Data-driven modeling to optimize the injection well placement for waterflooding in heterogeneous reservoirs applying artificial neural networks and reducing observation cost

Xinwei Xiong and Kyung Jae Lee

Abstract
Secondary recovery methods such as waterflooding are often applied to depleted reservoirs for enhancing oil and gas production. Given that a large number of discretized elements are required in the numerical simulations of heterogeneous reservoirs, it is not feasible to run multiple full-physics simulations. In this regard, we propose a data-driven modeling approach to efficiently predict the hydrocarbon production and greatly reduce the computational and observation cost in such problems. We predict the fluid productions as a function of heterogeneity and injection well placement by applying artificial neural network with small number of training dataset, which are obtained with full-physics simulation models. To improve the accuracy of predictions, we utilize well data at producer and injector to achieve economic and efficient prediction without requiring any geological information on reservoir. The suggested artificial neural network modeling approach only utilizing well data enables the efficient decision making with reduced computational and observation cost.

Keywords
Data-driven modeling, well placement optimization, waterflooding, heterogeneous reservoirs, artificial neural network, reducing observation cost
Introduction

Waterflooding is one of the most widely used secondary recovery methods to improve the oil and gas production. When reservoir pressure significantly decreases after the primary production, external energy is needed to drive the remaining oil to the production well. The basic idea of waterflooding is to inject water into the reservoirs, to increase the formation pressure and maintain it around the initial value. Thus, the performance of waterflooding is controlled by the sweep efficiency, which is closely related to the reservoir heterogeneity.

Prediction of waterflooding performance is complex and challenging in heterogeneous reservoirs, due to the uncertainty and variability. Uncertainty refers to the insufficiency of knowledge and information on geological properties; variability refers to the reservoir heterogeneity such as spatial differences of porosity and permeability in the reservoir (Siirila et al., 2012). Because of these challenges, numerical simulations of heterogeneous reservoirs usually require a large number of discretized grid blocks in simulation models. Thus, it is expensive and time-consuming to run full-physics simulations of heterogeneous reservoirs for every possible condition of operation.

To improve the efficiency of the simulation of heterogeneous reservoirs, machine learning methods have been introduced to significantly decrease the simulation time while reliably predicting the fluid production. Machine learning is a data analysis method that can automatically learn the relationships of input and output from the sampled training data (Michie et al., 1994). Statistical learning methods and artificial neural network (ANN) are widely used machine learning technologies in various scientific and engineering problems.

Statistical learning methods are based on relationship formulation between input and output datasets by applying mathematical, logical, and statistic methods (James et al., 2013). There are various statistical learning methods used to build proxy models of reservoir simulations, such as support vector regressions, which utilize hyperplanes to maximize the margin of tolerance (Smola and Schölkopf, 2004). On the other hand, ANN models build the computing systems that imitate the biological neural networks (Jain et al., 1996). These neural networks can efficiently combine the multiple algorithms and process the complicated input data. Each neuron contains multiple parameters relating input and output data, and the prediction results can be improved by optimizing these parameters. As neural networks become larger, the more complex functions can be processed (LeCun et al., 2015). If there is a nonlinearity between input and output datasets, and the sizes of input and output datasets are huge, ANN shows better performance than statistical learning methods (Meyfroidt et al., 2009).

By their remarkable applicability in datasets with nonlinearity and complexity, ANN models are actively applied in the fields of heterogeneous reservoir simulation in recent years. Gaganis and Varotsis (2012) developed a non-iterative method to determine the phase equilibrium of reservoir simulation using ANN methods. Here, molar fractions of \(C_1\), \(nC_4\), and \(nC_{10}\), pressure, and temperature were trained as input data, while equilibrium coefficients of components were applied as output data. This ANN model greatly reduced the simulation time of phase equilibrium calculations. Bansal et al. (2013) developed ANN models to predict the production performance in a discontinuous tight oil reservoir, where well log data, well location, and completion data were applied as input data; maximum production rate, average production rate, and decline curve parameters were used as output data. Guérillot and Bruyelle presented ANN models to forecast the hydrocarbon production considering uncertainty in oil production by utilizing geological properties such as
porosity and permeability of reservoir as input data, and reservoir simulation solutions such as cumulative oil production and oil production rate as output data, respectively. This ANN model could increase the accuracy of simulation results by minimizing the effect of uncertainty and reduce the simulation time as well (Guérillot and Bruyelle, 2017). BuKhamseen and Ertekin proposed ANN models to predict the properties of hydraulic fractures in reservoir simulations, where fracture half-length, effective permeability, and productivity index were utilized as input data, and production data as output data, respectively. This ANN model predicted the fracture half-length and effective permeability with a good accuracy (BuKhamseen and Ertekin, 2017). Nwachukwu et al. (2018) proposed a tree-based learning method to estimate the optimized injection well placement, by utilizing connectivity between injection and production wells as input data and net present value as output data.

As stated, most of the previous works on data-driven models as proxy models to solve the complex reservoir problems utilized the geological data of reservoirs as input parameters, such as connectivity between wells, porosity and permeability fields, seismic data, time of flight, and radius of investigation, which could cause significant monitoring cost by drilling new wells for data observation. In this research, data-driven models based on ANN are proposed by utilizing data obtained at production and injection wells only, to efficiently predict the productivity in heterogeneous reservoirs where waterflooding processes are applied.

For the efficient prediction of productivity in heterogeneous reservoirs using ANN models, production history as an input dataset obtained with relatively short observation time is favored. However, it is challenging to reliably predict the ultimate productivity with production history obtained with short observation time. In order to improve the prediction performance of ANN models still utilizing short observation time of input, we propose an inclusion of additional informative monitoring data obtained at production and injection wells.

In this regard, this paper is structured as follows. In the following Methodology section, a two-dimensional (2D) numerical simulation model with different injector positions, which has been built to generate training data, is introduced. In the numerical simulation and data-driven modeling approaches, homogeneous and heterogeneous reservoir models are presented, where ANN approaches are applied to predict the fluid productions. In the training of ANN data-driven models, optimization by Levenberg–Marquardt backpropagation algorithm is introduced. In the Results and discussion section, the results of numerical simulation of waterflooding and the performance of ANN models are presented. Here, the improvement of ANN prediction performance by utilizing informative input data is discussed, and two criteria to decide the optimized injector position are proposed. In the Conclusions, summarizing remarks are addressed to provide insights for relevant applications.

**Methodology**

**Numerical model for reservoir simulation**

In this study, reservoir simulation code of TOUGH2 was used to generate training data of ANN models (Pruess et al., 1999). TOUGH2 is designed for simulating multiphase and multicomponent fluid flow in porous and fractured media under nonisothermal conditions. In this study, black oil and water systems (two-phase and two-component conditions) were
simulated using Equation of State Module 8 in TOUGH2. As the results of TOUGH2 simulations, system behavior including productions of fluid phases was obtained, by solving the discretized equations of mass balance.

A nine-spot waterflooding 2D simulation model was created as presented in Figure 1. This model included $51 \times 51$ cells, where reservoir length and width were both 1020 m. Each injector was evaluated to find the best oil production performance, where only one injector and one producer were open for each case of simulation runs. Reservoir properties and initial conditions are presented in Table 1.

**Reservoir permeability**

In order to take account for the uncertainty caused by sparse information of permeability, multiple equi-probable realizations of permeability fields were considered. Using the permeability values at the nine wells as in Table 2, 100 different permeability fields for each injector case were created; and the total of 800 different cases were simulated ($=8$ injector cases $\times 100$ permeability fields). Sequential Gaussian simulation methods were utilized to create the different permeability fields, by using iTOUGH2 GSLIB (Finsterle and Kowalsky, 2007). Here, ordinary kriging method was applied to generate the modification

![Figure 1. Reservoir geometry of 2D model.](image)

| Reservoir properties                                      | Values                      |
|-----------------------------------------------------------|-----------------------------|
| Reservoir temperature                                     | 60°C                        |
| Bulk density of formation                                 | 2650 kg/m$^3$               |
| Grid size ($\Delta x$, $\Delta y$, $\Delta z$)            | 20 m                        |
| Porosity                                                  | 0.2                         |
| Water injection rate                                      | 0.5520 kg/s (=300 bbl/day)  |
| Initial pressure                                          | $2.07 \times 10^7$ Pa (=3000 psia) |
| Bottomhole pressure                                       | $2.0 \times 10^7$ Pa (=2900 psia) |
| Waterflooding and production period                        | 10 years                    |
coefficient of permeability, which were log-normally distributed. Permeabilities for every grid block can be calculated by the following equation (Jung et al., 2011)

$$k(x) = \tilde{k}(x) \times \zeta(x)$$  \hspace{1cm} (1)

where $\zeta$ is the modification coefficient; $\tilde{k}$ is the mean value of the permeability distribution.

Figure 2 shows the samples of generated permeability fields of heterogeneous reservoirs. For the comparison with the heterogeneous cases, 100 homogeneous fields were also considered. Only injectors 1 and 2 were considered in the simulations of homogeneous cases, by the principle of symmetry. The range of permeabilities in the homogeneous cases was from $10^{-14} \text{ m}^2$ to $10^{-11} \text{ m}^2$ ($=10 \text{ mD}$–$10 \text{ D}$), which were evenly divided into 100 intervals in log-scale.

**Artificial neural network**

In order to build proxy models to predict the oil and water productions of stated reservoir simulation problems, ANN models were developed. ANNs are composed with a number of neurons, where every neuron includes three parameters: weights ($w$), biases ($b$), and activation functions ($f$) (Demuth and Beale, 1993). These parameters can optimize the input signals received from the previous neural layer and send the optimized signals to the next neural layer as an output signal. Figure 3 shows the structure of ANN models used in this study (Lee, 2019). We used two-layer feed-forward network composed with one hidden layer and one output layer. The numbers of neurons in the hidden and output layers were chosen to avoid the overfitting problem, by applying optimized suggestions (Heaton, 2008).

The relationships of the three parameters in each neuron are described as follows

$$v_i = b_i + \sum^n_i w_i x_i$$ \hspace{1cm} (2)

$$y_i = f(v_i)$$ \hspace{1cm} (3)

| Well     | Permeability                           |
|----------|----------------------------------------|
| Injector 1 | $2.000 \times 10^{-11} \text{ m}^2$ ($=20 \text{ D}$) |
| Injector 2 | $5.023 \times 10^{-12} \text{ m}^2$ ($=5090 \text{ mD}$) |
| Injector 3 | $2.000 \times 10^{-12} \text{ m}^2$ ($=2 \text{ D}$) |
| Injector 4 | $3.169 \times 10^{-12} \text{ m}^2$ ($=321 \text{ mD}$) |
| Injector 5 | $1.262 \times 10^{-12} \text{ m}^2$ ($=1279 \text{ mD}$) |
| Injector 6 | $3.990 \times 10^{-13} \text{ m}^2$ ($=404 \text{ mD}$) |
| Injector 7 | $1.002 \times 10^{-12} \text{ m}^2$ ($=1016 \text{ mD}$) |
| Injector 8 | $7.962 \times 10^{-13} \text{ m}^2$ ($=807 \text{ mD}$) |
| Producer  | $6.324 \times 10^{-13} \text{ m}^2$ ($=641 \text{ mD}$) |
Figure 2. Permeability fields in heterogeneous reservoirs (log k(m²)). (a) Field 11, (b) field 21, (c) field 31, (d) field 41, (e) field 51, (f) field 61, (g) field 71, (h) field 81.
where \( b_i, w_i, \) and \( f \) are the bias at neuron \( i \), weight at neuron \( i \), and the activation function, respectively (Cybenko, 1989).

Activation functions need to be selected carefully, because the failure to choose a proper activation function may lead to improper optimization of input data. In our ANN models, sigmoid function and linear function were utilized in hidden layer and output layer, respectively, considering both stability and efficiency (Agostinelli et al., 2015). Sigmoid curve is the special type of logistic function given by the following equation (Demuth and Beale, 1993)

\[
f(x) = \frac{1}{1 + e^{-x}}
\]

In our ANN models, input datasets included the cumulative oil and water productions obtained from 0 to 1500 days, with the observation frequency of 1 day. Output datasets included the cumulative oil and water productions after 10 years of waterflooding and production. MATLAB Neural Network Toolbox\textsuperscript{TM} was used to build ANN models in this study (Demuth and Beale, 1993). Here, the datasets were divided into three different categories—training, validation, and test data. In the training stage, the neural network was adjusted with respect to the evolution of training error. In the validation stage, the performance and improvement of neural network were checked to determine when the training procedure needed to be stopped. In the test stage, there was no alteration in the neural networks, and only the network performance was measured. In this study, 70%, 15%, and 15% of data were used for training, validation, and test, respectively.

**Levenberg–Marquardt backpropagation algorithm**

In the ANN modeling with backpropagation for training the datasets, optimization algorithms need to be applied to compute optimal values of weights and biases of neurons. In this study, Levenberg–Marquardt backpropagation algorithm was applied to predict the output results. Levenberg–Marquardt backpropagation algorithm has been widely used to provide numerical solution of a non-linear function (Moré, 1978). Levenberg–Marquardt
backpropagation algorithm combines the Steepest descent algorithm and Gauss–Newton algorithm (Hagan and Menhaj, 1994).

Steepest descent algorithm provides stable prediction results, but takes long time to converge because of the small step sizes (Widrow and McCool, 1976). On the other hand, Gauss–Newton algorithm takes second-order derivatives of error functions, automatically finds the proper step size, and converges rapidly (Wedderburn, 1974). However, Gauss–Newton algorithm can be applied only when the error functions are approximated as quadratic functions. In this regard, Levenberg–Marquardt backpropagation algorithm has been developed to automatically analyze the error function and selects more suitable algorithm for ANN models. Here, Steepest descent algorithm is applied when complex curvatures exist; Gauss–Newton algorithm is utilized when the quadratic approximations of error functions are reasonable.

**Results and discussion**

*Production behavior of homogeneous and heterogeneous reservoirs*

Simulation results of homogeneous reservoir cases are provided in Figures 4 and 5. Cumulative oil and water productions increased with increasing average permeability. Injector 2 cases showed shorter water breakthrough times than injector 1 cases, because

![Figure 4](image-url). Production rates of oil and water in homogeneous reservoirs utilizing each injector. (a) Oil production rate in injector 1 cases, (b) oil production rate in injector 2 cases, (c) water production rate in injector 1 cases, (d) water production rate in injector 2 cases.
the distance between injector and producer was shorter in the injector 2 cases. Injector 1 cases showed higher cumulative oil production, by the higher sweep efficiency. From the production rates and cumulative productions of oil and water, both reservoir permeability and the distance between injector and producer were influential for production behavior. This implies that it will be critical to include the information on the reservoir permeability (or geological property of reservoir affecting fluid flow) and the injector responses as an input dataset for the reliable prediction by ANN models.

Simulation results of heterogeneous reservoir cases are provided in Figures 6 to 9. Irregular patterns of oil and water productions and water breakthrough were observed by the heterogeneity-induced complex connectivity between injector and producer. From the production behavior, it is expected that the inclusion of informative input dataset on the reservoir connectivity affecting sweep efficiency in ANN modeling will be more critical in heterogeneous reservoir cases than homogeneous reservoir cases.

**ANN prediction performance using only production history as an input dataset**

As mentioned in the Introduction, the objective of this study is to successfully predict the long-term performance of production by using the input data obtained at wells only during short monitoring time. In this regard, the observation time of 1500 days with

![Figure 5](image-url)  
**Figure 5.** Cumulative productions of oil and water in homogeneous reservoirs utilizing each injector. (a) Oil production rate in injector 1 cases, (b) oil production rate in injector 2 cases, (c) water production rate in injector 1 cases, (d) water production rate in injector 2 cases.
1 day-observation frequency was applied to predict the cumulative oil and water productions after 10 years of waterflooding processes.

ANN prediction results of the cumulative oil and water productions after waterflooding processes, obtained with only production history as an input dataset, are shown in Figures 10 (homogeneous reservoirs) and 11 (heterogeneous reservoirs). Injector 2 cases of homogeneous reservoirs and injector 1 cases of heterogeneous reservoirs are presented representatively. Both in homogeneous and heterogeneous reservoirs, ANN prediction performances were not satisfactory. It indicates that the inclusion of input datasets implying the information of influential factors on production behavior will be necessary to improve the ANN prediction performances.

**Addition of input datasets to improve ANN prediction performances**

In order to improve the prediction performance of ANN models by still utilizing the short observation time, highly informative data needed to be included as an input of ANN models. We selected the five additional data as follows.

1. Ratio of oil–water production rates

   \[ R = \frac{q_o}{q_w} \]  

   \( R \)
where $R$ is the ratio of oil–water production rates; $q_o$ is the oil production rate at any time during observation; $q_w$ is the water production rate at the same time to $q_o$ measurement.

2. Ratio of cumulative oil–water productions

$$R_C = \frac{Q_o}{Q_w}$$

(6)

where $R_c$ is the ratio of cumulative oil–water productions; $Q_o$ is the cumulative oil production at any time during observation; $Q_w$ is the cumulative water production at the same time to $Q_o$ measurement.

3. Injectivity

$$I = \frac{q_{inj}}{\Delta P}$$

(7)

where $I$ is the injectivity; $q_{inj}$ is the water injection rate; $\Delta P$ is the difference of injector pressure between any time during observation and the initial time.
4. Pressure transient data at injector

\[ \Delta P = P_{\text{inj},i} - P_{\text{inj},0} \]  
\[ \Delta P' = \frac{\partial P}{\partial (\ln(t))} = \frac{P_{\text{inj},i} - P_{\text{inj},i-1}}{\ln(t_i) - \ln(t_{i-1})} \]

where \( \Delta P \) is the pressure difference; \( \Delta P' \) is the pressure difference derivative; \( P_{\text{inj},i} \) is the injector pressure at time \( i \); \( P_{\text{inj},0} \) is the injector pressure at initial time; \( P_{\text{inj},i-1} \) is the injector pressure at time \( i - 1 \) (Lee et al., 2003).

Utilization of oil–water ratios as an additional input dataset was proposed, because it was the direct indicator of relative productivity of oil and water. Utilization of pressure transient data at injector as an additional input dataset was proposed, because it included the information of reservoir connectivity (Lee et al., 2003). Pressure transient data at injector gives the information of permeability around the injector, which significantly affects the sweep efficiency of waterflooding. By applying the dimensionless variables of radial distance from
Figure 9. Cumulative water productions in heterogeneous reservoirs utilizing each injector. (a) Injector 1 cases, (b) injector 3 cases, (c) injector 6 cases, (d) injector 8 cases.

Figure 10. Prediction performances of ANN models in homogeneous reservoirs with injector 2. (a) Cumulative oil productions at 10 years, (b) cumulative water productions at 10 years.
Figure 11. Prediction performances of ANN models in heterogeneous reservoirs with injector 1. (a) Cumulative oil productions at 10 years, (b) cumulative water productions at 10 years.

Figure 12. Improved prediction performances of ANN models in homogeneous reservoirs utilizing each injector, after applying additional input datasets of injector. (a) Cumulative oil productions at 10 years (injector 1 cases), (b) cumulative water productions at 10 years (injector 1 cases), (c) cumulative oil productions at 10 years (injector 2 cases), (d) cumulative water productions at 10 years (injector 2 cases).
wellbore, time, and pressure, as defined in following equations (10) to (12), we have a dimensionless pressure solution as shown in equation (13) (Lee et al., 2003)

\[ r_D = \frac{r}{r_w} \]  \hspace{1cm} (10)

\[ t_D = \frac{kt}{\mu C_r r_w^2} \]  \hspace{1cm} (11)

\[ P_D = \frac{2\pi kh}{q\mu} (P - P_i) \]  \hspace{1cm} (12)

\[ P_D(r_D, t_D) = \frac{1}{2} \ln \left( \frac{4t_D}{e^{r_D^2}} \right) \]  \hspace{1cm} (13)

**Figure 13.** Error histograms of ANN models in homogeneous reservoirs utilizing each injector, after applying additional input datasets of injector. (a) Cumulative oil productions at 10 years (injector 1 cases), (b) cumulative water productions at 10 years (injector 1 cases), (c) cumulative oil productions at 10 years (injector 2 cases), (d) cumulative water productions at 10 years (injector 2 cases).
where $\gamma$ is the Euler’s constant ($\approx 0.577216$); $r_P$, $t_D$, and $P_D$ are the dimensionless variables for radial distance from wellbore, time, and pressure, respectively; $r_w$ is the wellbore radius; $k$ is the reservoir permeability; $h$ is the reservoir thickness; $q$ is the injection rate; and $\mu$ is the fluid viscosity. By combining equations (9), (12), and (13), we can obtain the following relationship for reservoir permeability and pressure difference derivative, $\Delta P'$:

$$k \propto \frac{q}{4\pi h \Delta P'}$$

(14)

As can be seen from equation (14), the inclusion of pressure difference derivative data ($\Delta P'$) will contain the information of permeability around the injector. The improved prediction performances of ANN models applying these additional input datasets are provided in Figures 12 to 17. Most of the data points fell on the dashed line in the homogeneous reservoir cases, which indicated the good performance of ANN predictions. In the heterogeneous reservoir cases, the additional input datasets made a significant

![Figure 14](image)

Figure 14. Improved prediction performances of ANN models in heterogeneous reservoirs utilizing each injector, after applying additional input datasets of injector. (a) Cumulative oil productions at 10 years (injector 1 cases), (b) cumulative oil productions at 10 years (injector 3 cases), (c) cumulative oil productions at 10 years (injector 6 cases), (d) cumulative oil productions at 10 years (injector 8 cases).
contribution of improving prediction performances as well. In both problems, error histograms show that the ANN models were built well, with the most frequent errors concentrated around zero.

**Selection of optimized injection well placement**

In order to design the optimized waterflooding operations in heterogeneous reservoirs, injector well placement with high cumulative oil production and relatively low cumulative water production was considered. In our selection, two criteria were utilized—P90 and P50. P90 was the cumulative oil production which was higher than 90% of whole cases; P50 was the cumulative oil production which was higher than 50% of whole cases. Figure 18 indicates the P90 and P50 lines of cumulative productions at injector 1. The cases of other injectors applied the same method to select the P90 and P50 productions.

Tables 3 and 4 present the P90 and P50 results from ANN models and actual simulation results, respectively. To select the optimized injection well placement, we first consider the

![Figure 15. Error histograms of ANN models in heterogeneous reservoirs utilizing each injector, after applying additional input datasets of injector. (a) Cumulative oil productions at 10 years (injector 1 cases), (b) cumulative oil productions at 10 years (injector 3 cases), (c) cumulative oil productions at 10 years (injector 6 cases), (d) cumulative oil productions at 10 years (injector 8 cases).](image-url)
highest cumulative oil production. However, if multiple injection well placements show very close cumulative oil production results, we selected the case with the least cumulative water production as the optimized injection well placement. More specifically, we considered the cumulative water production as additional criteria for the selection of optimized injection well placement, if the differences of cumulative oil productions in different cases were smaller than $0.15 \times 10^7$ kg.

As can be seen from the tables, injector 6 and injector 3 were the optimized injector well placement from ANN models using P90 and P50 criteria, respectively. Since we aim at higher cumulative oil production and lower cumulative water production, injector 8 and injector 3 were the optimized injector well placement from actual simulation results using P90 and P50 criteria, respectively. By using the P90 criteria, the optimized injection well placement was different in the ANN models and actual simulation results, which was caused by the very close cumulative oil productions in the cases with injectors 6 and 8. By considering P50 criteria, the optimized injection well placement by the ANN models showed conformity with the actual simulation results.

Figure 16. Prediction performances of ANN models in heterogeneous reservoirs utilizing each injector, after applying additional input datasets of injector. (a) Cumulative water productions at 10 years (injector 1 cases), (b) cumulative water productions at 10 years (injector 3 cases), (c) cumulative water productions at 10 years (injector 6 cases), (d) cumulative water productions at 10 years (injector 8 cases).
Figure 17. Error histograms of ANN models in heterogeneous reservoirs utilizing each injector, after applying additional input datasets of injector. (a) Cumulative water productions at 10 years (injector 1 cases), (b) cumulative water productions at 10 years (injector 3 cases), (c) cumulative water productions at 10 years (injector 6 cases), (d) cumulative water productions at 10 years (injector 8 cases).

Figure 18. Selection of P90 line and P50 line at injector 1 using simulation results.
Conclusions

In this study, data-driven ANN models efficiently forecasting waterflooding production performance in heterogeneous reservoirs have been developed, which successfully proposed the optimized injector well placement in accordance with the actual simulation results. Application of additional input data measured at the injector, which implied geological information and sweep efficiency of reservoirs, made significant contributions to improve the performance of ANN predictions. This suggested data selection method for the ANN modeling enabled the efficient prediction of production performance with relatively short observation time. The optimized ANN modeling approach is expected to contribute to reduce the computational and monitoring cost, because the data only obtainable at already drilled wells, such as production history and injector pressure transient data, were utilized as an input of ANN models. This is a noticeable improvement from the relevant previous works, where the majority of them required the geological properties at undrilled positions as an input of data-driven models with significant cost for geological survey.

The envisioned work will include the application of suggested data-driven modeling approach to actual field data, where different strategy needs to be taken. In the actual field data application, there is no future production data, which would be used as output data in the training of ANN models. In order to overcome the absence of future production data, the production history data until current time can be divided into input data (early time period production among measured data) and output data (late time period production among measured data) for the training of ANN models, which will be used to predict the future production. As soon as new production data are obtained, ANN models will be updated by including newly measured data. As stated, even though the time periods of input and output data for ANN training are different, the scientific methodology of data-driven modeling is the same in the cases applying simulation-generated data and actual field data.

Table 3. P90 and P50 for each injector case using ANN results.

| Mass (kg)a | Inj 1   | Inj 2   | Inj 3   | Inj 4   | Inj 5   | Inj 6   | Inj 7   | Inj 8   |
|------------|---------|---------|---------|---------|---------|---------|---------|---------|
| P90 oil    | $6.62 \times 10^7$ | $5.67 \times 10^7$ | $6.62 \times 10^7$ | $5.55 \times 10^7$ | $5.54 \times 10^7$ | $6.91 \times 10^7$ | $5.33 \times 10^7$ | $6.24 \times 10^7$ |
| P90 water  | $5.45 \times 10^7$ | $7.75 \times 10^7$ | $6.90 \times 10^7$ | $8.10 \times 10^7$ | $9.42 \times 10^7$ | $7.60 \times 10^7$ | $9.77 \times 10^7$ | $8.51 \times 10^7$ |
| P50 oil    | $5.01 \times 10^8$ | $4.21 \times 10^8$ | $5.25 \times 10^7$ | $4.02 \times 10^7$ | $3.84 \times 10^7$ | $5.38 \times 10^7$ | $4.17 \times 10^7$ | $4.99 \times 10^7$ |
| P50 water  | $1.02 \times 10^8$ | $1.08 \times 10^8$ | $8.25 \times 10^7$ | $1.33 \times 10^8$ | $9.44 \times 10^7$ | $1.37 \times 10^8$ | $1.14 \times 10^8$ | $6.31 \times 10^7$ |

*a1 STB oil = 135.13 kg oil, 1 STB water = 158.98 kg water.

Table 4. P90 and P50 for each injector case using simulation results.

| Mass (kg)a | Inj 1   | Inj 2   | Inj 3   | Inj 4   | Inj 5   | Inj 6   | Inj 7   | Inj 8   |
|------------|---------|---------|---------|---------|---------|---------|---------|---------|
| P90 oil    | $6.55 \times 10^7$ | $5.50 \times 10^7$ | $6.54 \times 10^7$ | $5.53 \times 10^7$ | $5.63 \times 10^7$ | $6.58 \times 10^7$ | $5.35 \times 10^7$ | $6.71 \times 10^7$ |
| P90 water  | $7.03 \times 10^7$ | $9.15 \times 10^7$ | $8.59 \times 10^7$ | $1.04 \times 10^8$ | $9.34 \times 10^7$ | $1.00 \times 10^8$ | $9.77 \times 10^7$ | $8.74 \times 10^7$ |
| P50 oil    | $4.96 \times 10^8$ | $4.11 \times 10^8$ | $5.11 \times 10^7$ | $4.11 \times 10^7$ | $3.82 \times 10^7$ | $4.98 \times 10^7$ | $4.10 \times 10^7$ | $5.12 \times 10^7$ |
| P50 water  | $9.25 \times 10^8$ | $1.21 \times 10^9$ | $6.82 \times 10^7$ | $6.38 \times 10^7$ | $7.86 \times 10^7$ | $8.51 \times 10^7$ | $9.75 \times 10^7$ | $1.10 \times 10^8$ |

*a1 STB oil = 135.13 kg oil, 1 STB water = 158.98 kg water.
Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The author appreciates the financial support for this research from the National Research University Fund of University of Houston.

ORCID iD
Kyung Jae Lee https://orcid.org/0000-0002-5631-9809

References
Agostinelli F, Hoffman M, Sadowski P, et al. (2015) Learning activation functions to improve deep neural networks. arXiv preprint arXiv:1412.6830. Available at: https://arxiv.org/pdf/1412.6830.pdf
Bansal Y, Ertekin T, Karpyn Z, et al. (2013) Forecasting well performance in a discontinuous tight oil reservoir using artificial neural networks. In: SPE unconventional resources conference-USA. Woodlands, TX: Society of Petroleum Engineers.
BuKhamseen NY and Ertekin T (2017) Validating hydraulic fracturing properties in reservoir simulation using artificial neural networks. In: SPE Kingdom of Saudi Arabia annual technical symposium and exhibition. Saudi Arabia: Society of Petroleum Engineers.
Cybenko G (1989) Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals, and Systems 2(4): 303–314.
Demuth H and Beale M (1993) Neural Network Toolbox For Use with Matlab – User’s Guide. Version 3.0. Natick, MA: The MathWorks Inc.
Finsterle S and Kowalsky MB (2007) iTough2-GSLIB User’s Guide. Berkeley, CA: University of California.
Gaganis V and Varotsis N (2012) Machine learning methods to speed up compositional reservoir simulation. In: SPE Europec/EAGE annual conference. Denmark: Society of Petroleum Engineers.
Guérillot D and Bruyelle J (2017) Uncertainty assessment in production forecast with an optimal artificial neural network. SPE Middle East oil & gas show and conference. Kingdom of Bahrain: Society of Petroleum Engineers.
Hagan MT and Menhaj MB (1994) Training feedforward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks 5(6): 989–993.
Heaton J (2008) Introduction to Neural Networks with Java. Washington: Heaton Research, Inc.
Jain AK, Mao J and Mohiuddin KM (1996) Artificial neural networks: A tutorial. Computer 29(3): 31–44.
James G, Witten D, Hastie T, et al. (2013) An Introduction to Statistical Learning. New York: Springer.
Jung Y, Imhoff P and Finsterle S (2011) Estimation of landfill gas generation rate and gas permeability field of refuse using inverse modeling. Transport in Porous Media 90(1): 41–58.
LeCun Y, Bengio Y and Hinton G (2015) Deep learning. Nature 521(7553): 436–444.
Lee J, Rollins JB and Spivey JP (2003) Pressure Transient Testing. Richardson, TX: Society of Petroleum Engineers.
Lee KJ (2019) Characterization of type and maturity of organic matter in source rock by in-situ electrical heating and temperature transient analysis. In: SPE international conference on oilfield chemistry. Galveston, TX: Society of Petroleum Engineers.
Meyfroidt G, Güiza F, Ramon J, et al. (2009) Machine learning techniques to examine large patient databases. Best Practice & Research Clinical Anaesthesiology 23(1): 127–143.
Michie D, Spiegelhalter DJ and Taylor C (1994) Machine learning. *Neural and Statistical Classification* 13.

Moré JJ (1978) The Levenberg-Marquardt algorithm: Implementation and theory. In: Watson GA (ed.) *Numerical Analysis*. New York: Springer, pp.105–116.

Nwachukwu A, Jeong H, Pyrcz M, et al. (2018) Fast evaluation of well placements in heterogeneous reservoir models using machine learning. *Journal of Petroleum Science and Engineering* 163: 463–475.

Pruess K, Oldenburg CM and Moridis G (1999) *TOUGH2 User’s Guide*. Version 2. Berkeley CA: Lawrence Berkeley National.

Siirila ER, Navarre-Sitchler AK, Maxwell RM, et al. (2012) A quantitative methodology to assess the risks to human health from CO₂ leakage into groundwater. *Advances in Water Resources* 36: 146–164.

Smola AJ and Schölkopf B (2004) A tutorial on support vector regression. *Statistics and Computing* 14(3): 199–222.

Wedderburn RW (1974) Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* 61(3): 439–447.

Widrow B and McCool J (1976) A comparison of adaptive algorithms based on the methods of steepest descent and random search. *IEEE Transactions on Antennas and Propagation* 24(5): 615–637.

**Appendix**

**Notation**

- \( b_i \): bias value at neuron \( i \), dimensionless
- \( c_t \): total compressibility, Pa\(^{-1}\)
- \( h \): reservoir thickness, m
- \( I \): injectivity, kg/Pa·s
- \( k \): absolute permeability, m\(^2\)
- \( P \): reservoir pressure, Pa
- \( P_D \): dimensionless reservoir pressure, dimensionless
- \( P_{\text{inj},0} \): injector pressure at initial time, Pa
- \( P_{\text{inj},i-1} \): injector pressure at time \( i-1 \), Pa
- \( P_{\text{inj},i} \): injector pressure at time \( i \), Pa
- \( q_{\text{inj}} \): water injection rate, kg/s
- \( q_o \): oil production rate, kg/s
- \( q_w \): water production rate, kg/s
- \( Q_o \): cumulative oil production, kg
- \( Q_w \): cumulative water production, kg
- \( r_D \): dimensionless distance from wellbore, dimensionless
- \( r_w \): wellbore radius, m
- \( R \): ratio of oil-water production rates, dimensionless
- \( R_c \): ratio of cumulative oil-water productions, dimensionless
- \( t \): time, s
- \( t_D \): dimensionless time, dimensionless
- \( v_i \): transformed input for activation function, dimensionless
- \( w_i \): weight value at neuron \( i \), dimensionless
- \( x_i \): input at neuron \( i \)
- \( y_i \): output node of layer \( i \)
\( \Delta P \) pressure difference, Pa
\( \Delta P' \) pressure difference derivative, Pa/ln(s)
\( \phi \) reservoir porosity, dimensionless
\( \gamma \) Euler's constant, dimensionless
\( \mu \) reservoir fluid viscosity, Pa·s
\( \zeta \) modification coefficient of permeability, dimensionless