Correlations of Permeability and Geological Characteristics Based on Mercury Intrusion Data and Hierarchical Statistical Models: A Case Study

Zihao Li
Department of Mining and Minerals Engineering,
Virginia Tech,
Blacksburg, VA 24060
e-mail: zihao91@vt.edu

Wenyu Gao
Department of Statistics,
Virginia Tech,
Blacksburg, VA 24060;
Statistical Applications and Innovations Group (SAIG),
Virginia Tech,
Blacksburg, VA 24060
e-mail: wenyu6@vt.edu

Xiangming Li[1]
College of Geosciences,
Yangtze University,
Wuhan, Hubei, 430100, China;
Hubei Cooperative Innovation Center of Unconventional Oil and Gas,
Wuhan, Hubei, 430100, China
e-mail: lixiangming@yangtzeu.edu.cn

The mercury intrusion technique is a crucial in-lab method to investigate the porous medium properties. The potentiality of mercury intrusion data has not been explored significantly in the traditional interpretation. Thus, a hierarchical statistical model that not only captures the quantitative relationship between petrophysical properties but also accounts for different geological members is developed to interpret mercury intrusion data. This multilevel model is established from almost 800 samples with specific geological characteristics. We distinguish the fixed effects and the random effects in this mixed model. The overall connection between the selected petrophysical parameters is described by the fixed effects at a higher level, while variations due to different geological members are accommodated as the random effects at a lower level. The selected petrophysical parameters are observed through hypothesis testing and model selection. In this case study, five petrophysical parameters are selected into the model. Essential visualizations are also provided to assist the interpretations of the probabilistically model. The final model reveals the quantitative relationship between permeability and other petrophysical properties in each member and the order of relative importance for each property. With this studied relationship and advanced model, the geological reservoir simulation can be greatly detailed and accurate in the future. [DOI: 10.1115/1.4047327]

Keywords: geological member, hierarchical data structure, petrophysical properties, mercury intrusion data, oil/gas reservoirs, petroleum engineering

1 Introduction

Fossil fuels will remain an important energy source for the next several decades [1]. As an essential petrophysical property, permeability represents the capacity of the reservoir to determine hydrocarbon recovery. It is essential for the petroleum investigators to have an accurate insight into reservoir permeability. The percentage of petroleum products from the conventional reservoir in China was about 80% in 2017 [2]. Therefore, more efforts need to be extended to investigate the controlling factors for the permeability of the conventional reservoir. Meanwhile, much of the effort focuses on correlating permeability with a broad range of sandstone petrophysical properties. The relationship between permeability and these petrophysical properties have various forms, including simple statistical correlations and property converted correlations [3]. However, there is no universal mathematical relationship between permeability and these petrophysical properties so far. Besides, the features of different geological members noticeably have a substantial impact on the quantitative relationship among the petrophysical parameters. A quantitative relationship allowing for multiple petrophysical properties is necessary to provide guidance on oil extraction. The oilfield, especially mature oilfield, will be benefitted a lot in product prediction, drilling design, and enhanced recovery from monitoring petrophysical properties and their variation. To practically apply these permeability predictions, detailed data on every single core sample is usually required, primarily through the in-lab techniques such as mercury intrusion experiment.

The mercury intrusion experiment is one of the most critical methods to investigate the pore structure of core samples by measuring petrophysical properties. Petrophysical properties have a great impact on reservoir characteristics and oilfield production. Except for permeability, these petrophysical properties include porosity, sorting coefficient, mean pore radius, and displacement pressure, to name a few. Meanwhile, all (or most) petrophysical properties usually interact with each other because they are dominated by the same or similar physical laws [1]. During the development, amounts of mercury intrusion data have accumulated based on the varieties of geological members. These data offer the possibility to be analyzed by a statistical model. In this paper, we study the permeability prediction in different geological members by mercury intrusion data using a hierarchical statistical model. With a better understanding of such a relationship, the reservoir simulation will yield a more accurate performance prediction.

1.1 Statistical Model. When data have different sources of variations, the traditional models will lose their power. The possession of hierarchy in the data causes correlations among observations and thus violates the independence assumption for conventional models. In our case, the variation comes from two sources: individual core samples and the geological members where the core samples are drilled from. It is reasonable to believe that core samples from the same member will share some characteristics in common, while dissimilar with the samples from other members. That is, the core samples are not independent to each other. Rather, they are correlated at a higher level (the member level). As a result, a linear mixed model [4] will be considered, as this is

[1]Corresponding author.

Contributed by the Petroleum Division of ASME for publication in the Journal of Energy Resources Technology. Manuscript received February 17, 2020; final manuscript received May 10, 2020; published online June 25, 2020. Assoc. Editor: Daoyong (Tony) Yang.

the conventional model to analyze continuous data with hierarchical correlations. A mixed model is a combination of fixed effects and random effects. The concept of random effects was first introduced by Fisher [5], and the estimation techniques for fixed and random effects were invented by the previous works [6]. With a mixed model, the variations at the geological member level can be accommodated by the random effects, allowing us to differentiate them from the individual core sample variations.

Hypothesis testing is one of the statistical inference techniques. We usually propose a "null hypothesis" and an "alternative" a priori. In regression models, if not specified, the null hypothesis is generally that the coefficient of the variable is equal to zero. We then arrive at a conclusion based on the observed data that either "reject" or "fail to reject" the null hypothesis [7]. A typical way is to derive a probability of observing the observation or more extreme values, which is called a p-value. If the p-value is smaller than 0.05, we will "reject" the null hypothesis. The statistical analyses are performed using R [8], which is one of the most commonly used statistical software for statisticians. This free software allows users to write their own codes, which provides more flexibility. In this paper, the linear mixed model is developed with the *lme* function in the *nlme* package, while the hypothesis testing uses the *linearHypothesis* function from the *car* package. Besides, the exploratory plots are generated with the *ggplot2* package, and the sensitivity analysis is performed with the *r2beta* function from the *r2glmm* package.

1.2 Permeability Prediction and Statistical Model. The combinations of permeability prediction and the statistical model are investigated as a promising research area. A variety of algorithms have been applied to permeability prediction from different research fields. Because of the enormous complexity in the oil and gas reservoir or coal seams, the high-qualified permeability model is a formidable task [9]. The predictive model consists of various mechanisms from fundamental understanding. Integrated with these conditions, the computer code can be developed for predictive models with certain initial and boundary conditions. Plenty of proposed permeability models attempts to account for the mechanisms within different background environments. Ramm and Bjørlykke [10] also considered the relationship between reservoir geological settings and petrophysical properties. They found that pre-burial mineralogy was a significant point in porosity prediction of sandstones. But most of the previous works take attention on the formation scale or in-lab scale, instead of geological member scale. The impact of specific characteristics of the members has not been focused on in the research.

Except for absolute permeability, relative permeability prediction is also super critical. The oil recovery is associated with relative permeability, especially in the water-flooding reservoir, because relative permeability is often represented as a function of water saturation. For the 2D/3D permeability experiments and tight formations, the traditional methods find their limitations in determining the absolute permeability and relative permeability. In this case, the history matching method shall be used [11–14]. The values from the model entirely agreed with three-phase relative permeability, which were used to evaluate the model. Blunt et al. [15] used a three-dimensional network model to compute relative permeability in porous media. The experimental data and proposed model predictions had a significant agreement.

1.3 Permeability Prediction With Mercury Intrusion Data. Mercury intrusion is a powerful technique for the evaluation of petrophysical properties to characterize a wide variety of porous media. Overall, the mercury intrusion experiment can provide detailed data on every single sample, which is required in the permeability prediction algorithm. That is why permeability prediction with mercury intrusion data is a unique interdisciplinary approach. Lapierre et al. [16] showed that there is a relationship between the pore-size parameters and the permeability of the clay. Al Hinai et al. [17] correlated permeability with capillary pressure from mercury intrusion measurement. They derived the equation between permeability, porosity, and pore-throat size at 75% mercury saturation.

The investigation of permeability prediction with mercury intrusion data is not limited in the aforementioned works in this section. More formulas have been derived from various algorithms. Moreover, the application of mercury intrusion experiment has expanded from traditional sandstone to unconventional rocks. Clarkson et al. [18] investigated nanopore structure in the tight reservoir with mercury intrusion techniques and low-pressure adsorption. Noorud-din et al. [19] presented a comparative study among the main soft computing algorithms. In addition, this technique is also applied in the investigation of the pore-throat structure with permeability in tight formations because the pore-throat structure dominates the permeability in tight formations [20–22]. Although the previous works about mercury intrusion experiment is numerous, the potentiality of mercury intrusion data has not been explored significantly in the traditional interpretation. Plenty of petrophysical information still has not been revealed under the mercury intrusion data.

1.4 Workflow. This section provides an overview of the workflow of the permeability and hierarchical statistical models with geological members based on mercury intrusion data. A statistical model/geological background-integrated approach (Fig. 1) is developed to investigate the quantitative relationship between permeability and other petrophysical properties, subjecting to specific geological member characteristics. First comes the exploratory analysis. Relationships between permeability and other petrophysical properties are visualized by a series of collected geological members. Visualization helps us identify the most appropriate model, as well as potential issues, such as outliers. The second stage is model fitting, we selected the linear mixed model and the general linear model, contributing to the quantitative relationship

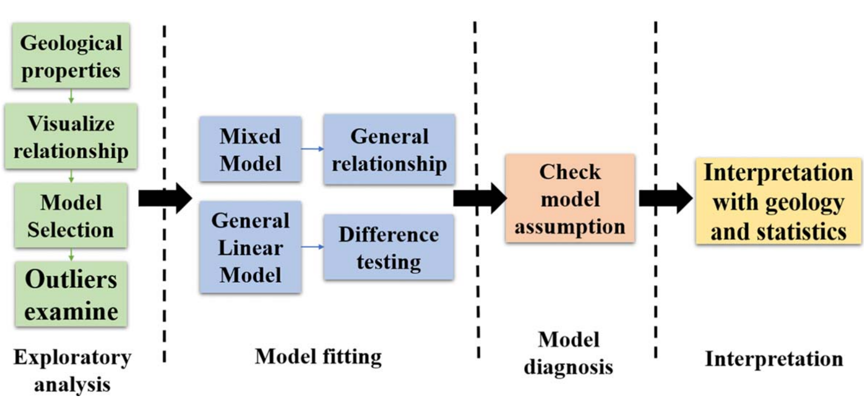

Fig. 1 A statistical model/geological background-integrated workflow

between the petrophysical properties, and the testing on differences among significant geological members, respectively. Third, model diagnosis is performed to check the validity of the models. The final stage is interpretation. A final most appropriate model is achieved to explain the relationship based on different geological behaviors. This model can be used for prediction of different geological settings.

We collected the mercury intrusion data of 797 core samples from a total of 11 members. The primary oil-producing members are member nj, qn1, q4, and qt, which belong to the same formation with the increasing depth from member nj to member qt. Meanwhile, the geological characteristics of the members that we selected are lithology, grain composition, interstitial, sedimentary microfacies, and diagenesis. For the hierarchical statistical model, the dependent variable is permeability (mD), and the potential independent predictors are five petrophysical parameters, which are porosity (%), sorting coefficient (dimensionless), mean pore radius (μm), homogeneous coefficient (dimensionless), and displacement pressure (MPa). Among the predictors, the sorting coefficient accounts for the distribution of the pore size. The higher sorting coefficient means the more uniform grain size. Also, the homogeneous coefficient describes the deviation between the pore radius and the maximum pore radius. The displacement pressure is the pressure at which the non-wettability fluid begins to enter into the maximum pore of the core sample. Through comparing geological characteristics with the coefficients of petrophysical properties in the statistical model, we can investigate the contribution quantitatively from geological characteristics to the values of petrophysical properties.

2 Geological Setting

Although we have the samples from 11 members, the main oil-producing members are only four of them: member nj, qn1, q4, and qt. We also have more samples from these four members than the others. Therefore, we will investigate the comparison of these four members specific geological settings in detail. The rock with different lithologies usually has different properties. However, the same lithology rock does not always have similar features, because the grain composition, sorting, and interstitial all have impacts on the reservoir properties. Except for the factors above, diagenesis also plays an essential role in the reservoir properties. Therefore, the geological characteristics of the members that we selected are lithology, grain composition, interstitial, sedimentary microfacies, and diagenesis phase.

Member nj, the sedimentary system is lacustrine-delta facies. The lacustrine facies include shore-shallow and semi-deep subfacies, which deposit the microfacies of shore-shallow lacustrine mud, shallow lacustrine mud, and beach-bar. The delta facies mainly includes delta front and prodelta subfacies, which deposited the microfacies of underwater distributary channel, estuarine bar, distal bar, front sheet sand, interdistributary bay, and prodelta mud. The reservoir lithology is siltstone and fine sandstone with the components of lithical arkose and feldspathic arenite. The sorting is medium-well. The psephicity is subangular-subrounded. The matrix-type is pelitic and quartz-feldspathic. The cementation type is mainly porous and contact-porous cementation with a little basal cementation. The adglutinate is calcite, kaolinite, and silica. The rock textural maturity is medium, and the mineral's maturity is medium-low.

Member qn1, the sedimentary system is delta facies. The delta facies mainly includes delta front and prodelta subfacies, which deposited the microfacies of underwater distributary channel, estuarine bar, underwater distributary channel overbank, front sheet sand, distal bar, and underwater interdistributary bay. The lacustrine facies also deposited a little, so the subfacies are semi-deep and deep lacustrine. The microfacies is turbidity sand. The reservoir lithology is siltstone and fine sandstone with the components of lithical arkose. The sorting is medium-good, and the psephicity is subangular-subrounded. The matrix-type is pelitic and calcareous. The cementation type is mainly porous and contact-porous cementation with a little basal cementation. The adglutinate is calcite, kaolinite, and silica. The rock textural maturity is medium, and the mineral maturity is medium-low.

Member q4, the central and lower member part is river-dominated nearshore shallow delta plain subfacies with the microfacies of distributary channel, point bar, crevasse splay, abandoned channel, natural levee, and interdistributary bay. The upper part is delta front subfacies. In the direction of lacustrine deposits, the microfacies is underwater distributary channel in the proximal delta front deposition, estuarine bar in the far deposition, and sheet sand and distal bar in the end area deposition. The reservoir lithology is siltstone and fine sandstone with the components of feldspathic arenite and little lithical arkose. The sorting is medium, and the psephicity is subangular. The matrix-type is pelitic and calcareous with thin film shape and recrystallization. The cementation type is mainly porous and contact-porous cementation with a little basal cementation. The adglutinate is calcite, kaolinite, and silica. The rock textural maturity is medium, and the mineral maturity is medium-low.

Member qt, the sedimentary system is far-shore delta distributary plain subfacies with the microfacies of distributary channel, point bar, crevasse splay, abandoned channel, natural levee, and interdistributary bay. The reservoir lithology is siltstone and fine sandstone with little coarse and medium sandstone. The rock components are feldspathic arenite and little lithical arkose. The sorting is medium, and the psephicity is subangular. The matrix-type is pelitic and quartz-feldspathic. The cementation type is mainly contacted cementation with a little basal cementation and porous cementation. The adglutinate is calcite, kaolinite, and silica. The rock textural maturity is medium, and the mineral maturity is medium-low.

3 Methodology

3.1 Exploratory Analysis. To find a proper model that can describe how permeability can be affected by porosity, sorting coefficient, mean pore radius, homogeneous coefficient, and displacement pressure, we start from exploratory analysis to examine the data by various geological members. If any member has a number of observations smaller than 5, we will combine it with other members if possible, or delete it, since too few observations cannot illustrate the relationship accurately. Extreme outliers with relatively high permeability are removed since such high permeability is not feasible. There are 11 members in total after data manipulation. Relationships between permeability and the other properties are plotted by the 11 members using scatterplot matrices. If all the members possess linear trends, a linear mixed model will be applied. Otherwise, necessary transformations on the variables will be used.

3.2 Linear Mixed Models. The linear mixed model is analogy to a linear regression model. If we fit a linear regression model, we can write the model as Eq. (1):

$$k_i = \beta_0 + \beta_1 \phi_i + \beta_2 a_i + \beta_3 b_i + \beta_4 c_i + \beta_5 d_i + \epsilon_i \quad (i = 1, 2, \ldots, n) \tag{1}$$

where k represents permeability, ϕ is porosity, a is sorting coefficient, b is mean pore radius, c is homogeneous coefficient, d is displacement pressure, $\beta_0 - \beta_5$ are fitting coefficients, $\epsilon_i \sim N(0, \sigma^2)$ represents the error from observation i, which follows a normal distribution with mean zero and shares a constant variance with other observations, and n is the total number of observations. The variables in the model may not be the raw data, but with transformations (hereafter). The linear regression model can express the general relationship among the variables, but it cannot capture the differences across members. Although the relationships are all linear across the 11 members, the intercepts and slopes may be different. Thus, we would like to use the linear mixed model to describe the general relationship among variables, as well as capture individual differences across members. If we only consider the differences in the intercept, then we can build an intercept-only linear mixed model (random intercept model) as Eq. (2):

$$k_{i,j} = \beta_0 + \beta_{0,j} + \beta_1\ \phi_{i,j}\ + \beta_2\ a_{i,j} + \beta_3\ b_{i,j} + \beta_4\ c_{i,j} + \beta_5\ d_{i,j} + \epsilon_{i,j}$$
$$(i = 1, 2, \ \ldots, n,\ j = 1, 2, \ \ldots, K) \tag{2}$$

Here, j represents the member where the specific observation comes from, and $K = 11$ is the total number of members. The difference compared with the linear regression model exists in β_0. Instead of one common value β_0, in a linear mixed model with random intercept, it becomes $\beta_0 + \beta_{0,j}$. Here, β_0 represents a mean value for all members, while $\beta_{0,j}$ represents the difference from member j to the mean value. In such a way, the general relationship among the variables can be represented using the mean value β_0, while the specific relationship for member j can be described by $\beta_0 + \beta_{0,j}$. Note the two parts of the intercept can also be combined into a single $\tilde{\beta}_{0,j} = \beta_0 + \beta_{0,j}$ notation wise.

To further increase the flexibility, we can consider a random slope model. Not only the differences in the intercept can be captured, the differences in the slopes (coefficients for the independent variables) are able to be detected as well. Similarly, we can write the model as Eq. (3), note the slopes can also be written as a single $\tilde{\beta}_{p,j},\ p = 1, 2, \ \ldots, 5$:

$$k_{i,j} = \beta_0 + \beta_{0,j} + (\beta_1 + \beta_{1,j})\ \phi_{i,j}\ + (\beta_2 + \beta_{2,j})\ a_{i,j}$$
$$+ (\beta_3 + \beta_{3,j})\ b_{i,j} + (\beta_4 + \beta_{4,j})\ c_{i,j} + (\beta_5 + \beta_{5,j})\ d_{i,j} + \epsilon_{i,j}$$
$$(i = 1, 2, \ \ldots, n,\ j = 1, 2, \ \ldots, K) \tag{3}$$

An important characteristic of the linear mixed model is that although accounting for the differences across members, the difference is not of particular interest. The model treats the differences of the intercepts and slopes, $\beta_{p,j}$, $p = 0, 1, 2, \cdots, 5$, as random. Instead of a constant number as β_p, we consider $\beta_{p,j} \sim N(0,\ \sigma_p^2)$ for different member j. That is, we can attain estimations for $\beta_{p,j}$, $p = 0, 1, 2, \cdots, 5$, but we cannot conduct hypothesis testing on them to provide evidence that the differences between specific members are not by chance. Thus, they are called random effects. However, we can perform hypothesis testing on the mean estimates of the parameters, β_p, to show the statistical significance of variable p in affecting permeability. The β_p's are called fixed effects. Although no testing can be performed on a particular $\beta_{p,j}$, they are able to accommodate the variations due to different members.

Model selection is based on criteria Akaike information criterion (AIC), Bayesian information criterion (BIC), log-likelihood, and likelihood ratio test (only available for nested models). Since we have assumptions that the errors and random terms have normal distributions, each model can obtain a likelihood. The number represents how likely the response variable takes the observed value, and thus the higher, the better. However, if we have more predictors in the model, the likelihood will tend to increase, so the model may be overfitted. Thus, AIC or BIC will be considered. They are calculated based on likelihood but with different penalties on the number of predictors in the model. The smaller the value of AIC or BIC, the better the model is. The likelihood ratio test provides a solid test to show whether two models are essentially the same. If two nested models are different only by chance, we can always choose the simpler model.

After the final model is determined, diagnoses will be performed on that final model to check for the validities of assumptions. We will want residuals from each member to have zero mean, random pattern, constant variances, and normal distributions. Besides, we will also want to check the multicollinearity among the predictors in the model. Multicollinearity means there is a linear relationship among the predictors. If this issue exists, the estimations of the parameters will not be reliable. Furthermore, sensitivity analysis is conducted to show the relative importance of each independent variable by semi-partial R^2. Partial R^2 in the linear regression model represents the percentage of variability explained by a particular variable from the total variation accounting for all the other variables. The semi-partial R^2 is an analogy used in linear mixed models. Variables with a higher semi-partial R^2 have more impact on explaining the variability of permeability.

3.3 General Linear Models. If all the assumptions are valid, we will confirm the final model to explain the general relationships among the variables. With the geological setting consideration, our next step is to conduct hypothesis testing to illustrate that the main members are different. To investigate this, we fit a general linear model, which treats members as a predictor in the model, which can be written as Eq. (4):

$$k = \beta_0\ + \beta_1\ \phi_i\ + \beta_2\ a_i + \beta_3\ b_i + \beta_4\ c_i + \beta_5\ d_i + \beta_6 M_i$$
$$+\ \beta_{1,6}\phi_i * M_i +\ \beta_{2,6}a_i * M_i +\ \beta_{3,6}c_i * M_i +\ \beta_{4,6}c_i * M_i$$
$$+ \beta_{5,6}d_i * M_i + \epsilon_i \quad (i = 1, 2, \ \ldots, n) \tag{4}$$

where M is geological member. This model looks like linear regression. However, since M is a categorical variable, it becomes different. Statistical significance of variable M means differences among members in the intercept level is not due to chance. Significance in the interactions means that the differences lie in the slope of the corresponding variable. We cannot replace this general linear model to the previous mixed model because the newly added term M and the corresponding interactions will affect the estimations and testing results of the other coefficients. Besides, this model does not reflect the general relationship among properties. Members nj, qn1, q4, and qt are the main oil-producing members. As a result, we would like to focus on comparisons among these members. We conduct hypothesis testing between each pair of them.

4 Results and Interpretation

4.1 Exploratory Analysis. Figure 2 includes the scatterplot matrix for permeability against porosity by the member. It is observable that the relationships between porosity and permeability are nonlinear. Other distributions are also considered for the goodness-of-fit. However, common distributions, such as log-normal or gamma, do not work well. As a result, we consider transformations on the variable. Natural log transformation is regarded as it is a common choice. After transformation (shown in Fig. 3), the relationships become linear, and the distribution of Log(permeability) also becomes normal.

Due to the page limit, we can only display one scatterplot matrix in the paper. The other variables follow similar patterns. Thus, we determine to use Log(permeability) as the response variable to fit the model. Since the units of independent variables do not align, we standardize each predictor, so they are on the same scale. Note that standardization will not change whether the coefficient of a particular predictor is significant or not. It only turns in the estimated values. Yet, it can be helpful in comparisons among predictors. No predictor will dominate the others, and the magnitudes of the estimated coefficients can also reflect the relative importance in predicting permeability. Also, since standardization centers the predictor values, the estimated intercept will have a significant change.

4.2 Linear Mixed Model. We start from the random intercept model with all five predictors. We determined from the results of the exploratory analyses that some predictors do not possess perfect linear relationships. Thus, we consider log transformations on these predictors and compare them with the original forms. If a model contains insignificant variables, we would also perform backward selection manually by removing the predictor with the highest p-value until all predictors are significant. Then, a

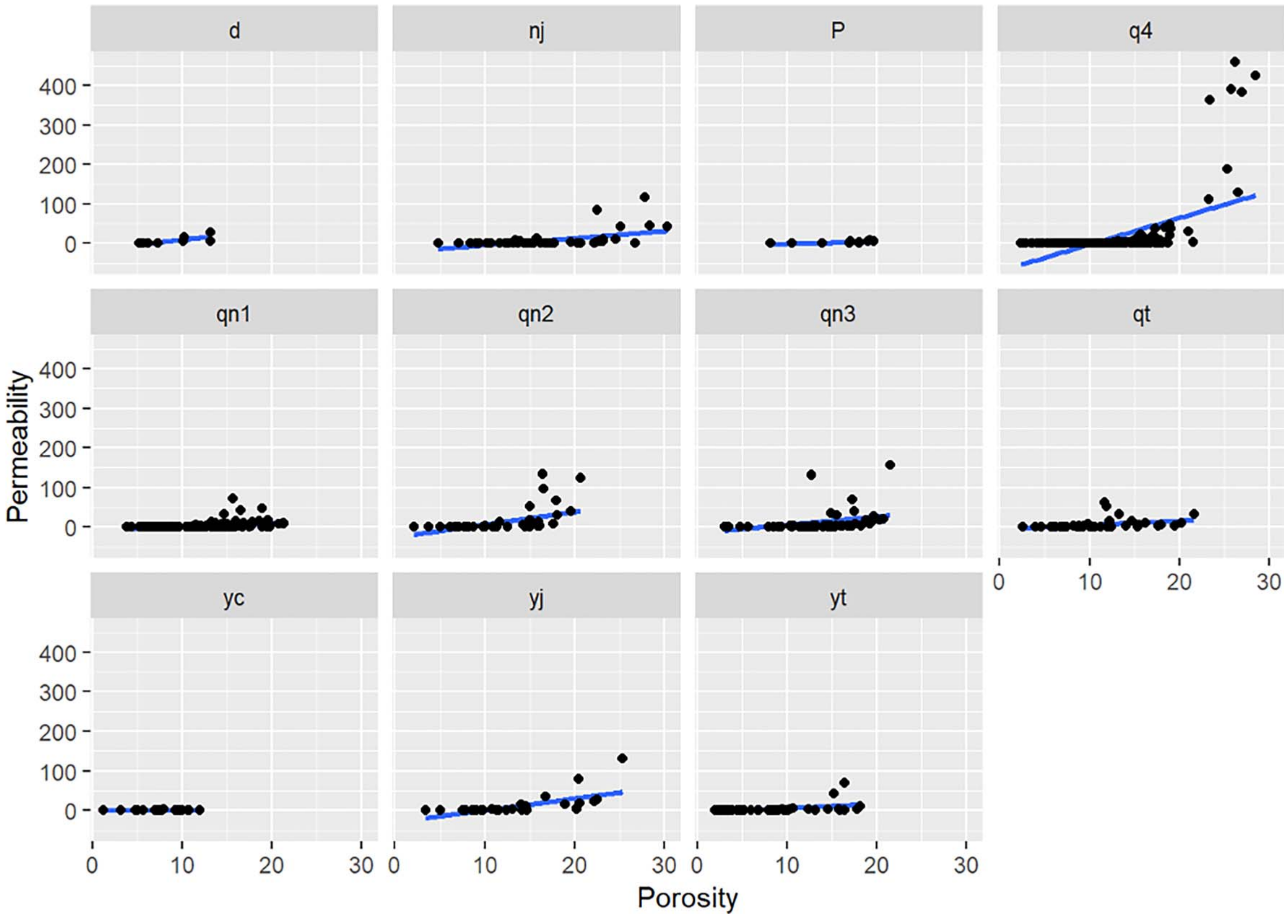

Fig. 2 Permeability versus porosity by all 11 members

likelihood ratio test is conducted on all nested models. We select the best model among various transformations and combinations of predictors, considering AIC, BIC, log-likelihood, and likelihood ratio test together. Finally, we choose the model with all five predictors included, taking the log transformation only on Predictor Displacement_Pressure. We then consider the random slope model and follow the same procedure. The selected model contains the same combination and transformation of predictors as the random intercept model.

The next step is to compare the random intercept and the random slope model. Since the two best models contain precisely the same set of predictors, they are nested. A likelihood ratio test is thus valid, with a p-value <0.0001, so we cannot reduce to the simpler model, which is the random intercept model. To be more cautious about the results, we also compare across other candidates from both random intercept and random slope models, although they are not selected within each category. Based on various criteria, our final model to explain the general relationship among variables is the full model with random slopes and log transformation on Displacement_Pressure. The corresponding estimates, standard errors, degrees-of-freedom, test statistics, and p-values for the coefficients from fixed effects are listed in Table 1. All five predictors from petrophysical properties are statistically significant.

Based on Eq. (3) and Table 1, we obtain Eq. (5):

$$\begin{aligned}\text{Log}(k) = &-0.0433 + 0.7700\ \phi_{i,j}\ + 0.1225\ a_{i,j} + 0.9763\ b_{i,j}\\ &-0.1391\ c_{i,j} - 0.7041\ \text{Log}(d_{i,j}) + \epsilon_{i,j}\\ &(i = 1, 2,\ \ldots, n,\ \ j = 1, 2,\ \ldots, K)\end{aligned} \tag{5}$$

We can have specific quantitative relationships in every member, as listed in Table 2. With random effects, the final weights from every predictor are significantly different in each member. In the fixed effect equation and all the random effect equations, the porosity, sorting coefficient, and mean pore radius usually have a positive impact on permeability. In contrast, homogeneous coefficient and Log(displacement pressure) typically hurt permeability.

We continue to check the validity of model assumptions. Figure 4 illustrates that our assumptions are generally satisfied except for a few outliers, by checking the residuals universally and by the member. Residuals are located around zero with no particular pattern. Although the variations are not precisely the same, they spread out in the same area randomly. The only problems are the outliers, most of which appear in member q4. This phenomenon makes sense since q4 contains more data points and thus would be more heterogeneous. The outliers are due to very large or small values of permeability and still fall in reasonable ranges. Thus, we decide to keep them in the model after careful consideration. Except for the outliers, the mean zero and constant variance assumptions are valid. Even with the outliers, the deviations are not significant, and no particular pattern indicates us to change the model.

As Fig. 5 illustrates, to further investigate the performance of our model, we compare the fitted values to the truth. Here, fitted values mean the estimated or calculated permeability values using our model with estimated coefficients and observed values for the predictors. Although the observed permeability values have been used to determine the coefficients, it is still reasonable to compare the fitted ones with the truth, since all observations are used in the estimation while one single value is compared. The observed versus fitted values plot is still a popular way for goodness-of-fit check. As compared with the observed permeability values (taking log-transformation), we can see that our estimations fit well with the truth, which means our model is reasonable. Besides, the variance inflation factor (VIF) is considered to check for multicollinearity.

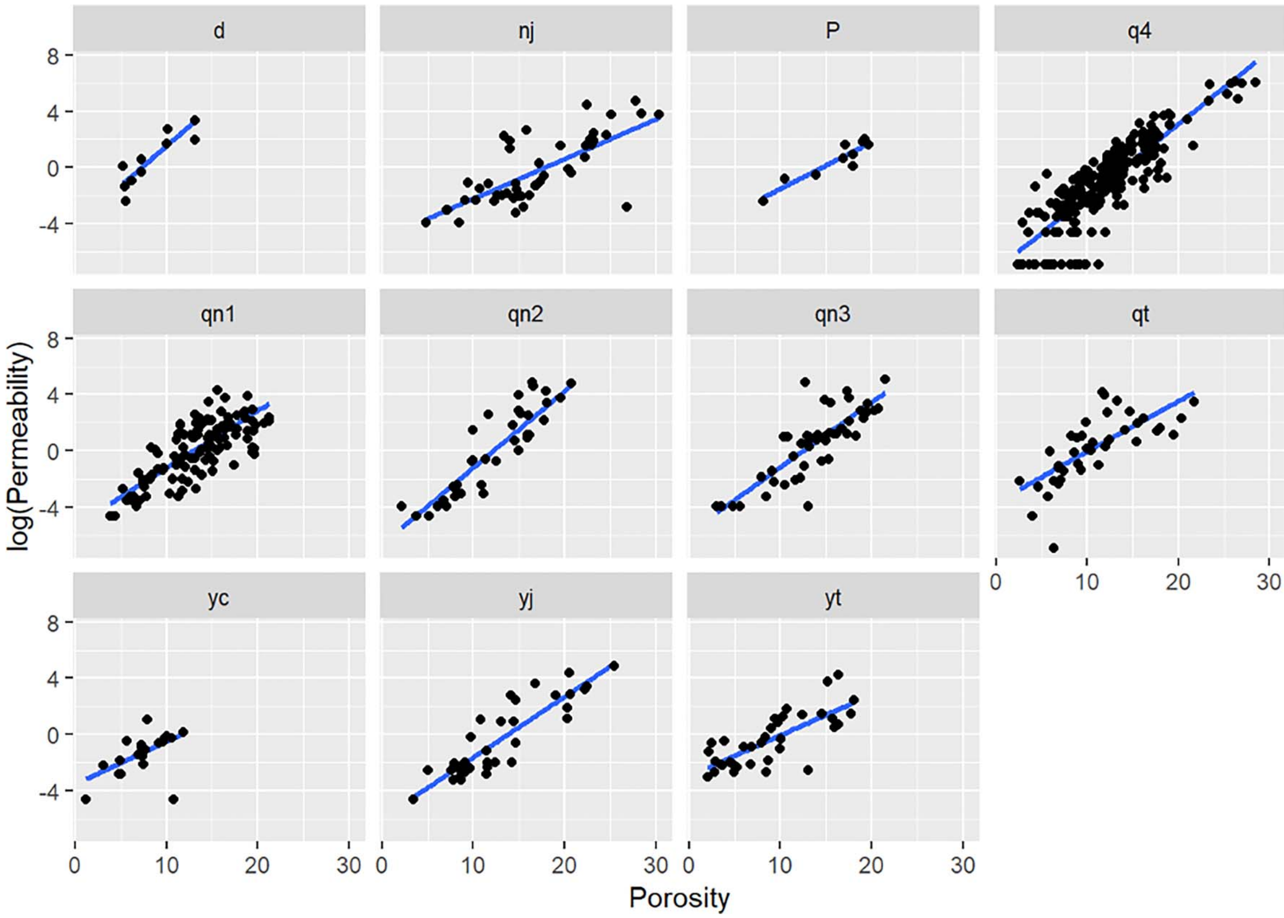

Fig. 3 Log(permeability) versus porosity by all the members

Table 1 Estimates, standard errors, degrees-of-freedom (DF), test statistics (*t*-value), and *p*-values for fixed effect coefficients for linear mixed model

	Estimate	Std. error	DF	*t*-value	*p*-value
Intercept	−0.0432758	0.14833405	781	−0.291746	0.7706
Porosity	0.7699809	0.07742211	781	9.945232	0.0000
Sorting coefficient	0.1225279	0.05514295	781	2.222005	0.0266
Mean radius	0.9763455	0.14121891	781	6.913703	0.0000
Homogeneous coefficient	−0.1390756	0.07052698	781	−1.971948	0.0490
Log (displacement pressure)	−0.7041038	0.16782522	781	−4.195459	0.0000

A value close to 1 is preferable for VIF, while a value greater than 10 is problematic. Our highest VIFs are around 5, indicating that there are some collinearities among the predictors, but they will not affect our results. A relatively high correlation among predictors also makes sense due to the physical laws of quantifying these petrophysical properties.

Based on our model and estimates, sensitivity analysis can be performed to identify the relative importance of each predictor affecting permeability. In terms of the mean of permeability, the dominating factors can be addressed through the absolute magnitude of the estimated coefficients, since our predictors are standardized. From Table 1, Mean_Pore_Radius, Porosity and Log(Displacement_Pressure) have relatively high magnitudes (>0.7) compared with the others (≈0.1). Thus, they are the dominating factors. In terms of the variance of permeability, we consider the semi-partial R^2, which compares the marginal interpretability of the permeability variations by each predictor, holding the others in the model. Figure 6 illustrates the semi-partial R^2 for the whole model and each predictor. The whole model has an R^2 above 0.8, indicating more than 80% of permeability variations are explained by the model. The values for individual predictors are not very large due to the correlations among predictors. Similar to the mean, Mean_Pore_ Radius, Porosity, and Log(Displacement_Pressure) also have higher relative importance (>0.1) than the other two predictors (<0.1).

4.3 General Linear Model. Hypothesis testing is performed for each pair of the primary oil-producing members. Corresponding

Table 2 Estimates of the coefficients of all predictors for four main members

	Intercept	Porosity	Sorting coefficient	Mean radius	Homogeneous coefficient	Log(displacement pressure)
nj	−0.0831	0.5835	0.1004	1.3342	−0.0785	−0.4573
q4	−1.0747	0.7724	−0.0653	0.3127	0.2374	−1.8347
qn1	−0.4569	0.8052	0.0281	0.8757	−0.0451	−1.0612
qt	0.3725	0.9569	0.2270	0.7231	−0.3096	−0.6223

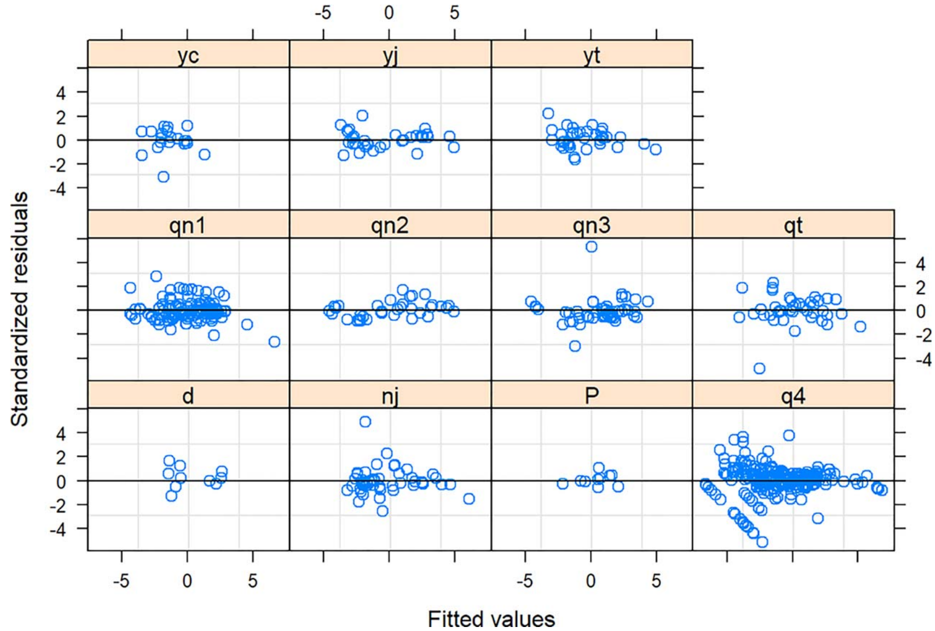

Fig. 4 Standardized residuals versus fitted values in each member

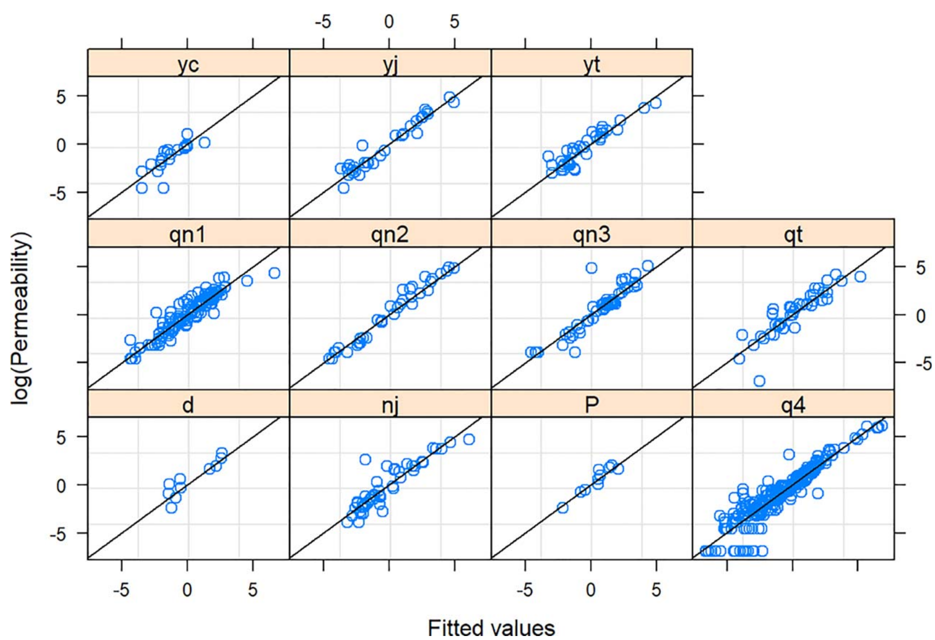

Fig. 5 Observed Log(permeability) versus fitted by each member

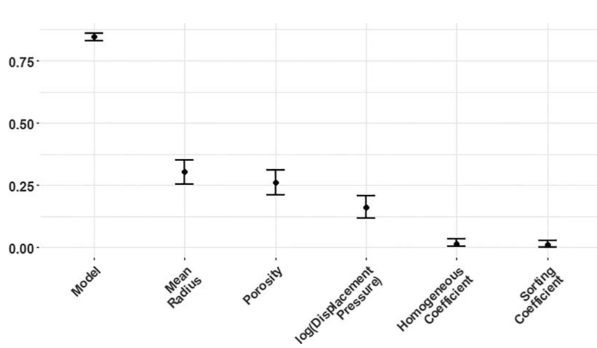

Fig. 6 Semi-partial R^2 for the model and each predictor

p-values for intercept and slopes are listed in Table 3. Bold numbers are *p*-values smaller than 0.05, indicating significance. The significant comparison means the two compared members are different in this specific predictor. From Table 3, the pairs of main members are almost significant in at least one predictor. Among the total six comparisons, the predictor Log(Displacement_Pressure) is substantial in four comparisons. Between the other two, one comparison, namely, nj versus qt, shows significance in Porosity and Mean_Radius. Results indicate that the overall pore structure of each member is different from the other members.

Considering the geological settings of the main oil-producing members, the relationship between specific geological characteristics and significant predictors can be identified. The predictor Porosity is primarily dominated by lithology, which controls the grain

Table 3 *p*-values for hypothesis testing in different members

	Intercept	Porosity	Sorting coefficient	Mean radius	Homogeneous coefficient	Log(displacement pressure)
nj versus qn1	0.8569	0.2419	**0.0422**	0.2266	0.4598	**0.0015**
nj versus q4	**0.0062**	0.8619	**0.0341**	**0.0002**	**0.0117**	0.0000
nj versus qt	**0.0017**	**0.0297**	0.6027	**0.0232**	0.8717	0.0918
qn1 versus q4	0.0000	0.3104	0.8061	**0.0002**	**0.0135**	0.0001
qn1 versus qt	0.0000	0.2023	0.1960	0.1267	0.4794	0.1863
q4 versus qt	0.0000	**0.0394**	0.1596	0.3008	**0.0502**	0.0000

Note: Bold numbers are *p*-values smaller than 0.05, indicating significance.

gradation in the member. The predictor Sorting_Coefficient is mostly related to the interstitial and diagenesis phase, which confirms that the interstitial has a significant impact on the permeability and other reservoir properties. The predictor Mean_Pore_Radius is mainly decided by grain composition and grain gradation, which implies that the lithology is the primary role in the mean pore radius. The predictor Homogeneous_Coefficient is governed by microfacies and the diagenesis phase. The final predictor Log(Displacement_Pressure) is influenced mostly by the cementation type and diagenesis phase. The result reveals that the diagenesis phase has a dominant role in three predictors. The identification of the diagenesis phase should be emphasized during the geological investigation and model fitting. Table 3 confirms the complicated derivation degree of the pore structure between different members, even in the same formation, which should be noticed in advanced geological reservoir simulation.

We notice that the group of qn1 versus qt is only significant in the intercept, which means they are of the statistical similarity in the predictors. We believe it is the post-diagenesis that causes this situation. The heterogeneity characteristics of post-diagenesis brought the isolated and discontinuous planar distribution of different members, such as member qn1 and member qt. Although we have the samples with the same depth and assume they are qualified, due to the different inspection wells location and various post-diagenesis in member qn1 and qt, the specific geological setting for each sample may be divergent to our assumption. Besides, the amounts of the samples should be increased to improve the model reliability. In the future, we will collect more samples to modify and check our model.

5 Conclusions

In this study, we proposed a hierarchical statistical model that both quantifies the relationship between permeability and other petrophysical properties and incorporates various geological characteristics using mercury intrusion data. To the best of our knowledge, no existing method builds this relationship with mercury intrusion data. By using the mixed model, we not only account for individual sample variations but also accommodate variations due to geological members. In addition, we identified that mean pore radius, porosity, and displacement pressure are relatively more important in explaining both the mean and variance of permeability, than sorting and homogeneous coefficients in this case study. Displacement pressure is also the most important variable to distinguish geological members. Geological settings behind each member are analyzed to explain the rationale of the comparison results and are also matched to petrophysical properties. This helps to fill the gap of the linear mixed model in predicting a new observation coming from a nonexisting member. The new observation can be assigned to a member with similar geological settings to its own. Though the model can always be improved with more samples, especially coming from new geological members, and more advanced techniques, our mixed model has the potential to provide guidance on mercury intrusion, as well as benefit the advanced geological simulation, model verification, and portable core sample 3D printing.

Acknowledgment

The authors thank the National Natural Science Foundation of China (No. 41972098) for their financial support to carry out this research. The insightful and constructive comments of the anonymous reviewers are also gratefully acknowledged.

Conflict of Interest

There are no conflicts of interest.

References

[1] Scheidt, C., Li, L., and Caers, J., 2018, *Quantifying Uncertainty in Subsurface Systems*, Vol. 236, John Wiley & Sons, New York.
[2] Zou, C., Yang, Z., Zhu, G., Tao, S., Zhu, R., Yuan, X., and Cai, B., 2019, "Establishment and Practice of Unconventional Oil and Gas Geology," Acta Geol. Sin., **93**(1), pp. 12–23. (In Chinese).
[3] Bryant, S., Cade, C., and Mellor, D., 1993, "Permeability Prediction From Geologic Models," AAPG Bull., **77**(8), pp. 1338–1350.
[4] Gelman, A., and Hill, J., 2006, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, New York.
[5] Fisher, R. A., 1919, "XV—The Correlation Between Relatives on the Supposition of Mendelian Inheritance," Earth Environ. Sci. Trans. R. Soc. Edinburgh, **52**(2), pp. 399–433.
[6] Van Vleck, L. D., 1998, "Charles Roy Henderson, 1911–1989: A Brief Biography," J. Anim. Sci., **76**(12), pp. 2959–2961.
[7] Stuart, A., Ord, K., and Arnold, S., 1994, *Kendall's Advanced Theory of Statistics, Classical Inference and the Linear Model*, 6th ed., Vol. 2A, Wiley, Chichester.
[8] R Core Team, 2018, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.
[9] Shiqian, X., Yuyao, L., Yu, Z., Sen, W., and Qihong, F., 2020, "A History Matching Framework to Characterize Fracture Network and Reservoir Properties in Tight Oil," ASME J. Energy Resour. Technol., **142**(4), p. 042902.
[10] Ramm, M., and Bjørlykke, K., 1994, "Porosity/Depth Trends in Reservoir Sandstones: Assessing the Quantitative Effects of Varying Pore-Pressure, Temperature History and Mineralogy, Norwegian Shelf Data," Clay Miner., **29**(4), pp. 475–490.
[11] Fan, Z., Zhang, Y., and Yang, D. T., 2016, "Estimation of Three-Phase Relative Permeabilities for a Water-Alternating-Gas Process by Use of an Improved Ensemble Randomized Maximum-Likelihood Algorithm," SPE Reservoir Eval. Eng., **19**(4), pp. 683–693.
[12] Wantawin, M., Yu, W., and Sepehrnoori, K., 2017, "An Iterative Work Flow for History Matching by Use of Design of Experiment, Response-Surface Methodology, and Markov Chain Monte Carlo Algorithm Applied to Tight Oil Reservoirs," SPE Reservoir Eval. Eng., **20**(3), pp. 613–626.
[13] Zhang, Y., He, J., Yang, C., Xie, J., Fitzmorris, R., and Wen, X. H., 2018, "A Physics-Based Data-Driven Model for History Matching, Prediction, and Characterization of Unconventional Reservoirs," SPE J., **23**(4), pp. 1105–1125.
[14] Zhang, Y., Fan, Z., Yang, D., Li, H., and Patil, S., 2017, "Simultaneous Estimation of Relative Permeability and Capillary Pressure for PUNQ-S3 Model With a Damped Iterative-Ensemble-Kalman-Filter Technique," SPE J., **22**(3), pp. 971–984.
[15] Blunt, M., King, M. J., and Scher, H., 1992, "Simulation and Theory of Two-Phase Flow in Porous Media," Phys. Rev. A, **46**(12), pp. 76–80.
[16] Lapierre, C., Leroueil, S., and Locat, J., 1990, "Mercury Intrusion and Permeability of Louiseville Clay," Canadian Geotech. J., **27**(6), pp. 761–773.
[17] Al Hinai, A., Rezaee, R., Saeedi, A., and Lenormand, R., 2013, "Permeability Prediction From Mercury Injection Capillary Pressure: An Example From the Perth Basin, Western Australia," APPEA J., **53**(1), pp. 31–36.
[18] Clarkson, C. R., Wood, J., Burgis, S., Aquino, S., and Freeman, M., 2012, "Nanopore-Structure Analysis and Permeability Predictions for a Tight Gas Siltstone Reservoir by Use of Low-Pressure Adsorption and Mercury-Intrusion Techniques," SPE Reservoir Eval. Eng., **15**(06), pp. 648–661.

[19] Nooruddin, H. A., Anifowose, F., and Abdulraheem, A., 2014, "Using Soft Computing Techniques to Predict Corrected Air Permeability Using Thomeer Parameters, Air Porosity and Grain Density," Comput. Geosci., **64**, pp. 72–80.

[20] Gao, H., and Li, H. A., 2016, "Pore Structure Characterization, Permeability Evaluation and Enhanced Gas Recovery Techniques of Tight Gas Sandstones," J. Nat. Gas Sci. Eng., **28**, pp. 536–547.

[21] Liu, G., Bai, Y., Gu, D., Lu, Y., and Yang, D., 2018, "Determination of Static and Dynamic Characteristics of Microscopic Pore-Throat Structure in a Tight Oil-Bearing Sandstone Formation," AAPG Bull., **102**(9), pp. 1867–1892.

[22] Liu, G., Yin, H., Lan, Y., Fei, S., and Yang, D., 2020, "Experimental Determination of Dynamic Pore-Throat Structure Characteristics in a Tight Gas Sandstone Formation With Consideration of Effective Stress," Marine Petroleum Geol., **113**, pp. 104–170.