UDC 004.85 *Original article*

https://doi.org/10.23947/2587-8999-2023-6-1-70-76

Machine learning in the analysis of the electromagnetic field influence on the rate of oilfield equipment's corrosion and salt deposition

Sh. R. Khusnullin[1], K. F. Koledina[1,2], S. R. Alimbekova[3], F. G. Ishmuratov[3]

[1] Ufa State Petroleum Technical University, 1, Kosmonavtov St., Ufa, Russian Federation

[2] Institute of Petrochemistry and Catalysis of Russian Academy of Sciences, 141, October Ave, Ufa, Russian Federation

[3] Ufa University of Science and Technology, 12, Karl Marx St., Ufa, Russian Federation

shamil.khusnullin@gmail.com

Abstract

Introduction. The formation of salt deposits and oilfield equipment's corrosion in most oil fields has become particularly relevant due to the increase in the volume of oil produced and the increase in its water content in recent years. The deposition of salts in the formation and wells leads to a decrease in the permeability of the oil reservoir, the flow rate of wells. The aim of the work is to use machine learning algorithms to simulate the effects of an electromagnetic field on the processes of salt deposition and corrosion. Prediction of experimental results will allow faster and more accurate experiments to establish the influence of electromagnetic fields on the processes of corrosion and salt deposition.

Materials and methods. Three groups of data were used, to train the models, differing in the composition of the studied initial model salt solution: the waters of the Vyngapurovsk's and Priobsk's deposits, as well as tap water. The following machine learning models were used: linear regression with Elastic-Net regularization, the k-nearest neighbors algorithm, the decision tree, the random forest and a fully connected neural network.

Results. It was found that the decision tree and the random forest have the best accuracy of predictions, from the experiments conducted. Python program has been developed to predict the output results of experiments. Modeling with various models and their parameters is carried out.

Discussion and conclusions. From the experiments conducted, it was found that the decision tree and the random forest have the best accuracy of predictions. Neural networks, on the contrary, predict with the least accuracy. This is due to the fact that there is too little data in the training samples. With the increase in the number of observations, it is worth using neural networks of various architectures.

Keywords: salt deposition, electromagnetic impact, oilfield equipment's corrosion, multiple regression methods, machine learning, neural network.

For citation. Machine learning in the analysis of the electromagnetic field influence on the rate of oilfield equipment corrosion and salt deposition / Sh. R. Khusnullin, K. F. Koledina, S. R. Alimbekova, F. G. Ishmuratov // Computational Mathematics and Information Technologies. — 2023. — Vol. 6, no. 1. — P. 70–76.
https://doi.org/10.23947/2587-8999-2023-6-1-70-76

Научная статья

Машинное обучение в анализе влияния электромагнитного поля на скорость коррозии и солеотложения нефтепромыслового оборудования

Ш. Р. Хуснуллин[1], К. Ф. Коледина[1,2], С. Р. Алимбекова[3], Ф. Г. Ишмуратов[3]

[1]Уфимский государственный нефтяной технический университет, Российская Федерация, г. Уфа, ул. Космонавтов, 1

[2]Институт нефтехимии и катализа УФИЦ РАН, Российская Федерация, г. Уфа, пр. Октября, 141

[3]Уфимский университет науки и технологий, Российская Федерация, г. Уфа, ул. Карла Маркса, 12

shamil.khusnullin@gmail.com

Аннотация

Введение. Образование солеотложений и коррозия нефтепромыслового оборудования на большинстве нефтяных месторождений в последние годы получила особую актуальность ввиду роста объемов добываемой нефти и увеличения ее обводненности. Отложение солей в пласте и скважинах приводит к снижению проницаемости нефтеносного пласта, дебита скважин. Целью работы является применение алгоритмов машинного обучения для моделирования воздействия электромагнитного поля на процессы солеотложений и коррозии. Предсказание результатов экспериментов позволит быстрее и точнее проводить опыты, устанавливающие влияние электромагнитных полей на процессы коррозии и отложения солей.

Материалы и методы. Для обучения моделей были использованы три группы данных, различающихся по составу изучаемого исходного модельного солевого раствора: воды Вынгапуровского и Приобского месторождений, а также водопроводная вода. Были рассмотрены следующие модели машинного обучения: линейная регрессия с регуляризацией Elastic-Net, метод *k* ближайших соседей, дерево решений, случайный лес и полносвязная нейросеть.

Результаты исследования. С помощью алгоритмов машинного обучения были смоделированы процессы воздействия электромагнитного поля на образования солеотложений и коррозию нефтепромыслового оборудования. Разработана программа на Python для предсказания выходных результатов экспериментов. Проведено моделирование с различными моделями и их параметрами.

Обсуждение и заключение. Из проведенных экспериментов установлено, что наилучшую точность предсказаний имеют дерево решений и случайный лес. Нейронные сети, напротив, предсказывают с наименьшей точностью. Связано это с тем, что данных в обучающих выборках слишком мало. С увеличением числа наблюдений стоит использовать нейросети различных архитектур.

Ключевые слова: солеотложение, электромагнитное воздействие, коррозия нефтепромыслового оборудования, методы множественной регрессии, машинное обучение, нейронная сеть, случайный лес.

Для цитирования. Машинное обучение в анализе влияния электромагнитного поля на скорость коррозии и солеотложения нефтепромыслового оболрудования / Ш. Р. Хуснуллин, К. Ф. Коледина, С. Р. Алимбекова, Ф. Г. Ишмуратов // Computational Mathematics and Information Technologies. — 2023. — Т. 6, № 1. — С. 70–76. https://doi.org/10.23947/2587-8999-2023-6-1-70-76

Introduction. The problem of the salt deposits formation and oilfield equipment's corrosion is particularly acute in most of the actively developed oil fields in recent years. This is due to an increase in oil production and an increase in its water content. The deposition of salts in the formation and wells leads to a decrease in the permeability of the oil-bearing formation, the flow rate of wells, an increase in operating costs and the failure of deep-pumping equipment [1].

Experiments were conducted to study the effect of the electromagnetic field generated by the resonance-wave complex on the corrosion of structural low-carbon steel and on the crystallization of calcium carbonate at the Pilot Research Center. It was found that exposure to EMF reduces the total mass of poorly soluble salts, and also provides protection against corrosion [2, 3].

In this paper, the use of machine learning algorithms is used to simulate the effects of EMF on the crystallization of calcium carbonate and on the corrosion process of structural steel. The object of the study was the influence of electromagnetic fields on the processes of salt deposition and corrosion. The subject of study and analysis is the possibility of using machine learning models to simulate the processes of salt deposition and corrosion under the influence of electromagnetic fields. The purpose of this study is to develop software for modeling experiments on the interaction of electromagnetic fields on salt deposits and corrosion. The forecast of the results will make it easier and faster to conduct such experiments. This will make it possible to more accurately determine the effect of the magnetic field on calcium carbonate deposits and on the oilfield equipment's corrosion rate.

Materials and methods. Experimental studies of the electromagnetic effect on the processes of salt deposition and corrosion were carried out for the water of the Vyngapurovsky and Priobsky deposits, as well as for tap water [4]. To train the algorithms, 3 groups of data were used, differing in the composition of the studied initial model salt solution. The analyzed data includes information about the composition of solutions, research conditions (flow rate, pressure, temperature), parameters of electromagnetic fields acting on solutions. Output parameters to be predicted: corrosion rate in the flow and static solution, mm/year; distribution of calcium carbonate by morphology (calcite, aragonite, vaterite). Different models were used for each group of reports, since each task has different input and output data [5].

To solve the problem of machine learning in the analysis of the influence of the electromagnetic field on the rate of corrosion and salt deposition of oilfield equipment, the learning algorithms discussed below were applied.

1. Linear regression with Elastic Net regularization.

It is a linear relationship between the target variable and the features:

$$\hat{y} = w_0 + w_1 * x_1 + \ldots + w_D * x_D = <x, w> + w_0$$

for finding optimal weights that minimize the root-mean-square loss function [6]:

$$MSE = \frac{1}{N} * \sum_i^N \left(y^i - \hat{y}^i\right)^2. \tag{1}$$

The gradient descent method is used in practice.

Elastic Net regularization method is used to combat retraining. The 1st and 2nd weight norms are added to the loss function *w:*

$$L(y, \hat{y}, w) = \frac{1}{N} * \sum_i^N \left(y^i - \hat{y}^i\right)^2 + \lambda_1 \|w\|_1 + \lambda_2 \|w\|_1^2, \tag{2}$$

where λ_1 and λ_2 are the regularization coefficients.

2. *k*-nearest neighbors method (*k-NN*).

The essence of the method is as follows: the prediction $\hat{y}$ for object is calculated as the average value of the target variable *y* among its *k* nearest neighbors [7]:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{k} y^i. \tag{3}$$

Various metrics (distance functions) are used to find the distance between objects:

− Euclidean distance: $\rho(a,b) = \sqrt[2]{\sum_i (a_i - b_i)}$; (4)

− manhattan metric: $\rho(a,b) = \sum_i |a_i - b_i|$; (5)

− cosine distance: $\rho(a,b) = 1 - \frac{a * b}{|a||b|}$. (6)

3. Decision tree.

The decision tree predicts the value of the target variable by applying a sequence of simple decision rules (which are called predicates). There is a predicate in each node of this tree. If it is correct for the current sample from the selection, the transition is made to the right child, if not, to the left. Often predicates are simply taking a threshold by the value of some attribute. Predictions are recorded in the leaves (for example, the values of the target variable *y*).

4. Random forest.

This is an ensemble of models (a composition of several algorithms), where decision trees are used as the basic algorithm. This method is based on bagging (meta algorithm designed to improve the stability of the solution) [8]. The essence of the method is as follows: from the initial sample, subsamples of the same dimension are obtained by random selection of objects. A decision tree is trained on each sample, and not all features of objects are used, but a random number of them (the method of random subspaces). To get one prediction, the predictions of all models are averaged:

$$\hat{y}(x) = \frac{1}{k}(b_1(x) + \ldots b_k(x)). \tag{7}$$

5. Neural network.

Networks of the following architecture were used in all three tasks: the input layer from D neurons (number of input parameters), a hidden layer and an output layer consisting of m neurons, according to the dimension of the target vector $y = (y_1,..., y_m)$ [9].

The activation function was used to ensure the nonlinearity of transformations, the activation function was used:

$$ReLU(x) = \max(x, 0). \quad (8)$$

6. Model quality assessment.

The following criteria (metrics) are used in regression tasks in order to determine which model best approximates the relationship between features x and dependent variables y:

– RMS error: $MSE = \frac{1}{N}\sum_{i=1}^{N}(y-\hat{y})^2$; (9)

– coefficient of determination: $R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}$; (10)

– mean absolute error: $MAE = \frac{1}{N}\sum_{i=1}^{N}|y-\hat{y}|$. (11)

Research results. Tables 1–3 show the results of using models for three groups of experiments.

Table 1

Metric values for the Vyngapurovsk field

Model/ Metric	Linear Regression	*k-NN*	**Decision tree**	Random forest	Neural network
MSE	0.01453	0.0133	**0.0064**	0.0065	20.6408
MAE	0.0839	0.0757	**0.0467**	0.0478	3.2109
R^2	–5.2844	–8.2677	**0.3169**	0.2266	–1.0221

Table 2

Metric values for the Priobsk field

Model/ Metric	Linear Regression	*k-NN*	**Decision tree**	Random forest	Neural network
MSE	0.3073	0.2623	**0.1435**	0.1724	13.9461
MAE	0.2489	0.1972	**0.1314**	0.1724	2.4746
$R2$	–18.8724	–2.1167	**0.5364**	0.1113	–0.9355

Table 3

Metric values for tap water

Model/ Metric	Linear Regression	*k-NN*	**Decision tree**	Random forest	Neural network
MSE	0.3232	0.2525	**0.1799**	0.1845	1.1029
MAE	0.4729	0.4022	**0.3251**	0.3362	0.8847
R^2	–6.9638	–1.7114	**0.0378**	–0.0283	–4.3788

Based on the described algorithms, a Preditct program has been developed — modeling the effects of EMF on the processes of salt deposition and corrosion on oilfield equipment [10] (Fig. 1).

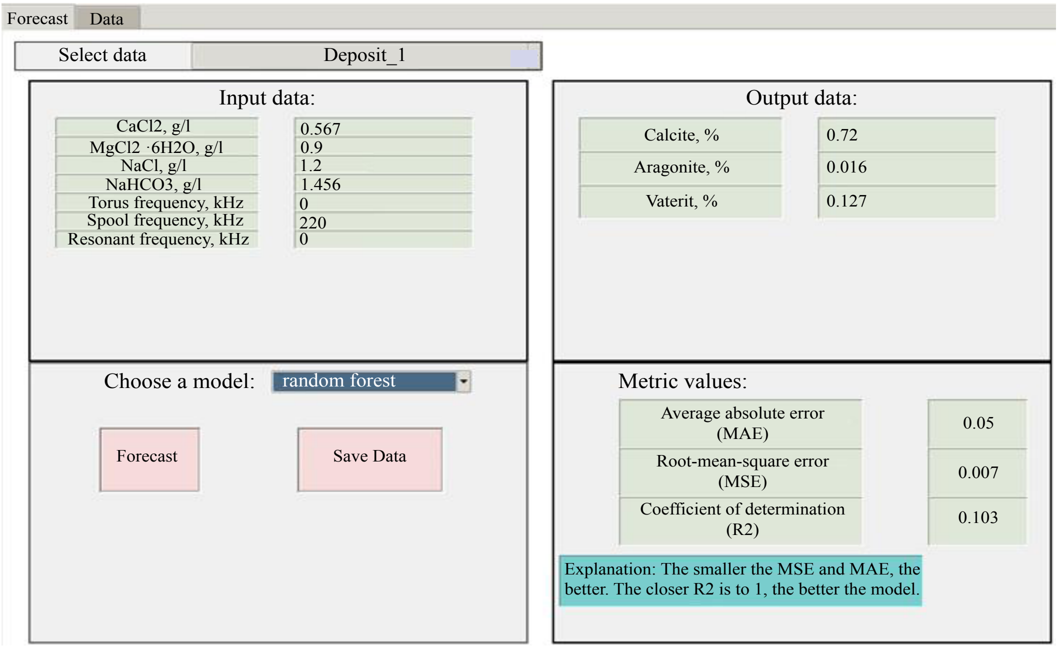

Fig. 1. The interface of the program "Predict" in Russian

The user can enter various input data (the composition of the solution, the frequency of electromagnetic fields, the presence of magnets, etc.) and get output data: the distribution of calcium carbonate by morphology and / or the rate of corrosion. The following machine learning methods can be used for prediction, at the user's choice: linear regression, the k nearest neighbors method, a decision tree, a random forest and a neural network. To select the best model, the program outputs the readings of quality metrics. Python version 3.10 was chosen as the programming language. Numpy, pandas, and sklearn libraries were used to use machine learning and data processing models. The openpyxl library was used to work with excel files.

Figure 2 shows graphs of the true and predicted values of the output parameters for the Priobsk field obtained by the random forest method. The corrosion rate in the flow (true and predicted value); the distribution of calcium carbonate in the form of calcite (true and predicted value) are considered. As can be seen from the graph above, the random forest algorithm approximates the output data quite well.

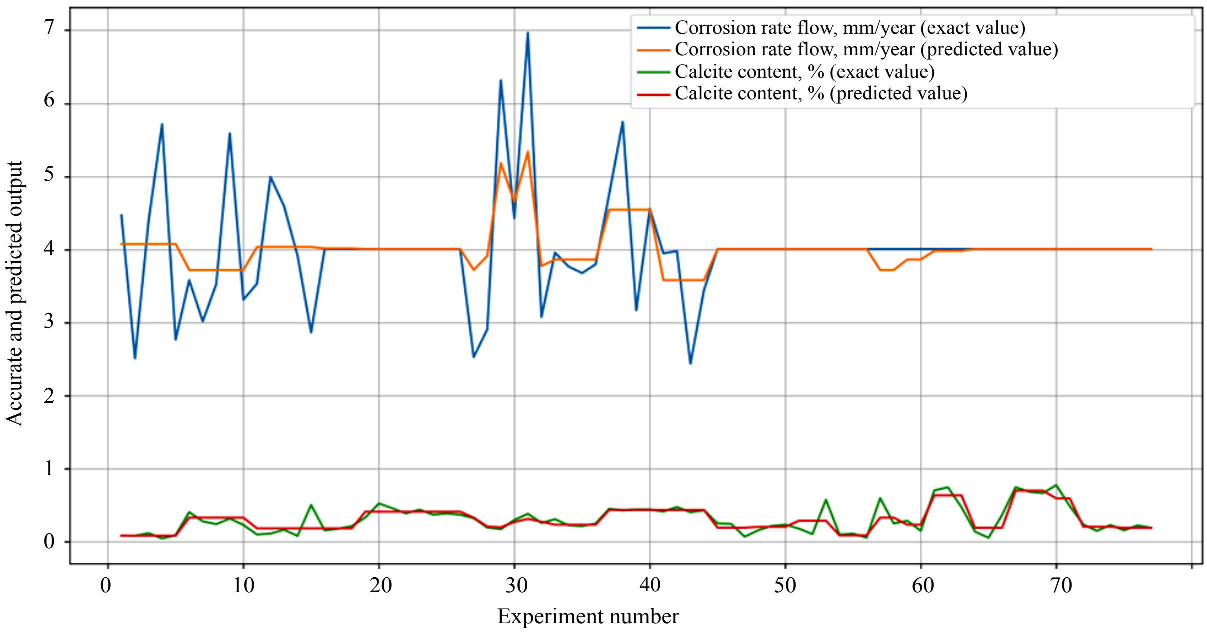

Fig. 2. Graph of true and predicted values by the random forest method

Discussion and conclusions. The decision tree and the random forest have the lowest values of the MSE and MAE metric and the values of R^2 closest to 1 in all three problems, as can be seen from the results (Tables 1−3, Fig. 2). Neural networks, on the contrary, have the worst error rates. This is due to the fact that there is too little data for training.

To analyze such a number of data, it is advisable to use decision trees and a random forest. With an increase in the number of observations, it is worth switching to the use of neural networks, as well as using other architectures (with a large number of hidden layers, recurrent networks, etc.).

References

1. Koledina, K. F. Solving the problem of multi-criteria optimization of the synthesis reaction of benzylalkyl esters by the method of "ideal" point and lexicographic ordering / K. F. Koledina, A. A. Alexandrova // Computational mathematics and information technologies. — 2022. — Vol. 1, no. 1. — P. 12–19. https://doi.org/10.23947/2587-8999-2022-1-1-12-19

2. Shaimardanova, G. F. Genetic algorithm for solving the inverse problem of chemical kinetics / G. F. Shaimardanova, K. F. Koledina // Computational mathematics and information technologies. — 2022. — Vol. 1, no. 1. — P. 41–49. https://doi.org/10.23947/2587-8999-2022-1-1-41-49

3. The influence of the electromagnetic field on the corrosion of low-carbon steel in aqueous mineralized media and on the crystallization of calcium carbonate in the presence of iron(II) ions / S. R. Alimbekova, F. G. Ishmuratov, V. V. Nosov [et al.] // SOCAR Proceedings Special Issue. — 2021. — No. 1. — P. 116−124. http://dx.doi.org/10.5510/OGP2021SI100515

4. Physical methods of preventing salt deposition during oil production / S. R. Alimbekova, R. N. Bakhtizin, A. I. Voloshin [et al.] // Oil and gas business. — 2019. — No. 6. — P. 31–38. http://ngdelo.ru/files/ngdelo/2019/6/ngdelo-6-2019-p31-38.pdf

5. Weisberg, S. Applied linear regression / S. Weisberg — New Jersey: John Wiley & Sons, 2005. — P. 352.

6. Khusnullin, S. R. Machine learning in the analysis of the influence of the electromagnetic field frequency at different solution flow rates on the corrosion rate / S. R. Khusnullin, K. F. Koledina, S. R. Alimbekova // In the collection: Technical and technological systems. Materials of the Thirteenth International Scientific Conference. — Krasnodar, 2022. — P. 462–464.

7. Sun S., Huang R. An adaptive k-nearest neighbor algorithm // 2010 seventh international conference on fuzzy systems and knowledge discovery. — IEEE, 2010. — Vol. 1. — P. 91−94.

8. Chistyakov, S. P. Random forests: an overview // Proceedings of the Karelian Scientific Center of the Russian Academy of Sciences. — 2013. — No. 1. — P. 117−136.

9. Gorbachevskaya, E. N. Classification of neural networks // Bulletin of the Volga State University. VN Tatishcheva. — 2012. — No. 2 (19). — P. 128−134.

10. Program for modeling the effects of electromagnetic field on the processes of salt deposition and corrosion on oilfield equipment (Predict). Khusnullin Sh. R., Koledina K. F., Alimbekova S. R. [et al.] Certificate of registration of computer programs. — No. 2023611354, dated January 19.

Received by the editorial office 08.02.2023.
Received after reviewing 01.02.2023.
Accepted for publication 02.03.2023.

About the Authors:

Khusnullin, Shamil R., student, Ufa State Petroleum Technological University (1, Kosmonavtov St., Ufa, 450064, RF), ORCID, shamil.khusnullin15@gmail.com

Koledina, Kamila F., Dr.Sci. (Phys.-Math.), Associate Professor, Ufa State Petroleum Technological University (1, Kosmonavtov St., Ufa, 450064, RF), researcher, Institute of Petrochemistry and Catalysis, Russian Academy of Sciences (141, Oktyabrya prospect, Ufa, 450075, RF), ORCID, koledinakamila@mail.ru

Alimbekova, Sofya R., Senior Researcher, Pilot Research Center, Ufa University of Science and Technology (12, Karl Marx St., Ufa, 450008, RF), ORCID, ms.sofia.al@gmail.com

Ishmuratov, Farid G., head of the laboratory, Ufa University of Science and Technology (12, Karl Marx St., Ufa, 450008, RF), ORCID, farid_ishmuratov@mail.ru

Claimed contributorship

Shamil Ramilevich Khusnullin: data preparation, machine learning algorithms training. Koledina Kamila Feliksovna: consultation on multiple regression methods. Alimbekova Sofya Robertovna: providing source data. Ishmuratov Farid Gumerovich: providing source data.

Conflict of interest statement

The authors declare that there is no conflict of interest.

All authors have read and approved the final version of the manuscript.