the core will not be a good representative of the formation and its properties. For analyzing water saturation, the preserved core samples must be transferred to the laboratory. Furthermore, migration and evaporation of fluids as well as oxidation of elements inside the samples must be ignored (American Petroleum Institute, 1998). Generally, water saturation can be obtained from core data during or after drilling. Considering the mentioned factors of environmental effects on samples, usually fresh ones are taken for analysis. Core data analysis to obtain water saturation information has been presented by different authors (Walther, 1967; Jia et al., 2020). Measurement of this parameter in the laboratory is costly and takes lots of time. Moreover, core data is not always available for the whole well. Therefore, there are variety of formulas that estimate water saturation from well logs such as resistivity and porosity (Luthi, 2001; Archie, 1942). One of the famous ones is the Archie equation that reads as follows:

$$S_w^{\,n} = \frac{R_w}{(\phi^m \times R_t)} \tag{1}$$

where, S_w is water saturation of the uninvaded zone, n is saturation exponent, R_w is formation water resistivity at formation temperature, ϕ is porosity, m is cementation exponent and R_t is true resistivity of the formation corrected for invasion, borehole, thin bed and other effects. However, one factor that mainly affect these methods is that these formulas highly depend on lithology and formation type. As a result, finding a model with capability of generalization to variety of situations is preferred.

Another popular formula to calculate water saturation is as follows (Pickett, 1966):

$$S_w = I^{-1/n} \tag{2}$$

$$I = \frac{R_t}{FR_w} = \frac{R_t}{R_o} \tag{3}$$

$$F = \phi^{-m} \tag{4}$$

where, S_w is the fractional part of the pore volume filled with water of resistivity R_w. I is resistivity index, n is saturation exponent, R_t is true formation resistivity, F is formation resistivity factor, ϕ is fractional porosity, m is cementation exponent. The accuracy of calculated water saturation depends on the accuracy of mentioned factors. More accurate inputs leads to more accurate water saturation calculation. For example, for determining I there are several methods such as: 1) conventional saturation calculations, 2) R_a VS ϕ_A plots, and 3) R_{wa} plots (Pickett, 1966).

Over the last decade, machine learning methods has been widely used to estimate reservoir parameters (Zhang et al., 2018; Okwu & Nwachukwu., 2019; Li et al., 2021). Water saturation has been estimated using different algorithms (Adeniran et al., 2009; Jafari Kenari & Mashohor, 2013; Bagheripour & Asoodeh., 2014). Each algorithm has its pros and cons. This paper applies SVR algorithm on well logs to obtain water saturation. The superiority of SVR to other algorithms is its high capability of model generalization and low percentage of model error. As the next step, membership functions was used to devote membership degrees to each data point. In other words, data is transformed to a fuzzy system in which each data is in a (0,1) interval (Zadeh, 1965). In this case, noise and outliers receive less degree of membership so their influence on the final model decreases. As a result, better output is produced and modification of SVR to FSVR notably improves the results (Lim et al., 2002; Le et al., 2009). In this paper, two wells within a gas reservoir are utilized; one of the wells is used for the training purpose and the other one is used for the testing operation. Well logs for this study include sonic (DT), neutron porosity (NPHI), density log (RHOB), photoelectric absorption factor (PEF), gamma ray, intensity of natural

radioactivity (GR), resistivity log both shallow and deep (LLD, LLS) and Micro Spherical Focused log (MSFL). Moreover, two sets of core data belonging to each well used as true response of SVR algorithm. Core intervals are in the range of 10120 to 10290 feet (3080-3140 meters) and 10145-10557 feet (3090-3220 meters). After running the algorithms, determination coefficient is calculated for water saturation core data and the predicted model obtained from FSVR illustrates better results compared to SVR.

2. Methodology

2-1. Support Vector Regression (SVR)

SVR is the application of SVM for regression purposes. SVM applies a hyperplane to classify data into two classes. There might be infinite hyperplanes to be able to classify the data. SVM finds a hyperplane for which the margin (the distance between support vectors and hyperplane) is maximum. Support vectors are the data points that are nearest to the hyperplane and affect position of the hyperplane. Regarding prediction problems, SVR estimates a function that is the center-line of a tube with radius of ε (Figure 1). The value of epsilon determines the width of the tube around the estimated function. In Figure 1, the predicted function is f (the central black line), which is in a tube (two black dotted lines) with radius of ε. SVR finds f with maximum distance from the tube. Data points outside this tube are called support vectors (green solid circles). Data points lie above the tube (ξ_i) or below the tube (ξ_i^*) receive penalty while the data fall inside the tube (green empty circles) are considered as true prediction and receive no penalty by the algorithm. ε-insensitive loss function is showed on the right side of Figure 1. Error or penalty considered for data points bigger than ε and error for values less than ε is equal to zero.

SVR is one of the machine learning algorithms that works based on statistical training theory. This method is a supervised learning algorithm that create a relationship between input data and dependent parameter due to minimizing structural risk. Structural Risk Minimization (SRM) is superior to Empirical Risk Minimization (ERM), which is used in Neural Network (NN). Algorithms based on SRM does not converge in local minimum (Yu & Kim, 2012). SVR predicts a function that is in the middle of a tube. The function has maximum distance from the tube body. SVR is based on Vapnik's ε-insensitivity loss function (Vapnik, 1995). In other words, the data located inside the tube receive no error, but the data out of the tube receive a penalization. This amount of error must be minimized. This type of algorithm is called ε -insensitive SVR. Figure 1 is an illustration of the mentioned explanations about SVR algorithm.

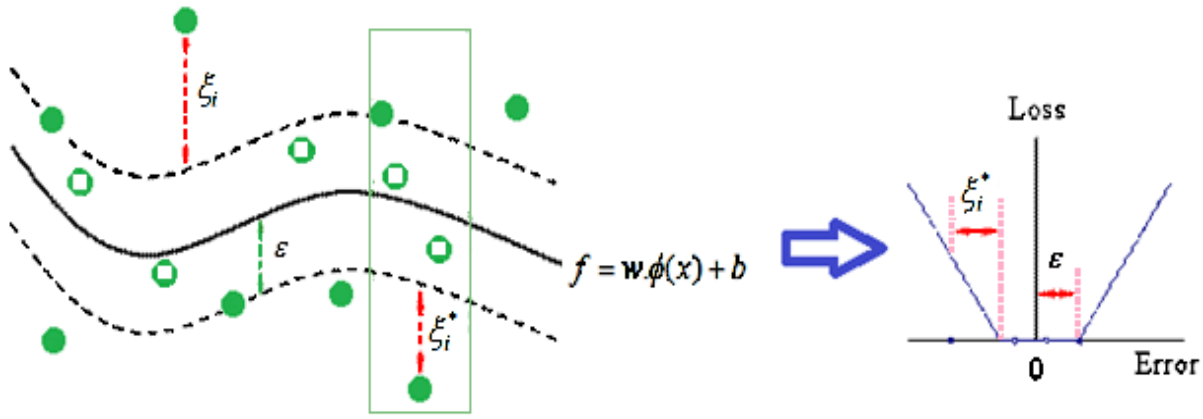

Figure 1. Illustration of non-linear SVR algorithm. Prediction function keeps the maximum distance from the data.

This distance is shown by ε. As the right part shows, data inside the ε-tube (dotted line) receive no error while the data out of the tube are penalized.

If we have L training data, x as input and y as related response, then, the final goal is to find a function that creates an appropriate relationship between input and output data.

$$f(x,w) = w^T x + b \tag{5}$$

To find a function f, parameters of w and b must be calculated. To find w and b the below relationship must be minimized (Vapnik, 1995).

$$R(c) = \frac{1}{2}\|w\|^2 + C\frac{1}{L}\sum_{i=1}^{L} L_\varepsilon(y_i, f_i(x,w)) \tag{6}$$

where C is a constant parameter and is identified by the user. The roll of C is to create a balance and to maximize the distance between data points and the function. L_ε is the Vapnik function and is described as below:

$$\begin{cases} 0 & if \quad |y - f(x,w)| \le \varepsilon \\ |y - f(x,w)| - \varepsilon & otherwise \end{cases} \tag{7}$$

The above formula is rewritten as maximizing the below formula:

$$L_p(\alpha_i, \alpha_i^*) = -\frac{1}{2}\sum_{i,j=1}^{L}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)x_i^T x_j - \varepsilon\sum_{i=1}^{L}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{L}(\alpha_i - \alpha_i^*)y_i \tag{8}$$

These conditions must be considered:

$$\begin{cases} \sum_{i=1}^{L}(\alpha_i - \alpha_i^*) = 0 \\ 0 \le \alpha_i \le C \quad i = 1,...,L \\ 0 \le \alpha_i^* \le C \quad i = 1,....,L \end{cases} \tag{9}$$

With solving Equation (9), function *f* can be obtained as follows:

$$f(x,w) = w_0^T x + b = \sum_{i=1}^{L}(\alpha_i - \alpha_i^*)x_i^T x + b \tag{10}$$

2-2. Fuzzy Support Vector Regression

Real data always contain noise and outliers. Concept of fuzzy sets first introduced by Zadeh (1965). Traditional mathematic concepts declare crisp numbers when a number belongs to a set or not. Thus there are only two choices representing 0 and 1 (Figure 2). However, fuzzy numbers receive degree of membership that means their presence in a set can be a degree between 0 and 1 interval [0, 1]. Different approaches are available to reduce the effect of noise in data. Modification of SVR to FSVR by mapping data into fuzzy system results in good results. In this approach, a coefficient of S_i is devoted to each data point. Figure 3 shows S_i for each log as input for the algorithm. In this case, each data point with S_i degree of membership has 1-S_i less contribution to the algorithm. In other words, with receiving different membership degrees, data find priority for their degree of importance. Outliers located farther compared to normal data receive less degree of contribution. Therefore, this can be helpful to prepare the data for making a more appropriate model using SVR algorithm (Le et al., 2009). S_i is calculated as follows:

$$S_i = \begin{cases} 1 - |X_+ - X_i|/(r_+ - \delta) \; where \; y_i = 1 \\ 1 - |X_- - X_i|/(r_- + \delta) \; where \; y_i = -1 \end{cases} \tag{11}$$

where X_+ is the mean of data points with symbol +1 and X_- is the mean of data points with symbol of -1. r_+ and r_- are as follow:

$$r_+ = \max|X_+ - X_i| \; where \; y_i = 1 \tag{12}$$

$$r_- = \max|X_- - X_i| \; where \; y_i = -1 \tag{13}$$

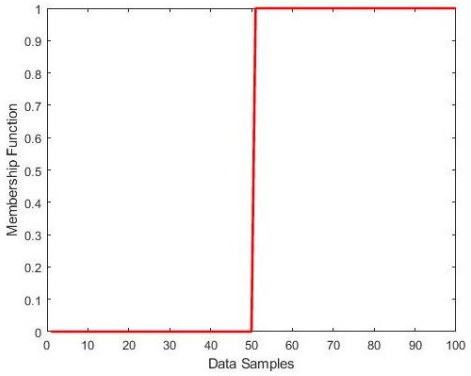

Figure 2. Displays on how membership function of crisp data has only two values 0 and 1. A crisp number can belong to a set or not.

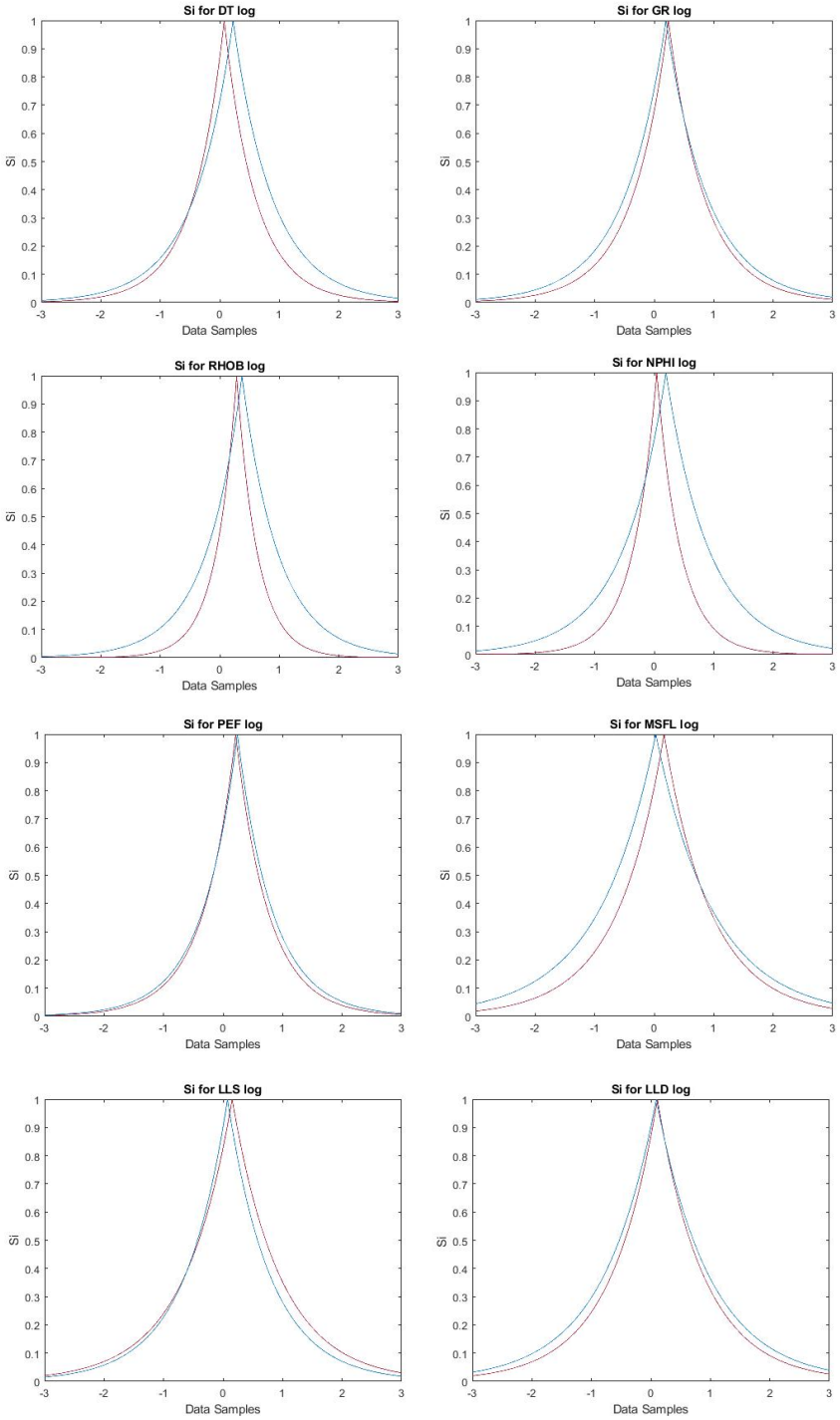

Figure 3. Different values of S_i coefficient for each log. It determines a Gaussian membership function for data point with labels +1 and -1.

3. Discussion and Results

3-1. Data Set

To carry out this study, we used data belonging to a gas reservoir located in the Persian Gulf. South Pars gas reservoir is located on the boundary of Iran and Qatar that is the largest gas reservoir in the world. The main formations containing gas devotes to Kangan and Upper Dalan. The composition of formations are mainly limestone, dolomite and anhydrites. Two wells are utilized in this study; one of them is for the training purpose and the other one is used for testing. Related well logs for water saturation estimation are chosen as DT, GR, RHOB, NPHI, LLD, LLS, MSFL, PEF and

water saturation from core data. In this study, water saturation has already been measured from core samples in the laboratory. In order to use SVR algorithm, we need to have input data and its related true response (output). Therefore, we used data points belong to well logs as input and core data at similar depth as true response to train SVR algorithm. Figure 4 is an illustration of different well logs near each other on the right side and the results of SVR and FSVR correlating on water saturation core data on the left side. As can be seen, each log changes over the depth due to lithology changes. Density log has decreased dramatically while NPHI has a slight reduction over the depth. Based on resistivity logs illustration, good separation of LLS and LLD from each other and from the MSFL represents a permeable formation containing hydrocarbon (if the amount of porosity reaches up to 15 percent, the hydrocarbon can be gas) (Asquith & Krygowski, 2004). In this interval amount of porosity changes between 6% up to 23%. The other sign of hydrocarbon presence in a formation is that if LLD has larger values than LLS and MSFL (LLD > LLS and MSFL) with good logs separation, then the area can be hydrocarbon zone, otherwise, if LLD has smaller values than LLS and MSFL (LLD < LLS and MSFL) the region can be a water zone (Schlumberger, 1991). Based on Figure 4, resistivity log LLD has smaller values compared to LLS and MSFL over the whole depth. Therefore, the observations confirm the existence of gas in this interval. Moreover, predicted values obtained from SVR and FSVR algorithms across core data show the superiority of FSVR to SVR.

3-2. Results

Data used in this study includes eight logs. Thus the input for SVR algorithm has eight features and the output of water saturation is obtained from core data. Coefficient of S_i is identified for data points with labels +1 and -1. Therefore, data points are turned into fuzzy sets in a [0, 1] interval by multiplication of coefficient S_i to each data point. Then, these data became ready as the input for SVR algorithm (Figure 5). Once more, this process is repeated with crisp data points (Figure 6).

As can be seen in Figures (5) and (6), water saturation predicted from SVR and FSVR algorithm is shown by dotted black line. Core data is displayed by a red dotted line. Determination coefficient obtained between SVR and core data is about 71% while the determination of coefficient is 95% between FSVR and core data. Therefore, fuzzy SVR has a better correlation with core data than SVR. As already mentioned, giving membership degrees to data samples accept the fact that data include uncertainty and imprecision. This fact can help the system with better results.

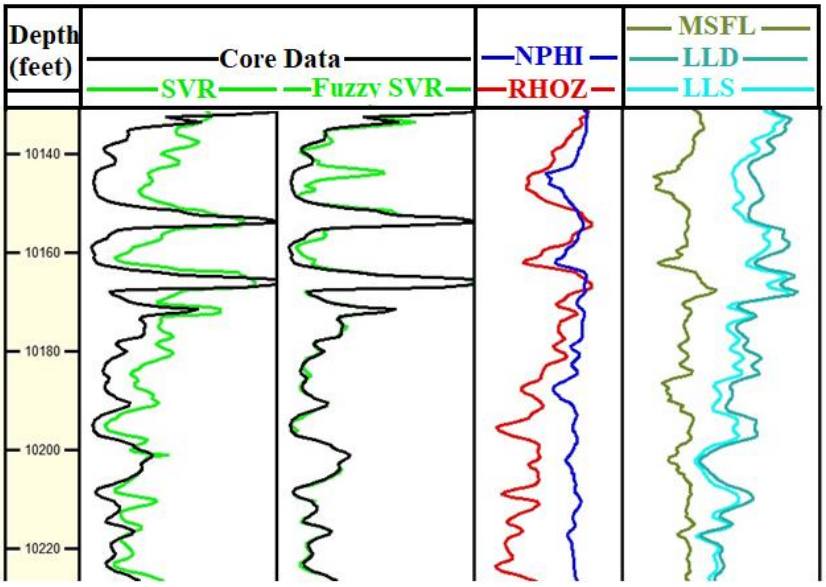

Figure 4. Illustration of various well logs over the reservoir zone. Predicted values obtained using SVR and Fuzzy SVR showed on the left side. Good separation of resistivity logs is a sign for permeable formation.

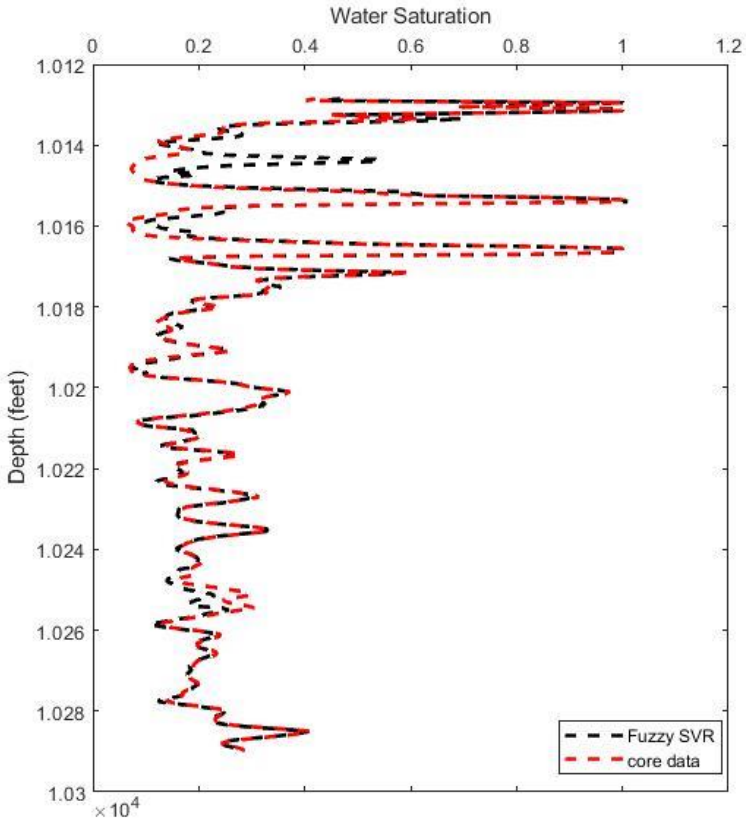

Figure 5. Water saturation prediction using FSVR; black dotted line shows FSVR output and red dotted line is the core data. Determination of coefficient between them is 95%.

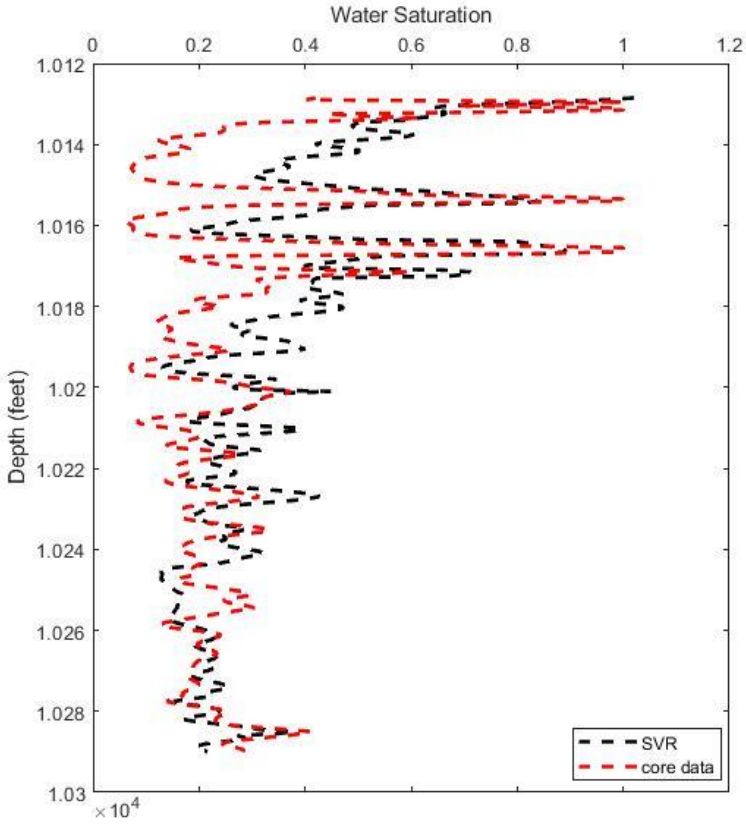

Figure 6. Water saturation prediction using SVR; black dotted line shows SVR output and red dotted line is the core data. Determination of coefficient between them is 71%.

SVR performance highly depends on several parameters (C, ε and γ or kernel type). Choosing appropriate parameters plays an important role on the accuracy of the model. There are three popular methods to choose optimal SVR parameters such as grid search (Hsu et al., 2003), gradient descent (Keerthi et al., 2007) and meta-heuristics algorithms (Blum & Roli, 2003; Talbi, 2009). In grid search method, different models are tested in a selected range of parameters. This evaluation takes place through cross-validation. The role of cross-validation is to measure how a model generalizes itself to an independent dataset. Cross-validation is the process to find an optimized model during k training steps; at first, data is divided into two sets of training and test parts. Partitions are equal in size which is called fold. The model is evaluated by test data in each step. This process is repeated k times until an optimized model is obtained. The model with highest accuracy is chosen as the best and the related parameters are the optimal parameters. For a better clarification, Figure 7 shows data set division and folds.

Figures (8) and (9) show the concentration of predicted data from SVR and FSVR and core data along a 45-degree line. Each figure has two sides; the left side is the scatter plot of the predicted data versus core data using linear SVM and the right-side scatter plot is related to medium Gaussian kernel as can be seen. As the figure shows, for FSVR, the concentration of data points along the 45-degreeline is better compared to scatter plot of medium Gaussian SVM. In Figure 9, for SVR, data points are concentrated in the left corner for linear SVM while for medium Gaussian SVM data points are randomly scattered toward right side with a less massive gathering in the left corner. Totally, data distribution along 45-degree line is more focused in FSVR compared to SVR.

Figure 7. K-fold cross-validation with k as 5.

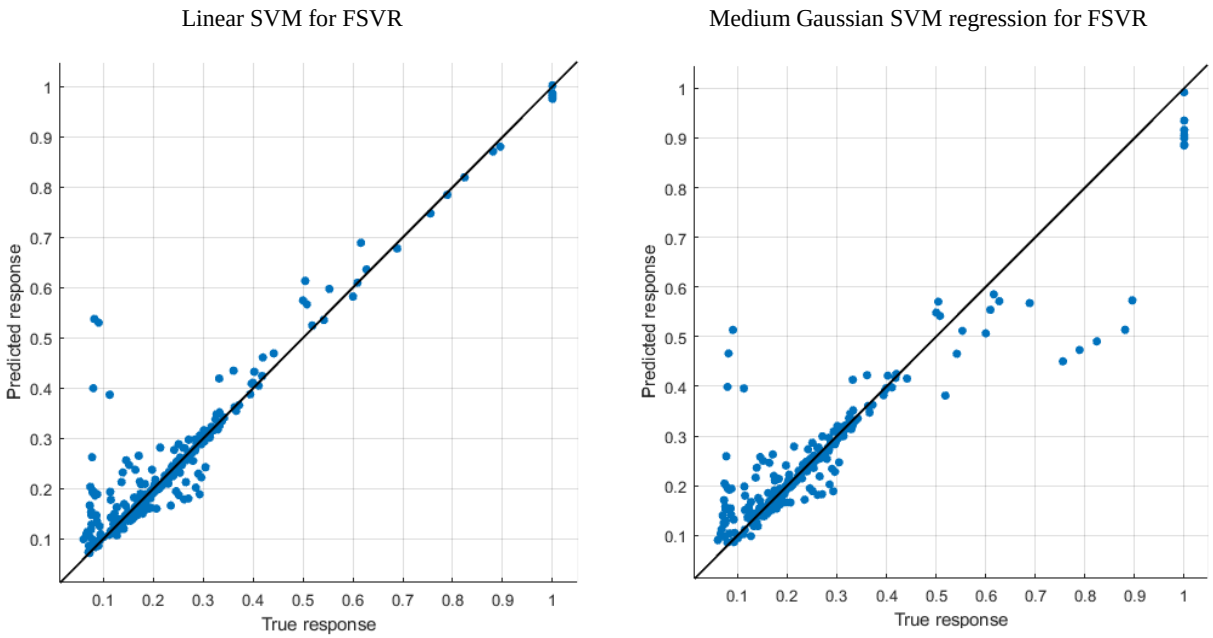

Figure 8. Evaluation of concentration of predicted water saturation data points obtained from FSVR and core data along a 45-degree line.

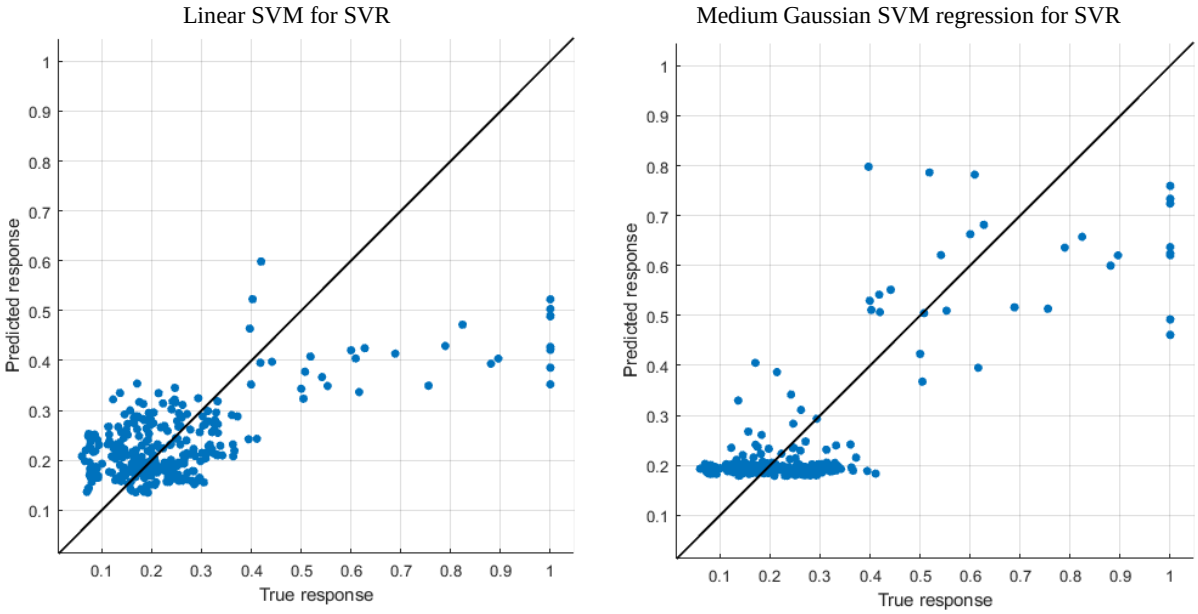

Figure 9. Evaluation of concentration of predicted water saturation data points obtained from SVR and core data along a 45-degree line.

4. Conclusion

Noise and outliers are inseparable part of well data. Well log measurements are always contaminated with noise and uncertainty that remain in data even after denoising and processing operations. Random noise is one of them which cannot easily get removed from data. Obviously, the presence of noise is able to ruin the results and lower the quality of output models. Applying degree of membership to each data point makes the system be able to consider imprecision and uncertainty of data. Consequently, this paper shows the modification of a well-known algorithm, SVR, to Fuzzy SVR yields better results in predicting water saturation in a gas reservoir in the south of Iran. Applying fuzzy functions on data boosts the robustness of model so that it is worth trying this method in more complicated well sets.

References

Adeniran, A., Elshafei, M., & Hamada, G. (2009). Functional network soft sensor for formation porosity and water saturation in oil wells, Instrumentation and Measurement Technology Conference, I2MTC '09. *IEEE, Singapore*, 113-1143.

American Petroleum Institute. (1998). Recommended Practices for Core Analysis Handbook, API publications, http://www.api.org.

Archie, G.E. (1942). The electrical resistivity log as an aid in determining some reservoir characteristics. *Petroleum Transactions of AIME*, 146, 54-62.

Asquith, G., & Krygowski, D. (2004). AAGP Methods in Exploration, No. 16, Chapter 5: Resistivity Logs. Pages 77-101.

Bagheripour, P., & Asoodeh, M. (2014). Genetic implanted fuzzy model for water saturation determination. *Journal of Applied Geophysics*, 103, 232-236.

Blum, C., & Roli, A. (2003). Metaheuristics in combinatorial optimization: overview and conceptual comparison. *ACM Computing Surveys (CSUR)*, 35, 268-308.

Hsu, C.W., Chang, C.C., & Lin, C.J. (2003). A practical guide to support vector classification.

Jafari Kenari, S.A., & Mashohor, S. (2013). Robust committee machine for water saturation prediction. *Journal of Petroleum Science and Engineering*, 104, 1-10.

Jia, J., Ke, Sh., Li, J., Kang, Zh, Ma, X., Li, M., & Cue, J. (2020). Estimation of Permeability and Saturation Based on Imaginary Component of Complex Resistivity Spectra: A Laboratory Study. *Open Geoscience*, 12, 299-306.

Keerthi, S.S., Sindhwani, V., & Chapelle, O. (2007). An efficient method for gradient based adaptation of hyperparameters in SVM models, in: *Advances in neural information processing systems*, 673-680.

Le, V.H., Liu, F., & Tran. D.K. (2009). Fuzzy Linguistic Logic Programming and its Applications. *Theory and Practice of Logic Programming (TPLP).*

Li, Z., Xie, Y., Li, X., & Zhao, W. (2021). Prediction and application of porosity based on support vector regression model optimized by adaptive dragonfly algorithm. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 43(9), 1073-1086.

Lim, K.M., Sim, Y.C., & Oh, K.W. (2002). A Face Recognition System Using Fuzzy Logic and Artificial Neural Network", IEEE. [1992 Proceedings] IEEE International Conference on Fuzzy Systems, USA.

Luthi, E.M. (2001). Geological well logs, their use in reservoir modeling, Springer-Verlag Berlin Heidelberg, Germany.

Okwu, M., & Nwachukwu, A.N. (2019). A review of fuzzy logic applications in petroleum exploration, production and distribution operations. *Journal of Petroleum Exploration and Production Technology*, 9(2), 1555–1568.

Pickett, G.R. (1966). A Review of Current Techniques for Determination of Water Saturation from Logs, SPE 1446.

Schlumberger. (1991). Log Interpretation Principles/Applications. Schlumberger Wireline & Testing, SMP- 7017, Sugar Land, Texas.

Talbi, E.G. (2009). *Methaheuristics: from Design to Implementation.* Volume 74. John Wiley & Sons.

Tonstad, S.L., Boordsen, H., & Ringen, J.K. (1990). Alternative Methods for Determining Water Saturation in Core Plugs, Advances in Core Analysis, Gorden and Breach, Sc., 345-409.

Vapnik, V.N. (1995). The Nature of Statitical learning Theory. Springer, New York.

Walther, H.C. (1967). Saturation from Logs-Laberatory Measurements of Logging Parameters, SPE 42nd Annual Meeting Huston, Tex. Oct, 1-4.

Yu, H., & Kim, S. (2012). SVM tutorial: classification, regression, and ranking, Handbook of Natural Computing. *Springer Berlin Heidelberg*, 479-506.

Zadeh, L.A. (1965). Fuzzy Sets, *Information and Control*, 8(3), 338–353.

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., Liu, X., & Zhuang, J. (2018), Modeling and Simulating of Reservoir operation Using the Artificial Neural Network, Support Vector Regression, Deep Learning Algorithm. *Journal of Hydrology*, 565, 720-736.