Research Article

Qiuyan Ji, Feilong Han*, Wei Qian, Qing Guo, and Shulin Wan

A machine learning-driven stochastic simulation of underground sulfide distribution with multiple constraints

https://doi.org/10.1515/geo-2020-0274
received June 23, 2020; accepted July 06, 2021

Abstract: The increase of sulfide (S^{2-}) during the water flooding process has been regarded as an essential and potential risk for oilfield development and safety. Kriging and stochastic simulations are common methods for assessing the element distribution. However, these traditional simulation methods are not able to predict the continuous changes of underground S^{2-} distribution in the time domain by limited known information directly. This study is a kind of attempt to combine stochastic simulation and the modified probabilistic neural network (modified PNN) for simulating short-term changes of S^{2-} concentration. The proposed modified PNN constructs the connection between multiple indirect datasets and S^{2-} concentration at sampling points. These connections, which are treated as indirect data in the stochastic simulation processes, is able to provide extra supports for changing the probability density function (PDF) and enhancing the stability of the simulation. In addition, the simulation process can be controlled by multiple constraints due to which the simulating target has been changed into the increment distribution of S^{2-}. The actual data test provides S^{2-} distributions in an oil field with good continuity and accuracy, which demonstrate the outstanding capability of this novel method.

Keywords: machine learning, stochastic simulation, sulfide, chemical estimation, multiple constraints

1 Introduction

With water flooding development in the oilfield, a series of security issues have been caused by the continuous increase of sulfide (S^{2-}) concentration. Obtaining the S^{2-} distribution is an essential foundation of taking precautions for high-benefit, low-risk, and long-term oilfield development. Previous studies have found that whether it is initially injected into seawater or freshwater (gradually back-injection the produced water), varying degrees of reservoir souring will occur in water flooding development [1]. Sulfate-reducing bacteria (SRB) play an important role in the generation of S^{2-}, which is the main reason for reservoir souring during water flooding [1–3]. In the process of dissimilatory sulfate reduction, SRB typically uses organic components as the electron donors, uses sulfate (SO_4^{2-}) as the terminal electron acceptor for respiration, and generates energy with the production of sulfide [2,3]. The existing forms of the sulfide ion (H_2S_{aq}, HS^-, and S^{2-}) in the aqueous solution will convert to each other depending on the pH of the solution, where fraction coefficients are commonly calculated from the pH of the solution and the ionization constants [4]. In oilfield systems, the concentration of sulfide in the water phase is usually tested by the iodometric method, which includes the sum of these forms. Hence, the sulfide represents the sum of three forms in the water phase with the chemical formula S^{2-} in this study.

Many negative effects have been caused by the continuous increase of sulfide concentration in the oilfield system. Corrosion of mild steel is a common effect in the environment of H_2S aqueous solution, which causes great economic losses in the oilfield system [5]. A possible

* **Corresponding author: Feilong Han,** Hohai University, School of Earth Sciences and Engineering, Nanjing 211000, China; School of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, China, e-mail: flhan@foxmail.com
Qiuyan Ji: Hohai University, School of Earth Sciences and Engineering, Nanjing 211000, China, e-mail: wqy123@hhu.edu.cn
Wei Qian: Hohai University, School of Earth Sciences and Engineering, Nanjing 211000, China, e-mail: wei.geoserve@gmail.com
Qing Guo: Department of Physics, Michigan Technological University, Houghton, Michigan 49931, United States of America, e-mail: qinguo@mtu.edu
Shulin Wan: Department of Physics, Michigan Technological University, Houghton, Michigan 49931, United States of America, e-mail: swan@mtu.edu

mechanism can be simply written as Fe(s) + H_2S → FeS(s) + H_2 [6]. In addition to the corrosion of petroleum pipelines and storage tanks, the reservoir souring due to S^{2-} generation, the plugging of machinery, and rock pores caused by the precipitation biomass and amorphous ferrous sulfide are also critical problems for oilfield development [7]. Besides, H_2S is soluble in water, alcohol, crude oil, and so on, which the solubility is affected by temperature and other environmental factors. The water solubility of H_2S at 20°C is 1 g in 242 mL [8]. In the late stage of water flooding development, the water content is higher than 80% and even more than 90% in most oilfields. In such an environment of temperature or other factors change, releasing H_2S gas from the water phase is worth to be concerned. For centuries, H_2S is known as toxic acidic gas. However, it has been reported that H_2S, as an endogenous gaseous mediator, might show a certain regulatory effect in neurotransmission, cardiovascular function, and cell metabolism in recent years [9]. There is no doubt that H_2S is harmful once it exceeds the safe concentration. According to the survey, we can smell and recognize H_2S with a characteristic odor of rotten eggs at a concentration of about 11 μg/m^3, and it can be fatal for a few breaths at 700 mg/m^3, where the conversion factors for H_2S in air (20°C, 101.3 kPa) is 1 mg/m^3 = 0.71 ppm [8]. Therefore, the release of H_2S gas from the liquid is another significant risk, once the concentration of H_2S_{aq} is overstandard. In general, the concentration of S^{2-} in the late stage of water flooding, attributed by SRB, is a threat to human life and generates serious pipeline corrosion. Also, the increased rate of S^{2-} concentration usually varies with different oil wells at different times during the water flooding process [10], and its value is mainly achieved by sample testing. The changeable and lateness information brings an unknown risk to safe production. Hence, it is important to study and predict the underground S^{2-} spatial distribution in the water phase.

The increase of sulfide is a concomitant problem in the late stage of water flooding development. So far, numerous studies have been done to reveal the mechanism of S^{2-} formation and inhibit the generation of S^{2-} during the water flooding process, which mainly includes the studies of growth kinetics, substrate utilization and metabolism of SRB, and chemical and biological inhibitors [11–14]. Many researchers have simulated the transport of S^{2-} to distinguish the effect of different inhibition methods in the lab scale and have applied the laboratory models to the simplified oil reservoir [15–17]. These simulations are based on biokinetic mechanisms and reactive transport models, in which S^{2-} is a product of SRB. Achieving precise simulations relies on various physical, chemical, and biokinetic parameters, while these parameters generally show dynamic instability behavior in a realistic oil developing area. Therefore, more sophisticated models than laboratory need to be considered for simulating the realistic S^{2-} distribution in the oilfield. In terms of estimating the spatial distribution of variables, geostatistics is an alternative or a supplement method to realize the element spatial distribution. A growing number of researchers have successfully applied geostatistics to investigate the spatial distribution of trace elements or other variables in soil and environmental science [18–22]. What is more, some researchers have realized the applications of geostatistics on reservoir modeling, porosity spatial distribution, and lithological distribution in the oilfield development [23,24]. Geostatistics is a common and effective tool for exploring and analyzing uncertain phenomena, which estimates unknown points by the sampling points. Currently, it has been widely applied in numerous fields, e.g., mining, oilfield exploration, geography, environment science, and soil science field [24–26]. Generally, according to the data types and the simulation target, both Kriging and stochastic simulations are widely used as geostatistics methods. The cognition of spatial distribution can be realized by Kriging with only the estimation result [27]. However, the ordinary Kriging method tends to smooth out the spatial variation of the unknown attributes [27]. Therefore, this method is more applicable for estimating geological parameters with gentle changes. Stochastic simulation is another geostatistical simulation method, which allows each variable to have multiple realities under the premise of the correctness of the overall trend [28,29]. It means that the local uncertainties are incorporated into the simulated attribute values at unsampled sites. This method overcomes some limitations of the ordinary kriging method and highlights the volatility of spatial distribution of the raw data [28,29]. Hence, stochastic simulation is a more suitable approach to estimate the S^{2-} spatial distribution in this study.

Nowadays, some researchers have integrated machine learning (ML) with the geostatistics method and have achieved good simulation and prediction results [30–33]. It is known that ML is an effective empirical approach to address regression and classification problems in science and engineering, especially in mathematics, meteorology, and computer sciences [34,35]. With the evolution and popularity of ML, a series of ML algorithms have been applied in different geoscience fields such as predicting the distribution of geochemical variables (including element concentrations) in soil, realizing the reservoir

simulation, and predicting the permeability of tight carbonates [36–38]. With the application of ML in geoscience, it gradually shows the ability in estimating physical/chemical variables that are difficult to monitor directly and forecasting long-term trends of geoscience variables [39]. Moreover, successfully combining ML and geostatistics method leads to realizing simulation and prediction of geoscience parameters, which provide a new perspective to address geoscience problems.

In this study, the conventional stochastic simulation method is not able to directly simulate and predict the continuous change of geochemical element distribution with very limited known datasets. The statistics of the probability density function are hard to be estimated, especially in the condition of sparse sampling points and insufficient sampling times of each point. To overcome the aforementioned limitations, we propose a novel method to simulate short-term changes of S^{2-} distribution in oil exploitation, which combines the stochastic simulation method and the modified probabilistic neural network (modified PNN). It possesses the advantages of stochastic simulation and the traditional ML method, so there is no need to compare it with a simple ML method. In this method, the modified PNN constructs the connections between multiple indirect datasets and the S^{2-} distribution at sampling points, which are treated as indirect data in the stochastic simulation processes. By modifying the structure, the output becomes continuous variables, and it is able to provide parameters for the next simulations. The association of the temporal structure was constructed by simulating the same layer in the continuous time range, which expands the selection range of constraints. In addition, the simulation target has been changed into the distribution of S^{2-} increment for enhancing the accuracy of the simulation. The case study illustrates the outstanding ability of this method. The flowchart of the proposed method is shown in Figure 1.

The rest of this article is organized as follows. Section 2 describes the detailed information of employed methodologies and techniques. Section 3 applies the proposed method into the field case examples. Section 4 concludes this article.

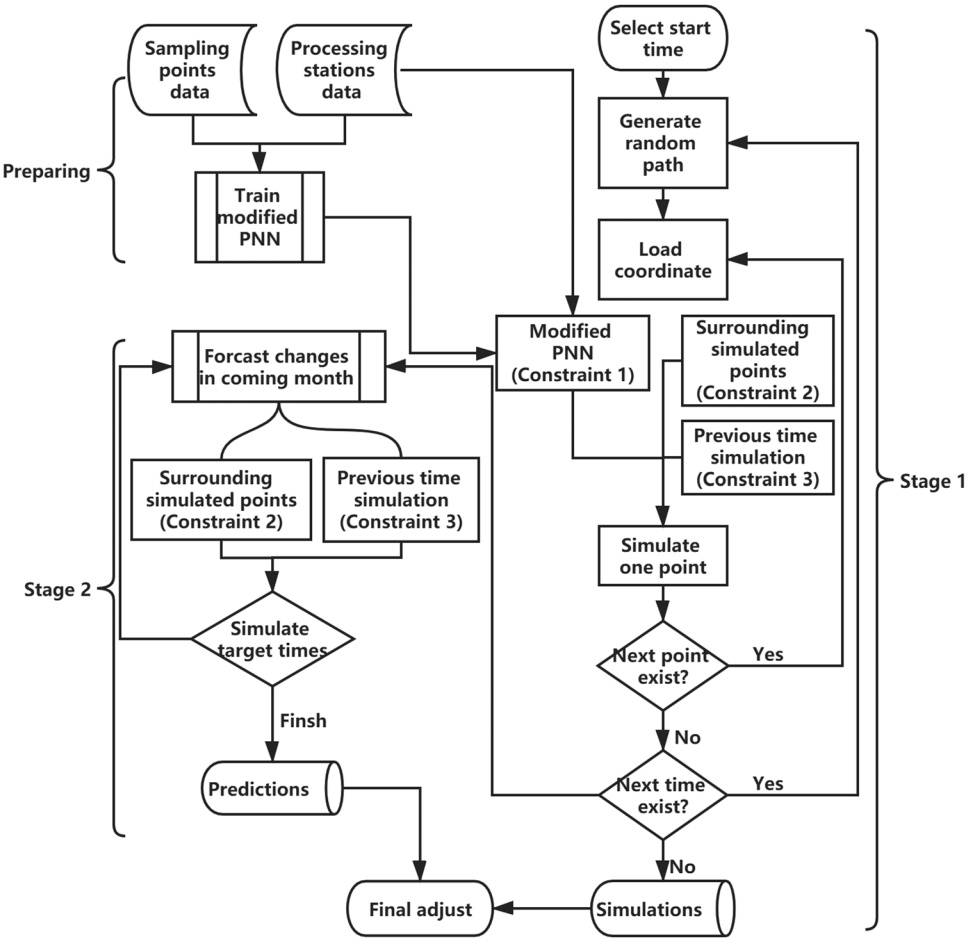

Figure 1: The flowchart of the conventional stochastic simulation and the modified PNN method.

2 Method

The detailed information of the modified PNN and adaptive probability distribution constrained stochastic simulation methods is introduced in this section. The process of the proposed method can be summarized as follows:

(1) Train the modified PNN for generating the mean value of PDF.
(2) Introduce three indirect constraints for stochastic simulation.
(3) Simulate the continuous changes of S^{2-} concentration distribution in the last 1 year by merging multiple constraints (Stage 1).
(4) Predict S^{2-} concentration distribution in the next 3 months through the previous simulation (Stage 2)
(5) Adjust the scope of results and validation.

2.1 Modified probabilistic neural network (modified PNN)

The PNN model is one of the supervised learning networks, and it can compute the nonlinear decision boundaries, which approach the Bayes optimal [40]. The training process of PNN is normally a forward propagation due to which the summation layers are highly related to the training datasets, and the output is a probabilistic expression, which will benefit the analysis of stochastic simulation. The outputs of PNN are the highest value of the competitive layer based on the distribution of probabilistic density function (PDF). As the principle of PNN, a simple PDF can be expressed as follows [41]:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} e^{\left(-\frac{|x_i-\bar{x}|}{2\sigma^2}\right)}, \quad (1)$$

where $f(x)$ is the PDF of the training datasets x, n is the number of x, and σ is the standard deviation of x. The hidden layer is able to evaluate each input array with a set of probabilities when a PDF is used in a neural network, and the output layer will select the highest value from these probabilities for the final result:

$$K(x) = \text{type}(\max(f_j(x'))), \quad (2)$$

where $K(x)$ is the final result of input data x', $f_j(\cdot)$ is the PDF of j category, and type$(\cdot)$ is the transformation from a probability to a category or a value.

PNN is a method that can provide an uncertainty evaluation of classification or regression problems through multiple parallel probability density functions. It has been effectively applied in classification, signal processing, evaluating seismic liquefaction potential, recognition system for language characters, and other fields [42–45]. However, the computational complexity of PNN is usually higher than a traditional ANN because it needs lots of comparison with training datasets to ensure the accuracy of output in each cycle of simulations. Besides, this neural network structure does not exactly match our purpose of chemical simulation in a complex layer. Therefore, we proposed the modified probabilistic neural network (modified PNN) to solve this problem for serving the simulation process. The structure of traditional PNN is modified to adapt the estimate tasks, and few layers of auto-encoder (AE) are combined to stabilize the training process. The structure of this modified PNN is shown in Figure 2. The pattern layer is replaced by the full connected layer to eliminate the tedious calculation with all training datasets, and the objective of this layer is to generate 10 possible values of the target point. Next, the summation layer will not only be a summation of former layers but will also be used to select the best value for the output layer. The decoder layer, which is designed to enhance the stability of the fully connected layer, is the same as the decoder of AE, but it do not required to be symmetry with the fully connected layer. The fully connected layer includes two layers, the summation layer includes three layers, and the decoder layer includes only one layer. The decoder layer is set to stabilize the training process, and the training will end within 300 iterations after using it, which is lower than 1,000 iterations without this structure. Besides, there is no need to enlarge the decoder because the weight of this loss function is lower than 0.3.

The main structure of this method is the same as the traditional PNN, but the objective function is changed

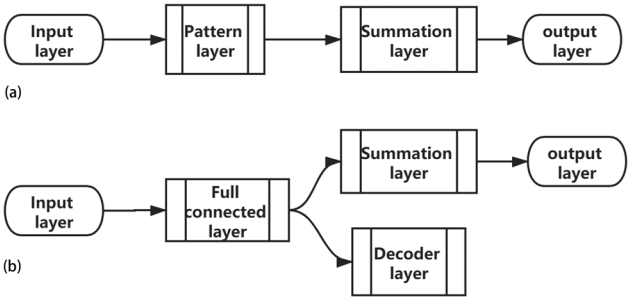

Figure 2: Structure of traditional PNN (a) and the modified PNN (b).

and divided into two parts. The first part is the loss function with the label value, which provides the prediction of a mean value for the next simulation, and it can be expressed as follows:

$$\text{Loss1}(x, y) = \min |P(x) - y|, \tag{3}$$

where x is the input data, y is the label value of x, and $P(\cdot)$ is the full connected layer. Both the training process and the prediction process select the minimum value from the summation layer. Besides, the second part of the loss function in the decoder, which is aim to stabilize the construction of the fully connected layer, can be defined as follows:

$$\text{Loss2}(x) = \sum_{i=1}^{m} |Q(P(x)) - x|, \tag{4}$$

where $Q(\cdot)$ is the decoder layer of the modified PNN. Therefore, the loss function of the modified PNN will be changed into:

$$\text{Loss3}(x) = \alpha_1 \text{Loss1} + \alpha_2 \text{Loss2},$$

where α is the weight subscripts that denote the two parts of the loss function. The loss function is optimized by iterative training, which gradually improves the effectiveness of the entire neural network, and then it can provide an average value that meets the stochastic simulation accuracy requirements. In this research, the training time of this structure is about 2 h, and it is only a small part of the total computation cost, and hence, it is not necessary to pay too much attention to reducing it.

2.2 Adaptive probability distribution constrained stochastic simulation

Bayesian sequential simulation ref. [46] is the most common method used to simulate geological and geochemical distributions. After generating a stochastic path for searching points, the posterior probability is needed to estimate a value for the current point. Gaussian probability density function is usually used for general simulation tasks:

$$p(a) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(a-\mu)^2}{2\sigma^2}\right), \tag{5}$$

where $p(a)$ is the PDF of a, μ is the mean value of a, and σ is the standard deviation of a. The mean and variance can be calculated by the Kriging method. The estimated value of the target point is obtained by random sampling in this PDF. Finally, the numerical distribution of the whole study area can be realized by simulating the entire area point by point. However, this method is difficult to simulate the distribution of the chemical component during the oilfield development. Since only a few wells have regular test data, the Kriging method hardly obtains the accurate mean and variance directly. Therefore, it is necessary to make full use of indirect data to constrain the simulation process.

In general, the changes of water components in the underground environment, such as common elements, nutrients, and bacteria content, have certain continuity in the time domain. However, changes in these components do not show a strong correlation at each time point due to the influence of water injection and other factors. Hence, the monthly variation of S^{2-} concentration is our simulation target. This study was divided into two stages: the first stage focuses on the simulation of current S^{2-} distribution in the target layer and the second stage predicts the short-term S^{2-} distribution.

In the first stage, we selected six-component datasets as indirect datasets, which are the long-term monitoring data in the three-phase separator outlet and the water injection well head. These components include sulfide content, oil content, total iron content, and three types of bacteria (sulfate-reducing bacteria (SRB), iron bacteria (FB or FEB), and putrefying bacteria (TGB)). The proposed modified PNN constructs the connection between indirect datasets of these six components and the S^{2-} distribution at sampling points. At the same time, the simulated point near the target point is used as a cooperative constraint, and the value change of this point in the previous month is used as a restriction condition. In this study, we also applied the Gaussian probability density function, which is commonly used in reservoir simulation. The mean of the function can be calculated by the following formula:

$$\begin{cases} \mu_1(z_0, r_1, u_l) = \mu_{11}(z_0, r_1, u_l) + \beta_3 H(z_0) \\ \mu_{11}(z_0, r_1, u_l) = \beta_1 \sum_{i_1=1}^{n_1} \delta_{i_1}(z_0) G(u_l) + \dfrac{\beta_2}{n_2} \sum_{i_2=1}^{n_2} a_l(z_0, r_1) \\ \beta_1 + \beta_2 + \beta_3 = 1, \end{cases} \tag{6}$$

where z_0 is the current simulating location, r_1 is the searching radius for simulated points, u_l is six kinds of relevant component changing values in the processing station at the current simulating time l, β is the weight for three kinds of restraints, n_1 is the number of processing station, δ is the inverse distance weight, $G(\cdot)$ denotes

the modified PNN, a_l is the simulated value at time l, and $H(\cdot)$ is the constraint of a former time, which is shown as follows:

$$H(z_0) = \begin{cases} 0, & \text{if } \left|\sum_{i_1=1}^{n_1} \delta_{i_1}(z_0)G(u_l)\right| \\ & \leq a_{l-1}(z_0) \\ \mu_{11}\dfrac{|\mu_{11}| - |a_{l-1}(z_0)|}{|\mu_{11}| + |a_{l-1}(z_0)|}, & \text{other} \end{cases}, \quad (7)$$

where μ_{11} is a temporary mean value. If $H(z_0) = 0, \mu_{11} = \mu_1$. This constraint is conducive to make full use of the neural network to construct the connection between the target reservoir S^{2-} concentration and indirect datasets. As a conventional random simulation, the influence of the simulated points on the unsimulated points is also considered to ensure each simulation result is smooth enough. In addition, the influence of the previous time point simulation results at the same location is also considered in the third constraint. Next, the standard deviation is calculated from the former layer to guarantee the stability of the simulation.

In the second stage, we attempt to simulate the dynamic changes of S^{2-} distribution in the next 3 months. The variance is still calculated based on the near-point value of the former layer. However, due to the lack of processing station data at these times, it is failed to calculate the mean value of the probability density function using the neural network as the main constraint. Thus, to simulate the short-term changes of S^{2-} distribution, the simulation mainly utilizes the near point value of the previous three time points and the value of the points that have been simulated in the simulating layer. At this time, the mean value calculation formula can be expressed as follows:

$$\mu_2(z_0, r_2, r_3) = \frac{\lambda_1}{3n_3}\sum_{i=1}^{n_3}\sum_{j=1}^{3} a_{l-j}(z_0, r_2) + \frac{\lambda_2}{n_4}\sum_{i=1}^{n_4} a_l(z_0, r_3), \quad (8)$$

where r_2 and r_3 are the searching radius for previous time and this time, respectively, λ is the weight of different constraints, and n is the number of known points in the searching scope. Many previously simulated datasets are introduced as constraints to reduce the influence caused by the processing station datasets deficient, which is helpful to enhance the stability of PDF and relatively accurately reflect the changes of overall data. Although the small-scale changes or fluctuations of the simulation target cannot be accurately realized by this prediction, the prediction results still have the function of evaluation and guidance for safe production in the oilfield.

In short, the adaptive PDF constrained random simulation is realized by the aforementioned processes with a series of indirect datasets. It obtains the S^{2-} distribution at the current time and the changes of S^{2-} distribution in the near future by conducting random simulations with equal probability. We will apply it to an oilfield development area case in the next section.

3 Result and discussion

The study area is located in the east of China. It has formed multiple sets of source bed, reservoir rocks, and caprocks due to the polycyclic tectonic movements and the cyclical change of climate. The types of oil and gas-bearing reservoirs that have been discovered are mainly sandstone reservoirs. In this study area, the S^{2-} concentration in the produced water was very low at the early stage of water flooding development. In recent years, the increase of S^{2-} concentration and serious pipeline corrosion gradually occur in some areas. Due to the high concentration of SO_4^{2-} in the formation water in this area, the possibility of generating high S^{2-}concentration could increase with the quantity of SRB. The sudden increase of S^{2-} concentration is a potential risk for operators and sewage treatment equipment, especially where the current detection of S^{2-} concentration is not overstandard. Hence, prediction of the short-term change trend of S^{2-} distribution can provide significant guidance for oilfield development and safety. There are 67 sample points in this study area, which are in the vicinity of four different process stations, as shown in Figure 3.

The variation range of the original data is very large as shown in Figure 4a and b, which exist as a large order

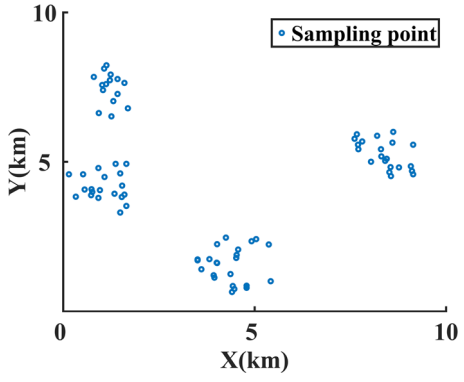

Figure 3: The distribution of sampling points in the region. *X* and *Y* represent the geographic distance.

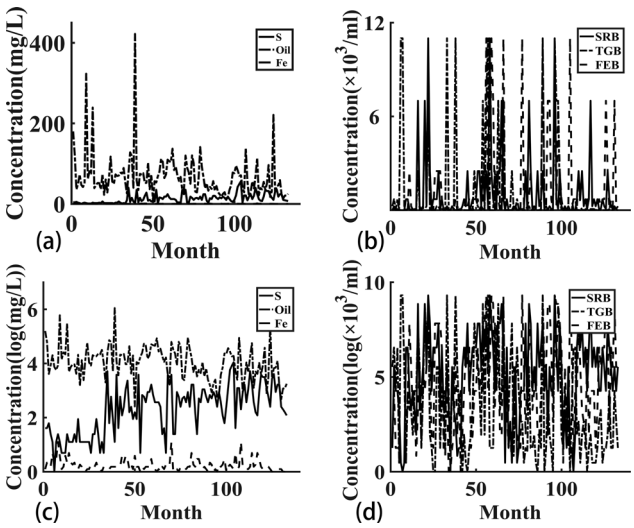

Figure 4: Traditional (a and b) and transformed (c and d) distribution of the processing station data. (a) and (c), The S, oil, and Fe represent sulfide content, oil content, and total iron content in the produced water, respectively. (b) The number of bacteria per milliliter in the produced water.

of magnitude difference between these data. In particular, the concentrations of Fe and Oil show relatively significant changes, while the variation range of S concentration is relatively insignificant (Figure 4a). It easily leads to a substantial imbalance in the neural network training by directly adopting these data. In this study, the log transformation is used to reduce the large variance of the original data. The transformed results are presented in Figure 4c and d, which increase the distinguishability of the transformed three chemical composition data. While the variation ranges of three types of bacterial concentration are very large, they are basically in a similar variation range. Thus, the normalized data meet the parameters input requirements of the neural network. The log transformation equation can be expressed as follows:

$$S_1(S_0, k) = \log(S_0 + k), \tag{9}$$

where S_1 is the transformed processing station data, S_0 is the original processing station data, and k is the hyperparameter about the minimum value of the overall data distribution to ensure that the transformed S_1 is greater than 0.

The modified PNN needs to be trained before we use it to estimate the distributions of S^{2-}. The input datasets include concentrations of sulfide, oil, total iron, SRB, FEB, and TGB, and the outputs are five probable concentration of S^{2-}. Eighty percent of these datasets collected in 2008–2011 are introduced for training the modified PNN, and the final residual is about 0.23, which is acceptable for stochastic simulations.

In the training process, the update weight is 0.01, batch size of the normalization is 20, and each iteration includes 100 batch. The activation function using in the training process is RELU. After 300 iterations, the trained model can be used for assisting the calculation of mean values.

There is an overall increasing trend of S^{2-} distribution from 2010 to 2011 (as shown in Figure 5), which is in line with the oilfield development status. Figure 5 shows that the rough distribution of S^{2-} is concentrated in three small areas in two distribution maps. However, the entire distribution map is obtained by simulation. Thus, the value of distribution points away from the sampling point is set as 0, which means that these two initial maps cannot fully reflect the S^{2-} distribution in the target layer. This is one of the defects of conventional simulation methods.

According to the sequential simulation process, the modified PNN will be introduced into equation (6), and six kinds of indirect datasets are considered to constrain the simulation. The weights of constraints are 0.5, 0.3, and 0.2 if the former layer exists, or the weights are 0.7, 0.3, and 0. This weight means that the mean value of the simulation mainly relies on the modified PNN, and the influence of the former layer is the minimum. Then, the mean value will be introduced into equation (5) to build the probability density function.

Then, we need about 20 h to run the simulation process for obtaining the S^{2-} concentration changes. In the research block, the changes of S^{2-} distribution in 12 months mainly show the slowly rising trend and the overall distribution is relatively stable (Figure 6). It is consistent with the change of S^{2-} concentration in the processing station. The variation of concentration appeared in 2011 is in coincidence with the oilfield development trend. From the first month to the sixth month, the increment of S^{2-} concentration shows some reduction, and even some areas have shown the negative growth. These might be attributed to the addition of

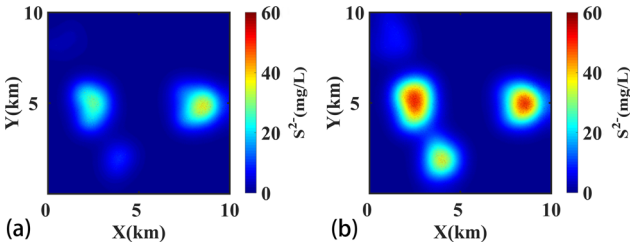

Figure 5: Distribution maps of S^{2-} concentration in 2010 (a) and 2011 (b).

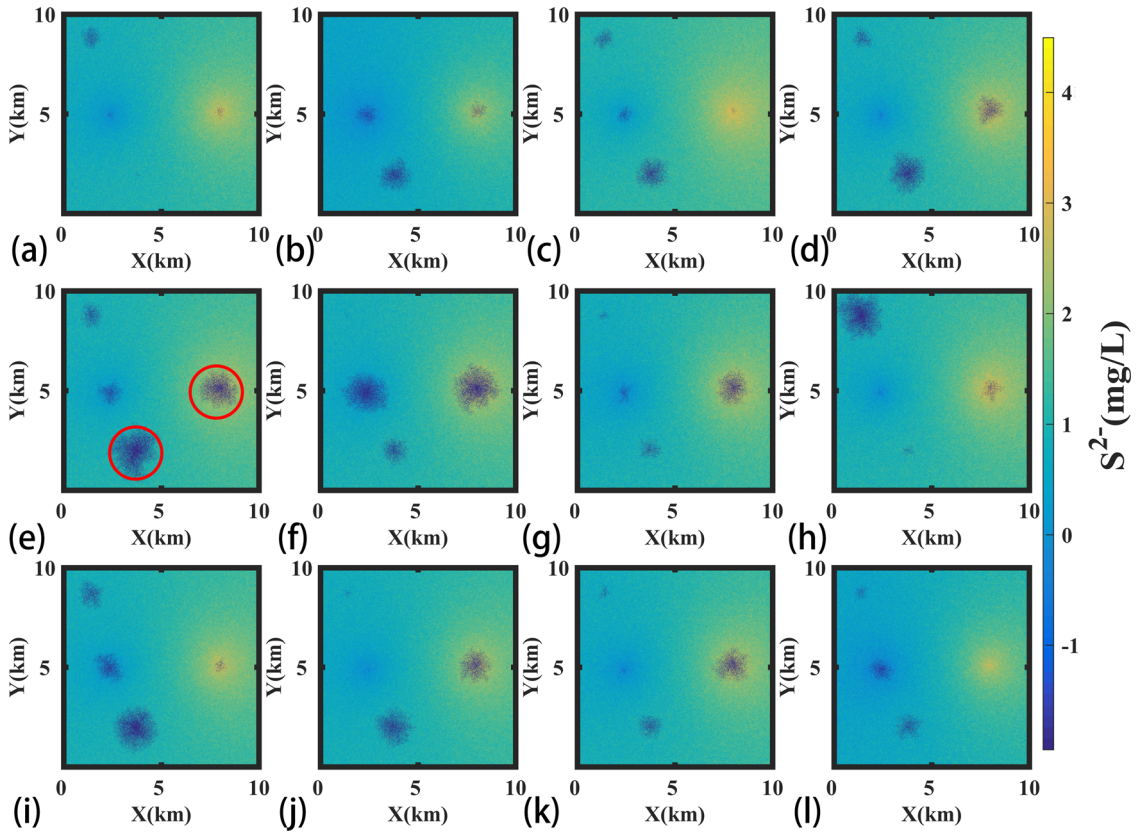

Figure 6: Distribution maps of S^{2-} concentration changes in 12 months. (a)–(l) January–December in 2012.

bactericide in the injection water of the test well, which effectively decreases the number of SRB and controls the increasing trend of S^{2-} concentration. However, with the process of water injection, the number of SRB is increasing again when the concentration of bactericide is exhausted. Hence, there is some increase in the increment of S^{2-} concentration in the later months. In Figure 6, a few outliers appeared in the red circle, which will not influence the following simulations. Since the absolute value of the data is not large, it has little effect on the final result.

Based on the simulation results presented in Figure 7, which ignore the outliers, the overall tendency of simulation results changes in 12 months, which is consistent with the trend shown in Figure 6. This also confirms that these abnormal points have a very limited impact on the simulation results. First, we test the overall precision by comparing the distribution of interpolation concentrations. The average accuracy of these results can be considered as 80%, which includes the points with less than 15% deviation. Besides, the total residual of one map is lower than 5%. In the study area, the distribution of S^{2-} concentration basically shows a different increasing trend. It can be found that S^{2-} distribution shows a significant increase in the red box (Figure 7). In contrast, the S^{2-} distribution has invaded the yellow box, and the concentration changes is not obvious. This may be related to the treatment scheme, reservoir characteristics, and water quality environment. Generally, the injection production form that was adopted in the process of water flooding development is that one injection well corresponds to multiple production wells. The addition of bactericide with injected water might have different effects on the S^{2-} distribution of the produced water due to the differences in reservoir connectivity, diffusion coefficient, water content, and so on. For example, the changing time of S^{2-} is earlier than before, and the changes of the numerical values and ranges are obvious in the reservoir with good connectivity. Besides, owing to the different types of SRB and water quality environment in different blocks, the growth rate of SRB and the drug resistance is different. These are possible reasons for the rapid increase of S^{2-} distribution in the eastern part of the research area, while the distribution of S^{2-} in the western part of the research area only gradually diffused with no obvious increase.

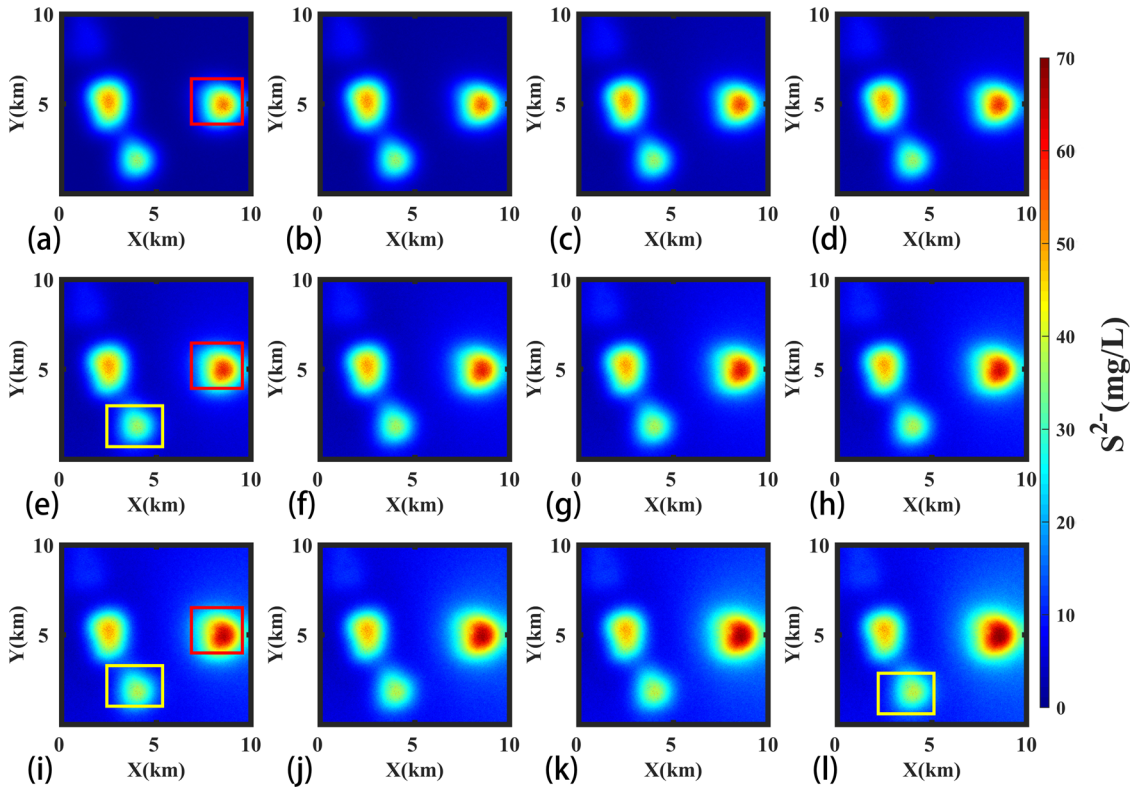

Figure 7: Distribution maps of S^{2-} concentration in 12 months. (a)–(l) January–December in 2012.

Moreover, the values of the processing station are also utilized for verifying the effectiveness of this method. We selected five sampling points near the processing station as verification points, one of which has a large deviation, and in the case of other four points, the deviation between the test value and the simulation result is within 13%. Therefore, the simulation results can effectively reflect the S^{2-} distribution in the target layer. It can be considered that the simulation accuracy of target points within 2 km of the original sampling points is higher than 85%, and they roughly reflect the variation trend of the target point located away from the sampling points. Thence, these simulation results can provide certain guidance for oilfield development and safety.

We can confirm that the modified PNN is effective again by these ideal simulations due to the weight of the major controlling factor. Although the first layer cannot be controlled by the former layer, its simulation still performs smoothly. Therefore, the influences of the same layer points and the former layer are very limited for the mean value, but the contribution of these constraints for stabilizing the probability distribution is still important. In terms of the current simulation results, it is necessary to take better inhibitory measures to prevent the continuous increase of S^{2-} concentration in the eastern part of the study area. The diffusion of S^{2-} distribution in the western part needs to be paid attention to, and some precaution measures should be taken.

Reasonable results of S^{2-} distribution are obtained in the simulation of 12 months in 2012. Therefore, we predict the distribution of S^{2-} distribution in the next 3 months using these data. In the simulation of the next 3 months, a large number of outlier blocks appear in the S^{2-} distribution maps because of the lack of constraints of the neural network and the processing station data. The failure of simulations in these locations may be due to few constraints. However, the overall change trend fits the general patterns of S^{2-} distribution change in the reservoir.

The monthly changes of S^{2-} distribution have shown the increasing trend in prediction 1 (Figure 8(a–c)), the decreasing trend in prediction 3(Figure 8(d–f)), and little changes in prediction 2, prediction 4, and prediction 5 (Figure 8(d–f), (j–l), and (m–o)). Since the prediction of S^{2-} concentration derived from the stochastic simulation is completely based on the previous statistical data, obtaining such changes also conforms the statistical law. However, the outliers need to be processed because the number of outliers is far more than the simulation results with neural network constraints. Here, the neatest neighbor value is used as the outlier value.

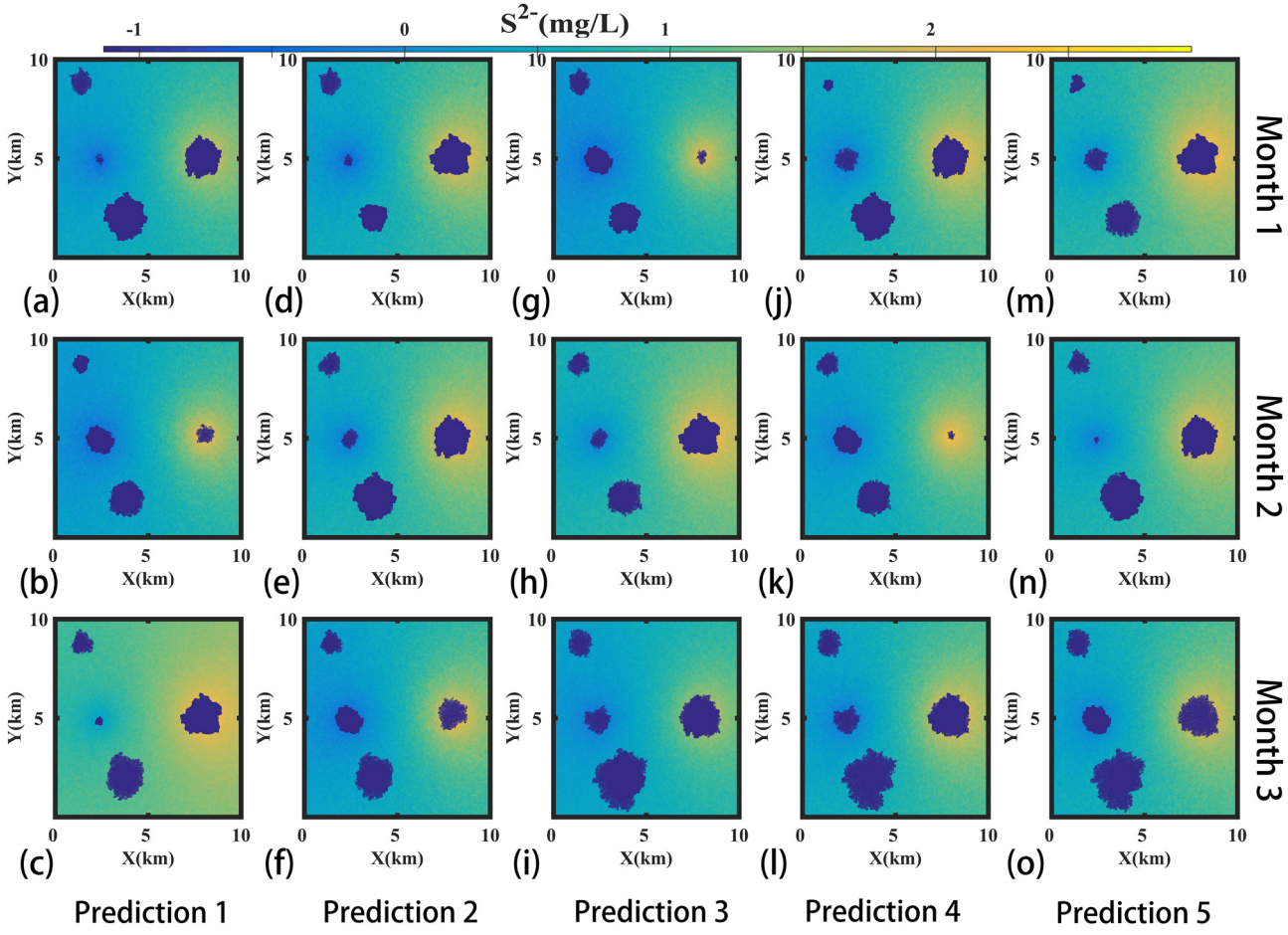

Figure 8: Predictions of S^{2-} concentration distribution on the next 3 months in 5 times stochastic simulations.

The five prediction results mainly reflect the statistical value changes because the prediction results are largely dependent on the previous simulation results. In Figure 9, the predictions of S^{2-} distribution presented by ladder diagrams due to the changes of S^{2-} distribution are insignificant.

According to the five predictions, they are not sensitive to the changes of S^{2-} distribution on a small scale. This might be because there is no obvious change in the last 12 months in the western part, and there is only a small change in the overall S^{2-} distribution in the eastern part. The changes of S^{2-} concentration looks similar to these 12 predictions because the constraints of the modified PNN are lost, and the former layers become the main factors for estimating the probability distribution. The weights in equation (8) are 0.8 and 0.2 when the known points of the same layer are lower than 20%, and the weights will be changed into 0.5 and 0.5 after the known points are higher than 20%. Hence, the prediction results mainly show small-range changes of S^{2-} distribution. However, these five predictions still have the indicative ability for the changes of S^{2-} distribution. For example, in prediction 1 (Figure 9(a–c)), the range of S^{2-} distribution shows the north–south extension and shrinkage in the second and third months, respectively. However, the prediction 3 shows a gradual shrink in the north–south direction for S^{2-} distribution range in the eastern part (Figure 9(j–l)). In five prediction results, the change is mainly reflected in the north–south direction extension of the S^{2-} distribution range in the eastern part of the study area, and the extended diffusion of the east–west direction is small. It indicates that these predictions have the ability to show the short-term changes in the diffusion and distribution range of S^{2-}, even though it is not sensitive to the absolute change of the value. Therefore, predictions can provide some guidance for safe production, which can be considered as the reference to assess the possibility of S^{2-} distribution overstandard in different areas in advance.

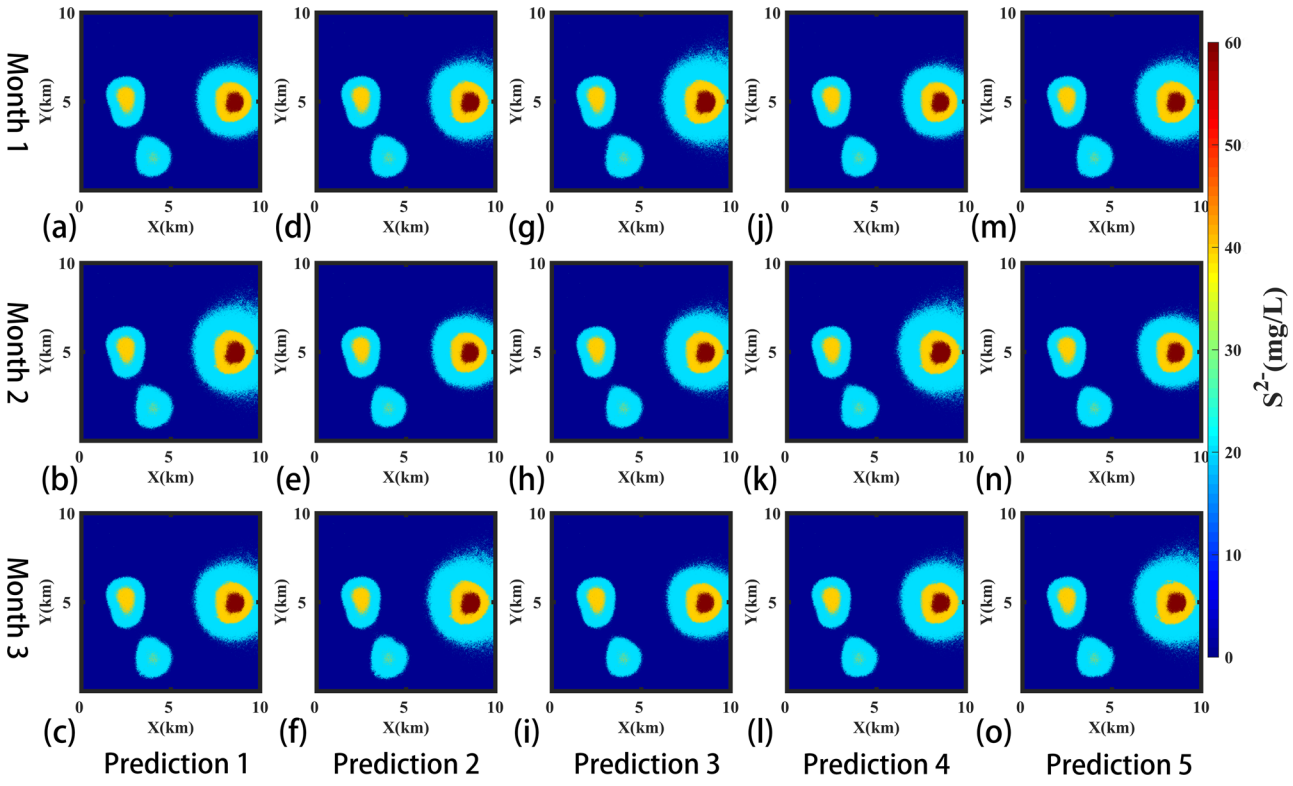

Figure 9: Prediction of S^{2-} concentration distribution maps over the next 3 months.

4 Conclusion

In this study, the machine learning-driven stochastic simulation method is presented and applied to realize the simulation of S^{2-} distribution. The simulation of the current S^{2-} distribution and the prediction of the short-term S^{2-} distribution are carried out successively. Based on the results of this study, several conclusions can be drawn as follows:

(1) The indirect information constructed the modified PNN, which is able to provide extra support for changing the PDF and enhance the stability of the simulation.
(2) The S^{2-} distributions possess good continuity, and the accuracy is around 80%, which relies on the two-stage strategy and multiple constraints, and these ideal results demonstrate the effectiveness of the proposed method.
(3) The results reveal the chemical and bacterial movements, so this proposed method has the ability to show the overall trends and the changes in value in an oil field, and even some small-scale changes or fluctuations of S^{2-} distribution is not displayed on these predictions

Results of this study suggest that the proposed machine learning-driven stochastic simulation has certain application prospects for the oilfield development and safety. The predictions show the distribution range of S^{2-}, but the value is not as accurate as of the simulations. Besides, it also has a great potential to apply in simulating and predicting the spatial distribution of different physical/chemical variables in other fields. Although the computational time of this approach is much higher than the traditional method, its unique capacity is still worth and acceptable for application.

Acknowledgments: The authors would like to thank Dr. Guo Qiang for the technical help. The authors gratefully appreciate the two anonymous revivers for offering valuable comments that led to great improvements in this article.

Funding information: This research is supported by Changzhou Sci & Tech Program Grant No. 20210340.

Conflict of interest: Authors state no conflict of interest.

References

[1] Cavallaro AN, Alberdi MI, Galliano GR. Overview of H2S souring cases in Argentina reservoirs: origin and mitigation scenarios. Buenos Aires, Argentina: Latin American and Caribbean Petroleum Engineering Conference. Society of Petroleum Engineers; 2007. doi: 10.2118/107376-MS.

[2] Song YC, Piak BC, Shin HS, La SJ. Influence of electron donor and toxic materials on the activity of sulfate reducing bacteria for the treatment of electroplating wastewater. Water Sci Technol. 1998;38(4–5):187–94.

[3] Eckford RE, Fedorak PM. Using nitrate to control microbially-produced hydrogen sulfide in oil field waters. Stud Surf Sci Catal. 2004;151:307–40.

[4] May PM, Batka D, Hefter G, Königsberger E, Rowland D. Goodbye to S^{2-} in aqueous solution. Chem Commun. 2018;54(16):1980–3.

[5] Obuka NSP, Okoli NC, Ikwu GRO, Chukwumuanya EO. Review of corrosion kinetics and thermodynamics of CO_2 and H_2S corrosion effects and associated prediction/evaluation on oil and gas pipeline system. Int J Sci Technol Res. 2012;1(4):156–62.

[6] Sun W, Nešic S. A mechanistic model of uniform hydrogen sulfide/carbon dioxide corrosion of mild steel. Corrosion. 2009;65(5):291–307.

[7] Cord-Ruwisch R, Kleinitz W, Widdel F. Sulfate-reducing bacteria and their activities in oil production. J Pet Technol. 1987;39(1):97–106.

[8] Selene CH, Chou J. Hydrogen sulfide: human health aspects. Geneva: Concise International Chemical Assessment Document; 2003. p. 53

[9] Hughes MN, Centelles MN, Moore KP. Making and working with hydrogen sulfide: the chemistry and generation of hydrogen sulfide in vitro and its measurement in vivo: a review. Free Radic Biol Med. 2009;47(10):1346–53.

[10] Vance I, Thrasher DR. Reservoir souring: mechanisms and prevention. Petroleum microbiology. Salt Lake City, USA: ASM Press; 2005. p. 123–42

[11] Mueller RF, Nielsen PH. Characterization of thermophilic consortia from two souring oil reservoirs. Appl Environ Microbio. 1996;62(9):3083–7.

[12] Hubert C, Nemati M, Jenneman G, Voordouw G. Containment of biogenic sulfide production in continuous up-flow packed-bed bioreactors with nitrate or nitrite. Biotechnol Prog. 2003;19(2):338–45.

[13] Gieg LM, Jack TR, Foght JM. Biological souring and mitigation in oil reservoirs. Appl Microbiol Biotechnol. 2011;92(2):263.

[14] Engelbrektson A, Hubbard CG, Tom LM, Boussina A, Jin YT, Wong H, et al. Inhibition of microbial sulfate reduction in a flow-through column system by (per) chlorate treatment. Front Microbiol. 2014;5:315.

[15] Ligthelm DJ, De Boer RB, Brint JF, Schulte WM. Reservoir souring: an analytical model for H_2S generation and transportation in an oil reservoir owing to bacterial activity. Offshore Europe. Aberdeen, UK: Society of Petroleum Engineers; 1991. doi: 10.2118/23141-MS.

[16] Haghshenas M, Sepehrnoori K, Bryant SL, Farhadinia M. Modeling and simulation of nitrate injection for reservoir-souring remediation. SPE J. 2012;17(3):817–27.

[17] Cheng Y, Hubbard CG, Li L, Bouskill N, Molins S, Zheng L, et al. Reactive transport model of sulfur cycling as impacted by perchlorate and nitrate treatments. Environ Sci Technol. 2016;50(13):7010–8.

[18] McGrath D, Zhang C, Carton OT. Geostatistical analyses and hazard assessment on soil lead in Silvermines area, Ireland. Environ Pollut. 2004;127(2):239–48.

[19] Pannecoucke L, Le Coz M, Freulon X, de Fouquet C. Combining geostatistics and simulations of flow and transport to characterize contamination within the unsaturated zone. Sci Total Environ. 2020;699:134216.

[20] Pozdnyakova L, Zhang R. Geostatistical analyses of soil salinity in a large field. Precis Agric. 1999;1(2):153–65.

[21] Vince T, Szabó G, Csoma Z, Sándor G, Szabó S. The spatial distribution pattern of heavy metal concentrations in urban soils – a study of anthropogenic effects in Berehove, Ukraine. Open Geosci. 2014;6(3):330–43.

[22] Islam AT, Shen S, Bodrud-Doza M, Rahman MA, Das S. Assessment of trace elements of groundwater and their spatial distribution in Rangpur district, Bangladesh. Arab J Geosci. 2017;10(4):95.

[23] Pyrcz MJ, Deutsch CV. Geostatistical reservoir modeling. New York, USA: Oxford university press; 2014.

[24] Doyen PM. Porosity from seismic data: a geostatistical approach. Geophysics. 1988;53(10):1263–75.

[25] Burrough PA. GIS and geostatistics: essential partners for spatial analysis. Environ Ecol Stat. 2001;8(4):361–77.

[26] Han F, Zhang H, Guo Q, Rui J, Ji Q. Lithological identification with probabilistic distribution by the modified compositional Kriging. Arab J Geosci. 2019;12(18):580.

[27] Yingjun S, Jinfeng W, Yanchen B. Study on progress of methods in geostatistics. Adv Earth Sci. 2004;19(2):268–74.

[28] Reis AP, Da Silva EF, Sousa AJ, Matos J, Patinha C, Abenta J, et al. Combining GIS and stochastic simulation to estimate spatial patterns of variation for lead at the Lousal mine, Portugal. Land Degrad Dev. 2005;16(2):229–42.

[29] Zhang HJ, Ma SC, Liu WK, Zhang HB, Yuan SH. Three-dimensional spatial simulation and distribution characteristics of soil organic matter in coal mining subsidence area. Materials Science Forum. Vol. 980. Switzerland: Trans Tech Publications Ltd; 2020. p. 437–48.

[30] Kanevski M, Parkin R, Pozdnukhov A, Timonin V, Maignan M, Demyanov V, et al. Environmental data mining and modeling based on machine learning algorithms and geostatistics. Environ Model Softw. 2004;19(9):845–55.

[31] Foresti L, Pozdnoukhov A, Tuia D, Kanevski M. Extreme precipitation modelling using geostatistics and machine learning algorithms. GeoENV VII – geostatistics for environmental applications. Dordrecht, Netherlands: Springer; 2010. p. 41–52.

[32] Li J, Potter A, Huang Z, Heap A. Predicting seabed sand content across the Australian margin using machine learning and geostatistical methods. Canberra, Australia: Geoscience Australia; 2012.

[33] Han F, Zhang H, Rui J, Wei K, Zhang D, Xiao W. Multiple point geostatistical simulation with adaptive filter derived from neural network for sedimentary facies classification. Mar Pet Geol. 2020;118:104406.

[34] Lary DJ, Alavi AH, Gandomi AH, Walker AL. Machine learning in geosciences and remote sensing. Geosci Front. 2016;7(1):3–10.

[35] Belchansky GI, Douglas DC, Eremeev VA, Platonov NG. Variations in the Arctic's multiyear sea ice cover: a neural network analysis of SMMR-SSM/I data, 1979–2004. Geophys Res Lett. 2005;32(9).
[36] Kirkwood C, Cave M, Beamish D, Grebby S, Ferreira A. A machine learning approach to geochemical mapping. J Geochem Explor. 2016;167:49–61.
[37] Han F, Zhang H, Chatterjee S, Guo Q, Wan S. A modified generative adversarial nets integrated with stochastic approach for realizing super-resolution reservoir simulation. IEEE Trans Geosci Remote Sens. 2019;58(2):1325–36.
[38] Al Khalifah H, Glover PWJ, Lorinczi P. Permeability prediction and diagenesis in tight carbonates using machine learning techniques. Mar Pet Geol. 2020;112:104096.
[39] Karpatne A, Ebert-Uphoff I, Ravela S, Babaie HA, Kumar V. Machine learning for the geosciences: challenges and opportunities. IEEE Trans Knowl Data Eng. 2018;31(8):1544–54.
[40] Specht DF. Probabilistic neural networks. Neural Netw. 1990;3(1):109–18.
[41] Mishra S, Bhende CN, Panigrahi BK. Detection and classification of power quality disturbances using S-transform and probabilistic neural network. IEEE Trans Power Deliv. 2007;23(1):280–7.
[42] Saritha M, Joseph KP, Mathew AT. Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. Pattern Recognit Lett. 2013;34(16):2151–6.
[43] Zaknich A. Introduction to the modified probabilistic neural network for general signal processing applications. IEEE Trans Signal Process. 1998;46(7):1980–90.
[44] Goh ATC. Probabilistic neural network for evaluating seismic liquefaction potential. Can Geotech J. 2002;39(1):219–32.
[45] Al-Omari FA, Al-Jarrah O. Handwritten Indian numerals recognition system using probabilistic neural networks. Adv Eng Inform. 2004;18(1):9–16.
[46] Golightly A, Wilkinson DJ. Bayesian sequential inference for nonlinear multivariate diffusions. Stat Comput. 2006;16(4):323–38.