ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Experimental and Statistical Investigation of Reservoir Properties with the Effect of Waterflooding Treatment

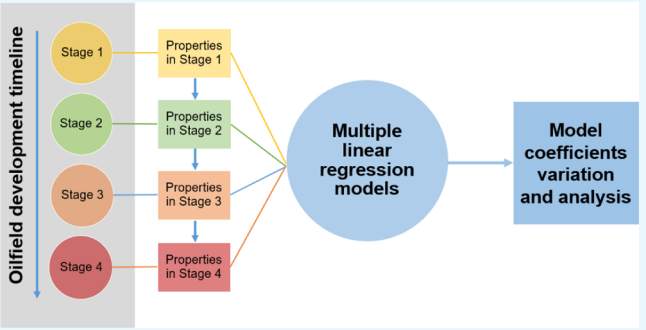
Zihao Li, Chenguang Du, Yongqiang Tang, and Xiangming Li*

Cite This: *ACS Omega* 2020, 5, 20922−20931

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: An oilfield reservoir over long-term operation may have different petrophysical information, which has a significant impact on oilfield maintenance and finance. Successful oilfield enhanced oil recovery benefits a lot from identifying and analyzing the variations of the critical properties after long-term waterflooding treatments. Since the inspection wells drilled within different development periods contain the core samples that have the petrophysical information at that period, it is necessary to collect and test the samples from different periods to investigate the overall tendency of the petrophysical properties. The samples from four inspection wells, which were drilled in four stages since the very beginning of development, were subjected to in-laboratory core analysis methods to illustrate the variation of some critical parameters in the reservoir. The permeability and porosity variation are revealed clearly by the experimental results. The migration and dissolution of clay minerals play a crucial role in the variation of petrophysical information and pore structure. To quantify the variations above, we applied the multiple linear regression model into our investigation. The dependent variable and all of the predictors in the model come from the experimental results. The quantitative results show the closed correlation between different parameters in the formation. With the development stage moving forward, the weight coefficients for different predictors have multiple trends. The experimental and statistical approach provides a novel understanding of the reservoir properties with the effect of waterflooding treatment



INTRODUCTION

The consumption of hydrocarbon energy in the world has been increasing steadily during the past several decades. To meet the rising energy demand, some oilfields that have been developed for several decades will still play an important role because the remaining hydrocarbon resources in place are still promising. To adequately recover the hydrocarbon resources in these reservoirs, the reservoir petrophysical properties after several decades of development should be assessed and characterized accurately before tertiary recovery or other enhanced oil recovery (EOR) processes. In many conventional oilfields, waterflooding, which is considered as the secondary recovery process, is an important method to enhance oil recovery. After waterflooding treatments for several decades, the petrophysical properties of the reservoir can vary in many aspects. It is crucial to identify and evaluate the variations of the petrophysical properties to optimize follow-on production strategies.

Waterflooding is the use of water injection to increase production from oil reservoirs. Water displaces oil from the pore space, and the efficacy of the displacement depends on many factors.[1] The key petrophysical properties include porosity, permeability, water and oil saturations, pore structure, etc. Decades of waterflooding treatments inevitably change the petrophysical properties of a reservoir. It is crucial to advance the understanding of the variations of the petrophysical properties resulting from long-term waterflooding treatments to optimize subsequent production strategies and to maximize the hydrocarbon recovery factor.

The permeability prediction with statistical models has already been investigated by many previous studies from simple algorithm studies[2−4] to advanced models such as Artificial Neural Networks (ANNs)[5−8] and hybrid models.[9−11] However, due to many limitations such as the issue of scale, the unknown initial weighted values, and overcomplexity, these models have not been widely applied in oilfields. As an important method of data mining and data analytics, linear regression is a linear approach to modeling the relationship between a dependent variable and one or more independent variables. The case of one independent variable is called simple linear regression. For more than one independent variable, the

Received: May 20, 2020
Accepted: July 23, 2020
Published: August 11, 2020

https://dx.doi.org/10.1021/acsomega.0c02374
ACS Omega 2020, 5, 20922−20931

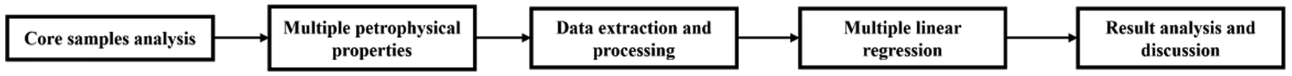


Figure 1. Experiment/data-mining integrated workflow for the assessment of the variations of petrophysical properties of the core samples.

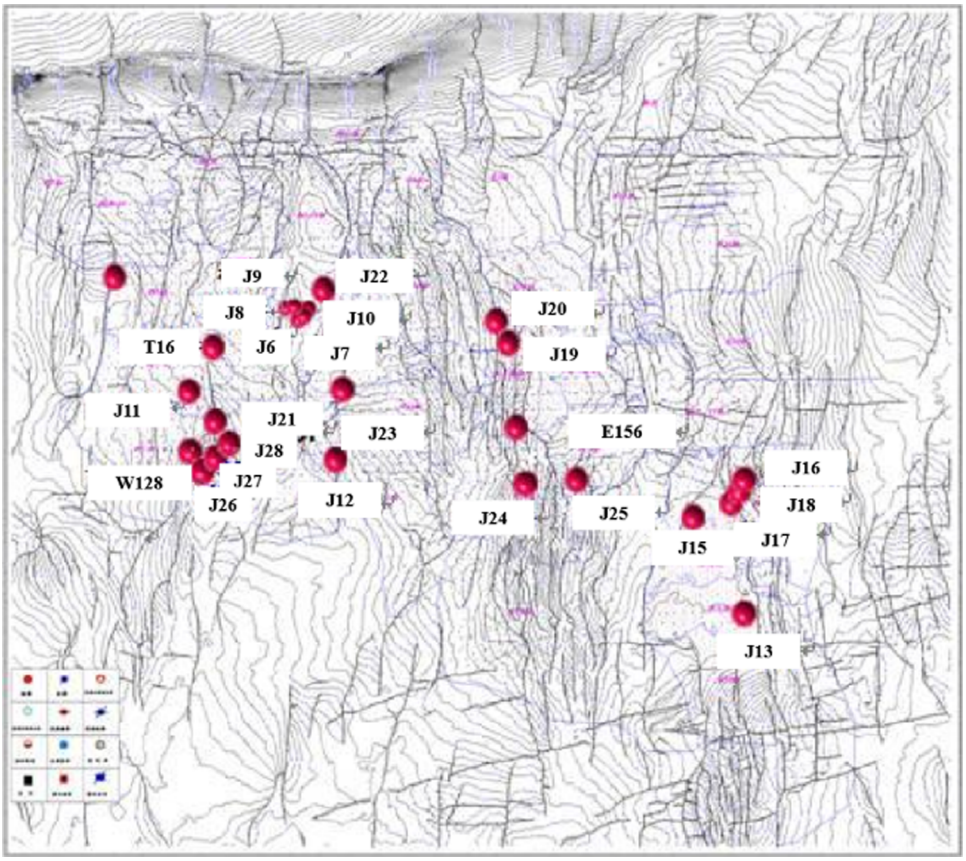


Figure 2. Inspection well map in the targeted reservoir. The whole area is about 46.6 km^2. The inspection wells from four stages were drilled respectively in 1980 (Stage 1), 2005 (Stage 2), 2011 (Stage 3), and 2014 (Stage 4). This map is modified from Li.[16]

process is called multiple linear regression.[12] The assumptions for conducting the MLR model are as follows:[13] (1) The relationship between the dependent variable and independent variables (i.e., the predictors) is linear; (2) The error term is assumed to have a normal distribution with zero mean and constant variance; (3) Homoscedasticity: the variance of the error term is constant regardless of the predictors; and (4) Non-multicollinearity: the regression model should not have two or more predictors moderately or highly correlated. Dahraj et al.[14] developed a linear mathematical model using statistical methods to predict permeability from porosity. However, a single predictor has considerable limitations in terms of prediction. The MLR model can avoid such limitations using multiple predictors.

This paper presents a comprehensive study of the changes in petrophysical properties during four different development stages under long-term waterflooding treatments. The investigated petrophysical properties include porosity, permeability, velocity sensitivity, and salinity sensitivity. Moreover, this study quantified the variations of the petrophysical data by a combination of mercury intrusion data and the MLR model, which leads to a methodology of multimeasurement investigation of the petrophysical properties of conventional reservoir development. Reservoir simulations can also benefit from this study with greatly detailed and improved reservoir petrophysical properties.

OVERVIEW AND WORKFLOW

This section aims to provide an overview of the experimental and data mining workflow. Details of the MLR model will be given in a later section. In this work, an experiment/data-mining integrated approach was developed to identify and quantify the variations of reservoir petrophysical properties of the core samples that were subjected to long-term waterflooding treatment in the field operation. First, a series of permeability and porosity measurements were conducted on the core samples that were taken from wells drilled at different times during the waterflooding operation in the same oil reservoir. Second, the mineral components of some samples from stages were investigated to confirm the variation. Third, the scanning electron microscopy (SEM) images of the samples by stages were taken. Fourth, the velocity sensitivity and salinity sensitivity of the core samples were tested to reveal the influence of waterflooding on the core samples. The details of the pore structure will then be investigated by mercury intrusion porosimetry (MIP). Finally, the MLR model was developed based on the mercury intrusion data from the four development stages. The variations in the petrophysical properties can be quantified using the MLR model. Figures 1

https://dx.doi.org/10.1021/acsomega.0c02374
ACS Omega 2020, 5, 20922−20931

and 2 illustrates a schematic workflow for the experimental and data mining workflow in this study.

More than five hundred core samples were collected from the oilfield and then classified by stages and sand units. In the laboratory, the porosity and permeability measurements were conducted by an automatic helium porosimeter and a nitrogen measuring permeameter. The testing temperatures were kept at about 90 °C. We also conducted sensitivity experiments on the core samples. The sensitivity tests include the velocity sensitivity test and the salinity sensitivity test. The purpose of the sensitivity test was to measure the triggering levels of flow velocity and salinity that cause abrupt changes in rock permeability. The velocity sensitivity is usually associated with grain migration, pore throat plugging, and permeability decrease. The salinity sensitivity refers to the permeability variation due to the change in water salinity. We conducted flooding experiments on samples extracted from the four different stages during the reservoir development history to measure the velocity and salinity sensitivity variations.

The result of mercury intrusion porosimetry (MIP) provides more details about the variation of the pore structure. The MIP is an effective instrument designed for the evaluation of porosity, pore volume, and many other parameters. The MIP utilizes a pressurized chamber to inject the mercury into the pore in the core samples. With the increase of applied pressure, mercury fills the larger pore first and then the smaller pore. During the process, the mercury injection capillary is tested. Therefore, the pore parameters can be calculated using Washburn's equation.[15] The pore parameter that MIP can evaluate includes the sorting coefficient (S_p), relative sorting coefficient (D_r), mean pore radius (r), structure coefficient (φ_p), feature structure index ($1/D_r\varphi_p$), homogeneous coefficient (α), displacement pressure (P_d), etc. The sorting coefficient (S_p) describes the distribution of grain size of sediments. The higher the sorting coefficient, the more uniform the grain size distribution is. The relative sorting coefficient (D_r) is the ratio of the sorting coefficient to the mean grain radius. The mean radius (r) is the average of pore radius values. The structure coefficient (φ_p) describes the curvature and connectivity of the pore channels. The feature structure index ($1/D_r\varphi_p$) is the reciprocal of the product of the relative sorting coefficient and the structure coefficient. The homogeneous coefficient (α) accounts for the deviation between the pore radius and the maximum pore radius. A higher value of α implies a more uniform distribution of pore size. The displacement pressure (P_d) is the pressure at which the nonwettability fluid begins to enter into the maximum pore of the core sample. All of the properties mentioned above can describe the pore structure from different perspectives. However, some of them, such as the feature structure index and relative coefficient, have functional relations with porosity and permeability. Therefore, we select five variables, which are independent of permeability, as the predictors in the MLR model. The five predictors are porosity, sorting coefficient, mean pore radius, homogeneous coefficient, and displacement pressure. After the predictors were determined, the MLR model can be developed. To have higher accuracy in this study, the high-low permeability percentage of the samples in MIP is the same as the whole samples' percentage. The MLR model was built through the following procedures: first, the assumption checks before the analyses; second, the transformation of dependent variables; third, processing of potential outliers; fourth, the goodness-of-fit test for the MLR model; and fifth, the final mathematical presentation of the MLR model. The details about the MLR model will be described in the following section.

GEOLOGIC SETTING

The studied reservoir is located in the central depression of the Songliao Basin. The top structure of the primary target strata was a monoclinal structure from NE to SW, and its east and west sides were blocked by the normal fault of nearly NS-trending. The distribution of oil, gas, and water was impacted by the structures. The reservoir-buried depth was about 500 m, the oil-bearing area was 46.6 km^2, and the geological oil reserves were 1.57×10^7 barrels. The depositional type was delta plain-subfacies deposition. The major reservoir lithology included mudstone, silty mudstone, and siltstone. The major reservoirs were six single layers, and hypomajor reservoirs were the other five single layers. The average thicknesses of the sandstone and oil layers were 57.1 and 24.5 m, respectively. The injection water is surface water. The development stage is divided by the development period. The inspection wells from four stages were drilled respectively in 1980 (Stage 1), 2005 (Stage 2), 2011 (Stage 3), and 2014 (Stage 4). The inspection wells' map is shown below (Li, 2015)

Stratigraphic Classification. The studied reservoir can be divided into four mid-term sedimentary cycles, and these cycles were composed of flooding surface steadily rising sequence frameworks. Every mid-term cycle can be divided into three short-term cycles. Based on the sedimentary cycles, the reservoirs can be divided into four sand units (Units I, II, III, and IV) or 13 single layers. The flooding surface of Unit I constantly rose and formed the predominantly developed cycle of the top mudstone section. Since the flooding surface changed frequently, Units II, III, and IV were composed of one entire short-term cycle and flooding surface rising cycle, forming the sand−mudrock frameworks. Unit I formed a retrograding delta sequence because of lacustrine transgressive during the depositional stage. At the initial depositing process of Unit II, the sequence architecture changed from continental facies to lacustrine facies, and then the stabilized lake-flooding mudstones that were developed at the mid-later stage of the depositing process composed of unified continental-facies and lake-flooding sedimentary transformation surfaces. The depositing type of Units III and IV was mainly rifted sedimentary, leading to different vertical sequence architectures in different rift units.

Depositional Characteristics. The depositional stage of the studied reservoir was the transition period of the Songliao Basin, which was from rift sedimentation to depression sedimentation. During the completely depositional process, which was illustrated as lacustrine transgression, several small-scale oscillatory subsidences used to happen. Combined with the mudstone color, the sedimentary structure, particle size distribution, biological characteristics, and regional deposition background, the delta sedimentation facies of the studied reservoir can be divided into two subfacies and 13 microfacies. The facies of this region is delta facies because the depositional process of this region was delta deposit. The two subfacies were delta plain and delta front. The delta plain can be further subdivided into a distributary channel, a small crevasse channel, an abandoned channel, a crevasse splay, a natural levee, channel margin sediment, and an interdistributary bay. The delta front can be subdivided into a subaqueous distributary channel, a channel mouth bar, sheet sand, a

https://dx.doi.org/10.1021/acsomega.0c02374
ACS Omega 2020, 5, 20922−20931

subaqueous levee, subaqueous sheet flood sediment, and a subaqueous distributary bay. Among these microfacies, fluviation was rather strong here. Therefore, the distributary channel was dominant in the delta plain, and the subaqueous distributary channel was dominant in the delta front.

Reservoir Characteristics. The test results from the thin section of core and grain size analysis illustrated that the rock types of the target zone were mainly feldspathic litharenite and a small amount of arkosic lithic sandstone. The sandstone fragment components included rock block, quartz, and feldspar with the volume fractions of 35, 30, and 30%, respectively. The dominant cement was argillaceous and limy cement, and the dominant type of cement was porous cementation. The separation of reservoir rock was medium-good. The grain size was fine, its mid-value was 0.1−0.4 mm, and the roundness was poor. Based on all of the results mentioned above, the maturity of sandstone components and the structure of this reservoir was medium.

The sand unit distribution and sedimentary microfacies development have a close relationship. The base level of the certain reservoir went through repeated shocks so that the sedimentary microfacies of single layers developed differently. Influenced by the base-level rising, the development of the subaqueous distributary channel of Unit I got decreased in the vertical direction. With a low degree of the overlay, the overlay type between the sand units was mainly shear and intraformational interbed and occasionally thin layer interbed. Influenced by the base-level dropping, the difference in the development of single layers in Unit II was significant in the vertical direction.

MULTIPLE LINEAR REGRESSION MODELING

The details of the MLR model fitting of the mercury intrusion data are provided in this section.

Assumption Checks before the Analyses. The hypothesized MLR model in this study can be written as:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \quad (1)$$

where Y is the permeability, X_1 is the porosity, X_2 is the sorting coefficient, X_3 is the mean pore radius, X_4 is the homogenous coefficient, and X_5 is the displacement pressure. It is worth noting that the model does not have intercepts because β_{1-5} are the standardized regression coefficients for the five predictors. We use standardized coefficients in all of the models because we are interested in comparing their relative effects directly. Three assumption checks were made through the MLR modeling process and described as follows:

Gaussian Distribution of the Random Error $\varepsilon \sim (0, \sigma^2)$. It is hypothesized that the error of the regression model is normally distributed. To check this assumption, the probability−probability (P−P) plot or histogram is produced, which compared the cumulative probability of the residual against the cumulative probability of the theoretical normal distribution.[17] If the data of the residuals are normally distributed or approximated to a normal distribution, they should fall into the straight diagonal line (reference line). However, if most of the empirical data deviate from the reference line, it indicates that the assumption of the normal distribution for the residual is not valid. In this case, the transformation of the dependent variable (i.e., Y, permeability) is needed. Next, the Kolmogorov−Smirnov statistic test for normality will be conducted. The p-value or the probability value is the probability for a given statistical model that, when the null hypothesis is true, the statistical summary (such as the sample mean, the difference between two compared groups) would be greater than or equal to the actual observed results.[18] The null hypothesis is a general statement or default position that there is no relationship between two measured phenomena or no association among groups.[19] If the check result shows that the p-value is less than 0.05, it indicates that the null hypothesis of the normal distribution should be rejected.[20]

Homoscedasticity. The homoscedasticity assumption requires equal variances of the dependent variable across a range of predictors. To check this assumption, a residual versus fitted value plot should be created. In this plot, the x-axis shows the fitted or predicted values, and the y-axis is for the errors of prediction. Ideally, this assumption is valid when the data points are randomly scattered over a horizontal line across the zero.[21]

No Perfect Collinearity. This assumption requires that all of the predictors are not perfectly linearly related. In the exploration stage, the pairwise correlations between independent variables are calculated to avoid collinearity.[22] The rule of thumb is that if the correlation coefficient is larger than 0.8, then severe multiple collinearity may be present. Another way to check this issue is to obtain the variance inflation factor (VIF). The VIF measures the extent to which multicollinearity has increased the variance of an estimated coefficient.[22] It looks at the extent to which an explanatory variable can be explained by all of the other explanatory variables in the equation. The VIF is often calculated simultaneously with the estimation of the regression coefficient, and it will be checked when the final regression model is conducted. A VIF value less than 5 indicates no issue of collinearity. A VIF value above 5 and less than 10 indicates a moderate issue of collinearity. A VIF value larger than 10 implies a severe issue of collinearity.

Transformation of Dependent Variables. If the result of the assumption check implies that the distribution of the error score is not normally distributed, some technics are needed to improve it. Variable transformation can be used for the dependent variable (Y). The logarithmic (Log 10) transformation is one of the most common transformations. The formula used in this study was "NEW $X = \mathrm{Log}_{10}(X + C)$", where the constant $C = 17$ was chosen because it creates the ideal P−P plot.[23] After the logarithmic transformation of the permeability, the regression model can be written as

$$\mathrm{Log}_{10}(Y + 17) = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \quad (2)$$

After this transformation, the P−P plot usually shows that the empirical error scores heavily fall on the reference line. The Kolmogorov−Smirnov test also shows that the null hypothesis of normal distribution should not be rejected if the p-value is above 0.05.

Processing of Potential Outliers. An outlier in the MLR model is an observation that deviates noticeably from the major trend of the data. It can further influence the normality and linearity of the model. Therefore, it is necessary to remove outlier data. There are several statistical methods, which can be used for identifying outliers. The most frequently used one is Cook's distance.[24] If some points in the upper right, lower right, and lower left of the graph imply that these points are outliers since they have the largest value of Cook's distance or residual.

Goodness of Fit for the MLR Model. To develop a meaningful model, we used the "forward selection" approach

https://dx.doi.org/10.1021/acsomega.0c02374
ACS Omega 2020, 5, 20922−20931

for the goodness of fit. This method adds one variable into the model at a time. In each step, each variable, which is outside the model, will be tested for its relevance to the current model. That is, the variable that is the most significant one in all of the outside variables will be added into the model, as long as the *p*-value of this variable is below 0.5 significant level.[25] It is then necessary to measure how well the model fits the data. The statistical approach for measuring the goodness of fit is called the coefficient of determination (i.e., R^2 or R Square), which ranges from 0 to 1. More specifically, R^2 means the percentage variation in the dependent variable, which is explained by independent variables or the entire regression model.[26] Therefore, the higher the value of R^2, the better the hypothesized model fits the data.

Final Mathematical Expression of the MLR Model. The final MLR model includes all of the significant predictors of permeability. Specifically, a positive coefficient for β_i represents a positive partial relationship between the predictor (independent variable) and permeability (dependent variable), whereas a negative coefficient indicates a negative partial relationship between them.

Because the logarithmic transformation was conducted only on the permeability, the interpretation of the partial effect resulting from the changes in the predictors was different from the scenario in which the logarithmic transformation is not performed. In addition, as mentioned before, the VIF value will be given associated with the estimation of the regression coefficient. If all VIF values, which are listed in the far-right column of Table 4, are smaller than five, then it suggests that there is no perfect collinearity issue.

RESULTS AND DISCUSSION

Figure 3 illustrates the porosity distribution of the four different development stages. Specifically, the core sample

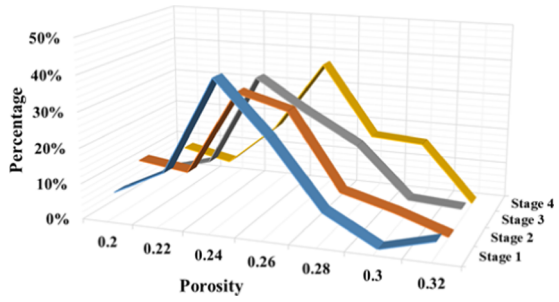


Figure 3. Porosity distribution in Stage 1 to Stage 4. The *x*-axis is the porosity, and the *y*-axis is the percentage of each porosity.

porosity ranged primarily from 22 to 26%. It is noticeable that the long-term waterflooding treatment increased the fraction of high porosity (e.g., comparing the fractions of 30% or above porosity between Stages 1 and 4), leading to an increasing average porosity in the reservoir. Figure 4 demonstrates the permeability distribution in the four development stages. The initial permeability of the reservoir (in Stage 1) was relatively low. The number fraction of cores that have permeability lower than 100 mD was about 66%, and the highest permeability was lower than 500 mD. After a period of waterflooding treatment, the number fraction of cores that have permeability lower than 100 mD decreased to 27%, and the average permeability of the entire reservoir increased noticeably (see Figure 4b). The peak permeability of Stage 2 reached 800 mD. With the water-

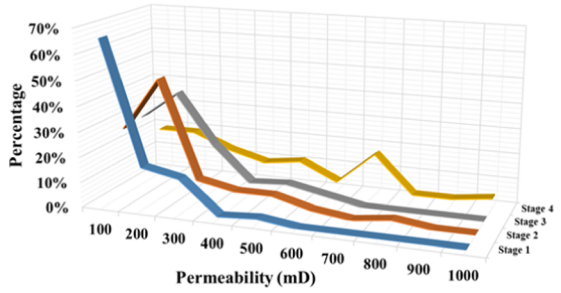


Figure 4. Permeability distribution in (a) Stage 1, (b) Stage 2, (c) Stage 3, and (d) Stage 4.

flooding continued through Stage 3, the number fraction of permeability in the 200−300 mD range increased. In Stage 4, the number fraction of low permeability decreased significantly, and the peak permeability had reached 1000 mD. This suggests that the long-term waterflooding treatment had effectively increased the average porosity and permeability of the reservoir rock.

Figure 5 below illustrates the relationship between porosity and permeability of the core samples during the four

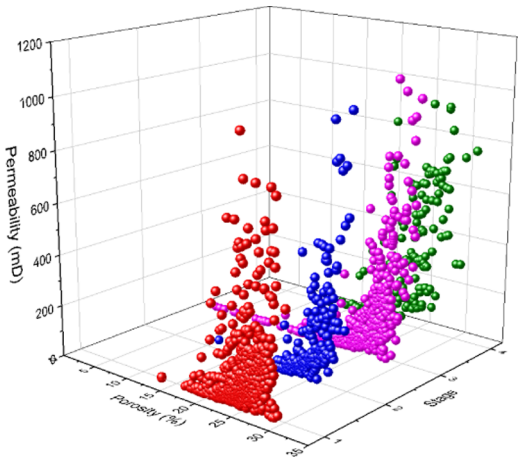


Figure 5. Relationship between porosity and permeability of core samples from Stage 1 to Stage 4.

development stages. Except for some extreme points, the overall porosity and permeability both have an increasing tendency from the initial stage to the final stage. The number of samples with low permeability is still more than the number of high permeability ones in Stage 4. The percentage of high permeability increase can indicate the preferential channel.

The overall variations of porosity and permeability reflect the overall effects resulting from the long-term waterflooding treatment. However, the effect of the waterflooding treatment is not the same for the four different sand units, even they are in the same reservoir. The waterflooding treatment can have an impact on the porosity and permeability mainly because of the migration of reservoir grains, and the water-swelling of some clay minerals. The velocity sensitivity and salinity sensitivity can reflect grain migration and clay swelling, respectively. The content of minerals with high velocity or salinity sensitivity was different in each sand unit. That can explain why the effect of the waterflooding treatment is not the same for the four different sand units even if they were in the same reservoir.

https://dx.doi.org/10.1021/acsomega.0c02374
ACS Omega 2020, 5, 20922−20931

Figure 6 illustrates the distribution of porosity in the four sand units in the reservoir. It is observed that the number

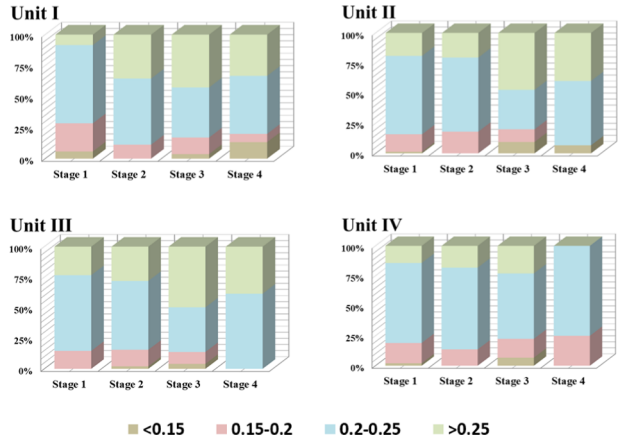


Figure 6. Porosity distribution in the four sand units from Stage 1 through 4.

fractions of high porosity (>0.25) in Unit I through III increased with the development of long-term waterflooding treatments (from Stages 1 through 4). Conversely, the number fraction of high porosity (>0.25) in Unit IV declined to a relatively low level when the waterflooding treatment developed into Stage 4. This confirms that the influence of long-term waterflooding on rock porosity was different on the four sand units even if they were in the same oilfield.

Figure 7 demonstrates the distribution of permeability in the four sand units in the reservoir. It is found that the number

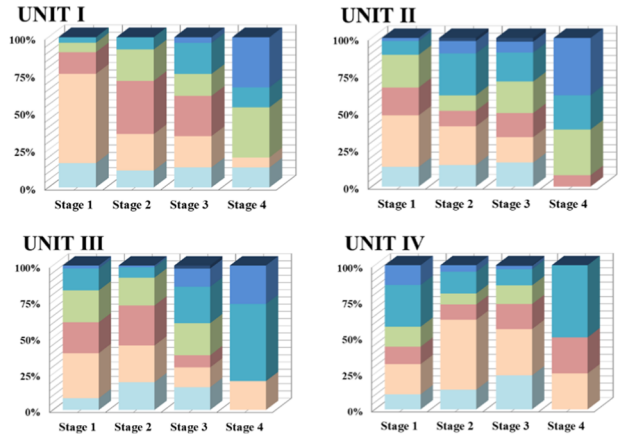


Figure 7. Permeability distribution in the four sand units from Stage 1 to Stage 4.

fractions of relatively high permeability (500−1000 mD) in Unit I through III increased with the development of long-term waterflooding treatments (from Stages 1 through 4). Conversely, the number fraction of 500−1000 mD permeability in Unit IV continuously declined from Stage 1 through Stage 4. This suggests that the influence of long-term waterflooding on rock permeability was different on the four sand units even if they were in the same oilfield.

The clay mineral contents in Stages 2 and 4 were tested to investigate the relationship between waterflooding and clay mineral. In Stage 2, the kaolinite content is 66.96%, which is quite high. The illite and illite/smectite contents are relatively low. The low-illite/smectite and high-kaolinite nature means the velocity sensitivity is strong and salinity sensitivity is weak in this reservoir. After long-term waterflooding treatment, the kaolinite content dropped from 66.96 to 49.70%. The illite content also dropped a lot. Kaolinite usually filled the intergranular pores in the reservoir with low adhesive force. With the shear force of the flowing fluid, kaolinite is extremely easy to be flushed out to create a preferential flow channel. However, the immigrated kaolinite may block the pore throats to decrease the permeability. The average permeability result shows that the permeability is increasing. This confirms that the effect of the preferential channel has a larger weight than throat blocking on the permeability variation in this particular reservoir (Table 1).

Table 1. Relative Percentage of Clay Minerals from Stage 2 to Stage 4

	relative percentage (%) of clay minerals			
stage	illite	kaolinite	chlorite	illite/smectite
2	22.55	66.96	4.49	5.99
4	15.45	49.70		34.85

Figure 8 illustrates the SEM images of the core samples extracted from different development stages. The images are grouped by four minerals from left to right, which are quartz, feldspar, kaolinite, and illite. The images illustrate that the primary pore structural changes of these core samples from Stage 1 through Stage 4 resulted from the dissolution by the waterflooding treatment. Since quartz is usually hard to be eroded, the dissolution brought about by decades of waterflooding cannot be ignored. The pore-scale erosion can be found noticeably in Stage 4. Anderson (1987) found that the wettability of rock surfaces can change and even reverse under the impact of long-term waterflooding treatments. This is because oil-wet materials on rock surfaces can be corroded and removed after long-term waterflooding, leading to more water-wet rock surfaces and consequently reduced contact angles.

The sensitivity testing results (Table 2) illustrate that the long-term waterflooding treatment had an influence on velocity sensitivity and salinity sensitivity. The velocity sensitivity was mitigated gradually with the development of long-term waterflooding, i.e., the reservoir rock became less and less sensitive to the waterflooding velocity from Stages 1 through 4. This was because in the earlier stages of waterflooding treatment, fine-sized clay grains in the reservoir rock were flushed and dislodged from the reservoir. Consequently, the average pore size, porosity, and permeability of the reservoir rock increased. When it came to Stage 4, the pore structure stabilized and the remaining rock minerals cannot be dislodged easily, leading to decreased velocity sensitivity. For salinity sensitivity testing, the result illustrates that salinity sensitivity increased with the development of long-term waterflooding treatment. This was because the reservoir rock contains various clay minerals such as smectite, illite, kaolinite, and chlorite. These clay minerals have different expansion factors. Specifically, smectite has the maximum expansion factor, illite has the middle one, and kaolinite and chlorite have the minimum one. In a later stage of the waterflooding treatment, low-expansion-factor minerals, such as kaolinite and chlorite, were dislodged by the waterflooding treatment because of their high velocity. The remaining minerals were mainly illite and

https://dx.doi.org/10.1021/acsomega.0c02374
ACS Omega 2020, 5, 20922−20931

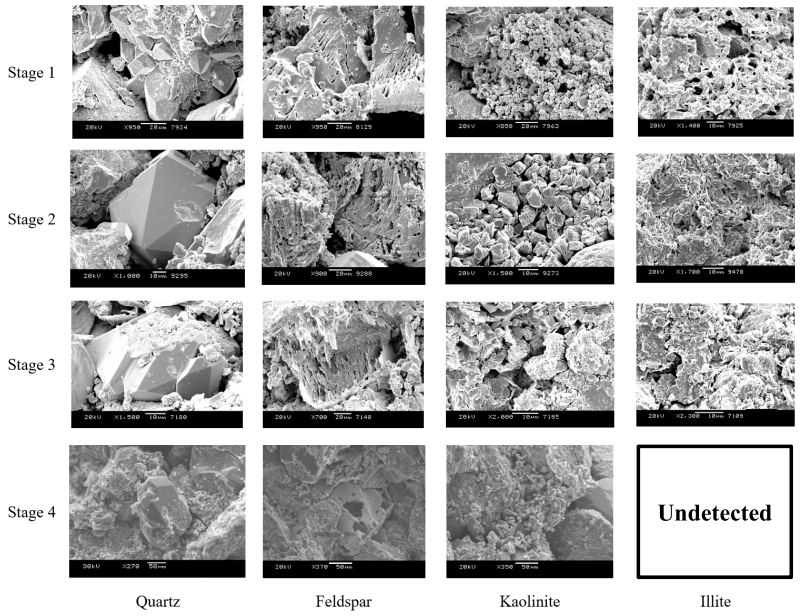


Figure 8. SEM images from Stage 1 through Stage 4. The four columns (from left to right) account for four mineral components.

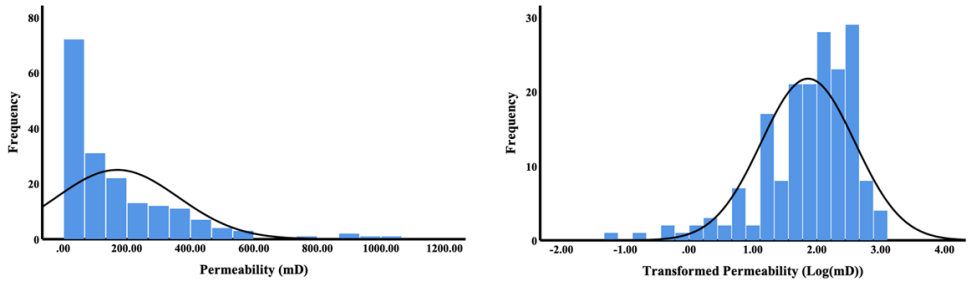
Table 2. Velocity and Salinity Sensitivity in Different Stages

stages	velocity sensitivity	salinity sensitivity
1	strong	medium weak
2	strong	medium strong
3	strong	medium strong
4	medium strong	strong

smectite with a high expansion factor. The content of these high-expansion-factor clay minerals increased so that the salinity sensitivity of the reservoir rock increased in the later stages of the waterflooding treatment.

Based on all of the experimental results above, the waterflooding practices need to be modified in the presence of certain clay types. Due to the variation of kaolinite and illite, the velocity sensitivity got weaker and salinity sensitivity got stronger. The original waterflooding practice will not be suggested to conduct. The enhanced oil recovery methods, such as low-salinity waterflooding,[27] polymer and surfactant flooding,[28] and alkaline surfactant−polymer flooding,[29] can be emphasized for this reservoir in the future development. For the preferential channel existing in the reservoir, the block technique should be applied, such as plugging particle[30] and preformed particle gel.[31]

The mercury intrusion data provide details about the pore structure. We conducted mercury intrusion experiments on more than two hundred samples from the four stages, which were selected from the five hundred samples. The mercury intrusion data demonstrate that the trend of porosity and permeability variations is the same as those measured by a laboratory porosimetry and a permeameter. From Stages 1 through 4, the average displacement pressure decreased continuously about 38.39% during the waterflooding process. This suggests that the waterflooding treatment increased the biggest pore throat radius and thus decreased the displacement pressure. Meanwhile, the average structure coefficient was also reduced by about 81.7%. The structure coefficient indicates the difference between the real pore structure and the pore structure of parallel capillary bundles having the same pore size. The reduction of the average structure coefficient illustrated that small-sized clay minerals were dislodged, leading to improved pore connection during the waterflooding process. The average sorting coefficient and the relative sorting coefficient decreased by 33.73 and 80.79%, respectively, from Stages 1 to 4. A smaller sorting coefficient implies a more heterogeneous distribution of pore size. Based on the above-mentioned results of mercury intrusion experiments, the mathematical relationships between these parameters will be



(a) before transformation **(b) after transformation**

Figure 9. Frequency distributions of permeability (a) before and (b) after the logarithmic transformation. The distribution of (b) is much closer to the normal distribution.

https://dx.doi.org/10.1021/acsomega.0c02374
ACS Omega 2020, 5, 20922−20931

investigated through MLR modeling. Figure 9 illustrates the permeability distributions before and after the logarithmic transformation. The frequency distribution of permeability is an extremely skewed distribution (Figure 7a). To avoid the possible estimation bias caused by this skewness in permeability, we chose to make a log transformation to permeability.[32] Specifically, Figure 9b shows that the permeability of the core samples approximately approaches a normal distribution after this log transformation, which confirms that the permeability of natural geological formations can be modeled using a log-normal distribution.

Figure 10 illustrates the standardized residuals versus fits the plot. It is a scatter plot of standardized residuals on the *y*-axis

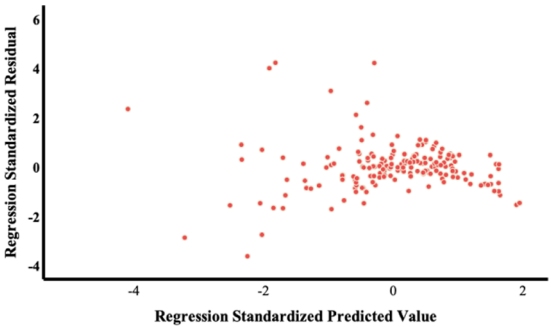


Figure 10. Regression standardized predicted value versus regression standardized residual.

and standardized predicted values on the *x*-axis. The standardized residual is defined as the residual divided by its standard deviation. Its value immediately tells us how many estimated standard deviations any point is away from the fitted regression model.[33] The predicted values are the responses predicted by the fitted MLR model. The standardized predicted value is calculated by subtracting the mean predicted value from the predicted values. Then, this difference is divided by the standard deviation of the predicted values. Similar to the standardized residuals, the standardized predicted values create a scale based on its standard deviation.

From the following figure, we could see that most of the standardized residuals "bounce randomly" around the 0 line without any clear pattern. This indicates that the fitted linear model is reasonable. An outlier is a point whose standardized residual is greater than 2.5 or smaller than −2.5 might be flagged as a potential outliner. These potential outliners could be influential points that bias the fitted linear trend. Therefore, we deleted these outliers from the dataset to meet the no-potential-outlier assumption.

When it turns to the noncollinearity assumption, Table 3 illustrates the bivariate correlation matrix. The range of the correlation coefficient ranges from 0.299 to 0.742, which are all below 0.8. Therefore, there is no perfect collinearity issue in this model.

The next step is to check the goodness of fit for the MLR model. Table 4 presents information about the goodness of fit. In Model 1, the predictor porosity is included first in the model. A total of 76.6% of the total variance in permeability (R^2 = 0.766 in Model 1) was explained by the predictor porosity. The mean pore radius was then added into the regression equation as the second variable in Model 2. A total of 82.4% of the variance in permeability was explained by porosity and the pore radius. Model 3 includes the predictors

Table 3. Correlation between Predictors Including Porosity (Φ), Sorting Coefficient (S_p), Mean Pore Radius (r), Homogeneous Coefficient (α), and Displacement Pressure (P_d)

	φ	S_p	$\overline{r}$	α	P_d
φ	1	0.625[a]	0.742[a]	0.661[a]	−0.607[a]
S_p		1	0.416[a]	0.403[a]	−0.665[a]
$\overline{r}$			1	0.721[a]	−0.369[a]
α				1	−0.299[a]
P_d					1

[a]Correlation is significant at the 0.01 level (2-tailed).

Table 4. Goodness-of-Fit Information[a]

model	*R* square	std. error of the estimate	VIF
1	0.766	0.35630	1.000
2	0.824	0.30977	2.223, 2.223
3	0.848	0.28876	3.044, 2.242, 1.656

[a]Model 1 includes the predictor of porosity; Model 2 includes the predictors of porosity and mean pore radius; and Model 3 includes the predictors of porosity, mean radius, and sorting coefficient.

of porosity, mean radius, and sorting coefficient, resulting in that a total of 84.8% of the variance in permeability was explained by all of the predictors. The fourth and fifth predictors, the homogeneous coefficient and displacement pressure, were not included in the final MLR model because including these two predictors in the model cannot significantly increase the overall variance of permeability explained by the model.

After all of the checks of the goodness of fit, the three predictors, the porosity, the sorting coefficient, and the mean radius, remained in the MLR model with the highest value of R^2. The coefficients of these predictors in the MLR model were then estimated using SPSS 24.0. As shown in Table 5, the standardized coefficients of the predictors are 0.468 for porosity (t = 9.118, p < 0.05), 0.199 for the sorting coefficient (t = 5.263, p < 0.05), and 0.381 for the mean radius (t = 8.652, p <0.05). All of these coefficients are significantly different from 0 when α = 0.05.

The column B under the unstandardized coefficients section shows all of the unstandardized regression coefficients for the predictors in different stages; the β column under the standardized coefficients section indicates the corresponding standardized regression coefficient; in fact, the standardized coefficients are equal to the regression coefficients estimated from using the normalized variables to fit the same model.[34] These standardized coefficients removed the unit of measurement of predictors and outcome variables so that we can compare the relative effects of predictors measured on different scales. t is the test statistics; Sig is the probability of obtaining the test results at least as extreme as the results actually observed, assuming that the null hypothesis is correct. Tolerance and VIF are statistics indicating multicollinearity. Therefore, the final mathematical regression equation using standardized coefficients for the core samples extracted from Stage 2 is written as follows

$$\log_{10}(Y + 17) = 0.468X_1 + 0.199X_2 + 0.381X_3 \quad (3)$$

Figure 11 illustrates the distribution of the standardized residual after the linear regression model was fitted to the data. The distribution shows a bell-shaped curve with its mean

https://dx.doi.org/10.1021/acsomega.0c02374
ACS Omega 2020, 5, 20922−20931

Table 5. Multiple Linear Regression Coefficients of Different Models

		unstandardized coefficients		standardized coefficients			collinearity statistics	
	model	B	std. error	β	t	sig.	tolerance	VIF
1	(constant)	−2.216	0.171		−12.939	0.000		
	porosity (%)	0.162	0.007	0.875	24.111	0.000	1.000	1.000
2	(constant)	−1.355	0.187		−7.263	0.000		
	porosity (%)	112	0.009	0.608	12.928	0.000	0.450	2.223
	mean radius (μm)	0.150	0.020	0.360	7.648	0.000	0.450	2.223
3	(constant)	−1.900	0.202		−9.387	0.000		
	porosity (%)	0.086	0.009	**0.468**	9.118	0.000	0.329	3.044
	mean radius (μm)	0.159	0.018	**0.381**	8.652	0.000	0.446	2.242
	sorting coefficient	0.299	0.057	**0.199**	5.263	0.000	0.604	1.656

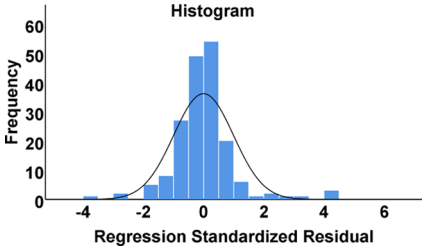


Figure 11. Histogram of the Regression Standardized Residual.

centering at zero and symmetrical on both tails. This verifies that the error terms that were left by fitting the linear model are normally distributed, which satisfies the assumption "Gaussian distribution of the random error".

Based on the residual versus fitted value plots (Figure 10), the distribution of residual plots, and the overall model fit statistics (Table 4), we decided to use the linear model to fit the data. The reasons are as follows: (1) Nonlinear trend among residuals could not be observed among residuals since all of them are evenly distributed along with the predicted values (Figure 10) and (2) The overall fitted statistics (R^2) of the linear model is very good, and the residual distribution itself totally satisfies the regression assumption.

Using the same model building procedure and evaluation criterion, the MLR equations using the standardized coefficients for the core samples extracted from Stages 1, 3, and 4 are written as follows

$$\text{Stage 1: } \log_{10}(Y + 17) = 0.861X_1 + 0.173X_2 \quad (4)$$

$$\text{Stage 3: } \log_{10}(Y + 17) = 0.557X_1 + 0.418X_3 \quad (5)$$

$$\text{Stage 4: } \log_{10}(Y + 17) = 0.452X_1 + 0.565X_3 \quad (6)$$

where Y is the permeability, X_1 is the porosity, X_2 is the sorting coefficient, and X_3 is the mean pore radius.

The variations of the coefficients in the MLR models through the four development stages during the waterflooding treatment agree with the petrophysical property variations measured in the laboratory experiments. In Stage 1, the reservoir had not been significantly influenced by the waterflooding treatment. In this situation, the original porosity and grain sorting coefficient can primarily decide the permeability of the reservoir rock. In Stage 2, the original pore structure had been influenced by the injection water. Fine-sized clay mineral particles that originally filled in the pore spaces were dislodged, leading to a larger pore radius. Therefore, the porosity, grain sorting coefficient, and mean pore radius all had an impact on the permeability. In Stage 3, more and more large-radius pores became connected so high-permeability flow channels started to develop, which greatly determine the permeability of the reservoir rock. The grain sorting coefficient played a negligible role and thus was removed from the MLR model. In Stage 4, the weight coefficient of the mean pore radius increased, whereas the weight coefficient of the porosity decreased, which implies that the permeability was controlled by the high-permeability flow channels to a larger extent.

Although the fitted statistics model is consistent with the findings from our experiment, caution should be raised when we try to use the fitted model to predict the effect of the relative weights of these variables on the permeability in the future stage. The current model may be more reliable under the current method. However, when a new method is invented and applied, this statistical relationship may also change. Therefore, future studies can continue to examine the feasibility of the current model under the application of the new method. Moreover, further research will consider other variables such as geological members and irreducible water saturation. The uncertainty quantification will also be applied in the future investigation.

CONCLUSIONS

A significant effect on the reservoir properties was brought by the waterflooding treatment, especially on the variations of permeability and porosity. Besides, the velocity sensitivity of the reservoir got decreased, whereas the salinity sensitivity increased during the development. The XRD and SEM results confirm the migration and dissolution of the clay mineral components. The mercury intrusion data not only provided more details of the pore structure but also contributed to the MLR model developing. The final MLR models agree with the petrophysical property variations of the reservoir rock measured in the laboratory. The variations of the predictor coefficients in the MLR model indicated the changes in the reservoir petrophysical characteristics through the four development stages. Further research will consider other model variables and field operations. The findings from this study advance the fundamental understanding of the role of long-term waterflooding treatment on oilfield developments. This work also provides insight into the combination of reservoir petrophysical properties and MLR modeling. The outcome of this research will benefit the reservoir development with similar conditions and geostatistical models by providing insightful correlations between various reservoir petrophysical properties.

https://dx.doi.org/10.1021/acsomega.0c02374
ACS Omega 2020, 5, 20922−20931

AUTHOR INFORMATION

Corresponding Author

Xiangming Li − *College of Geoscience, Yangtze University, Wuhan 430100, Hubei, China; Hubei Cooperative Innovation Center of Unconventional Oil and Gas, Wuhan 430100, Hubei, China*; Email: lixiangming@yangtzeu.edu.cn

Authors

Zihao Li − *Department of Mining and Minerals Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States*

Chenguang Du − *School of Education, Virginia Tech, Blacksburg, Virginia 24061, United States*

Yongqiang Tang − *Petroleum Exploration and Production Research Institute, Sinopec, Beijing 100083, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c02374

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the National Natural Science Foundation of China (No. 41972098) for their financial support to carry out this research. The authors also thank the support from the Virginia Tech Open Access Subvention Fund. The collaborators from the Statistical Applications & Innovations Group (SAIG) at Virginia Tech will also be thanked for their assistance and comments on the regression model. The insightful and constructive comments of the anonymous reviewers are also gratefully acknowledged.

REFERENCES

(1) Craig, F. F., Jr. *The Reservoir Engineering Aspects of Waterflooding, Monograph Series*; Society of Petroleum Engineers: Richardson, Texas, 1971; Vol. 3.

(2) Bryant, S.; Cade, C.; Mellor, D. Permeability prediction from geologic models. *AAPG Bull.* **1993**, *77*, 1338−1350.

(3) Katz, A. J.; Thompson, A. H. Quantitative prediction of permeability in porous rock. *Phys. Rev. B* **1986**, *34*, 8179−8181.

(4) Ramm, M.; Bjørlykke, K. Porosity/depth trends in reservoir sandstones: Assessing the quantitative effects of varying pore-pressure, temperature history and mineralogy, Norwegian Shelf data. *Clay Miner.* **1994**, *29*, 475−490.

(5) Van, S. L.; Chon, B. H. Effective prediction and management of a CO_2 flooding process for enhancing oil recovery using artificial neural networks. *J. Energy Resour. Technol.* **2018**, *140*, No. 032906.

(6) Huang, Z.; Shimeld, J.; Williamson, M.; Katsube, J. Permeability prediction with artificial neural network modeling in the Venture gas field, offshore eastern Canada. *Geophysics* **1996**, *61*, 422−436.

(7) Ali, M.; Chawathé, A. Using artificial intelligence to predict permeability from petrographic data. *Comput. Geosci.* **2000**, *26*, 915−925.

(8) Chen, C. H.; Lin, Z. S. A committee machine with empirical formulas for permeability prediction. *Comput. Geosci.* **2006**, *32*, 485−496.

(9) Helmy, T.; Fatai, A.; Faisal, K. Hybrid computational models for the characterization of oil and gas reservoirs. *Exp. Syst. Appl.* **2010**, *37*, 5353−5363.

(10) Ahangar-Asr, A.; Faramarzi, A.; Mottaghifard, N.; Javadi, A. A. Modeling of permeability and compaction characteristics of soils using evolutionary polynomial regression. *Comput. Geosci.* **2011**, *37*, 1860−1869.

(11) Naraghi, M. E.; Javadpour, F. A stochastic permeability model for the shale-gas systems. *Int. J. Coal Geol.* **2015**, *140*, 111−124.

(12) Seber, G. A.; Lee, A. J. *Linear Regression Analysis*, 2nd ed.; John Wiley & Sons: Hoboken, New Jersey, 2012; Vol. *329*.

(13) Yamane, T. *Statistics: An Introductory Analysis*, 2nd ed.; A Harper International Edition: Buffalo, New York, 1973.

(14) Dahraj, N. U.; Bhutto, A. A. In *Linear Mathematical Model Developed Using Statistical Methods to Predict Permeability from Porosity*, PAPG/SPE Pakistan Section Annual Technical Conference; Society of Petroleum Engineers, 2014.

(15) Cook, R. A.; Hover, K. C. Mercury porosimetry of cement-based materials and associated correction factors. *Constr. Build. Mater.* **1993**, *7*, 231−240.

(16) Li, X. Fuyu Core Data and Fine Sand Structured Analysis. Master Thesis, Northeast Petroleum University: Daqing, China, 2015.

(17) Park, H. M. Univariate Analysis and Normality Test Using SAS, Stata, and SPSS. Ph.D. Dissertation, Indiana University: Bloomington, IN, 2015.

(18) Pearson, K. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *London, Edinburgh Dublin Philos. Mag. J. Sci.* **1900**, *50*, 157−175.

(19) Everitt, B. *The Cambridge Dictionary of Statistics*; Cambridge University Press: Cambridge, U.K., 1998.

(20) Lopes, R. H. C. Kolmogorov-Smirnov Test. In *International Encyclopedia of Statistical Science*; Springer: Berlin, Heidelberg, 2011; pp 718−720.

(21) Garson, G. D. *Testing Statistical Assumptions*; Statistical Associates Publishing: Asheboro, NC, 2012.

(22) Bager, A.; Roman, M.; Algedih, M.; Mohammed, B. In *Addressing Multicollinearity in Regression Models: A Ridge Regression Application*, 11th International Conference of Applied Statistics, Brasov, 2017.

(23) Tabachnick, B. G.; Fidell, L. S.; Ullman, J. B. *Using Multivariate Statistics*, 7th ed.; Pearson: Boston, MA, 2007; Vol. *5*.

(24) Stevens, J. P. Outliers and influential data points in regression analysis. *Psychol. Bull.* **1984**, *95*, 334−344.

(25) Andersen, C. M.; Rasmus, B.. Variable selection in regression—a tutorial. *J. Chemom.* **2010**, *24*, 728−737.

(26) Chatterjee, S.; Hadi, A. S. *Regression Analysis by Example*, 5th ed.; John Wiley & Sons: Hoboken, NJ, 2015.

(27) Du, Y.; Xu, K.; Mejia, L.; Zhu, P.; Balhoff, M. T. Microfluidic Investigation of Low-Salinity Effects During Oil Recovery: A No-Clay and Time-Dependent Mechanism. *SPE J.* **2019**, *24*, 2841−2858.

(28) Shah, D. O. *Improved Oil Recovery by Surfactant and Polymer Flooding*; Academic Press: NY, 2012.

(29) Li, Z.; Zhang, W.; Tang, Y.; Li, B.; Song, Z.; Hou, J. Formation damage during alkaline-surfactant-polymer flooding in the Sanan-5 block of the Daqing Oilfield, China. *J. Nat. Gas Sci. Eng.* **2016**, *35*, 826−835.

(30) Zhao, F.; Li, Z.; Wu, J.; Hou, J.; Qu, S. New type plugging particle system with high temperature & high salinity resistance. *J. Pet. Sci. Eng.* **2017**, *152*, 317−329.

(31) Bai, B.; Zhang, H. Preformed-particle-gel transport through open fractures and its effect on water flow. *SPE J.* **2011**, *16*, 388−400.

(32) Xue, G.; Datta-Gupta, A.; Valko, P.; Blasingame, T. Optimal transformations for multiple regression: application to permeability estimation from well logs. *SPE Form. Eval.* **1997**, *12*, 85−94.

(33) Sheather, S. Diagnostics and Transformations for Multiple Linear Regression. In *A Modern Approach to Regression with R. Springer Texts in Statistics*; Springer: New York, NY, 2009.

(34) Johan, B. How to Standardize Regression Coefficients. *Am. Stat.* **1994**, *48*, 209−213.

https://dx.doi.org/10.1021/acsomega.0c02374
ACS Omega 2020, 5, 20922−20931