CrossMark

ORIGINAL ARTICLE

Characterizing interwell connectivity in waterflooded reservoirs using data-driven and reduced-physics models: a comparative study

Emre Artun[1]

Abstract Waterflooding is a significantly important process in the life of an oil field to sweep previously unrecovered oil between injection and production wells and maintain reservoir pressure at levels above the bubble-point pressure to prevent gas evolution from the oil phase. This is a critical reservoir management practice for optimum recovery from oil reservoirs. Optimizing water injection volumes and optimizing well locations are both critical reservoir engineering problems to address since water injection capacities may be limited depending on the geographic location and facility limits. Characterization of the reservoir connectivity between injection and production wells can greatly contribute to the optimization process. In this study, it is proposed to use computationally efficient methods to have a better understanding of reservoir flow dynamics in a waterflooding operation by characterizing the reservoir connectivity between injection and production wells. First, as an important class of artificial intelligence methods, artificial neural networks are used as a fully data-driven modeling approach. As an additional powerful method that draws analogy between source/sink terms in oil reservoirs and electrical conductors, capacitance–resistance models are also used as a reduced-physics-driven modeling approach. After understanding each method's applicability to characterize the interwell connectivity, a comparative study is carried out to determine strengths and weaknesses of each approach in terms of accuracy, data requirements, expertise requirements, training algorithm and processing times.

1 Introduction

Waterflooding is a secondary oil recovery method that is applied after the oil has been produced from a reservoir with its natural energy (known as the primary recovery period). Based on industry experience, it is known that the recovery factors (ratio of recovered oil volume to the original oil volume in the reservoir) with primary recovery ranges between 15–25 %, while a well-managed waterflood can increase this to 40–50 % depending on the reservoir characteristics and efficiency of the waterflooding operation. Although it can be classified as a common operation for secondary recovery, varying reservoir characteristics and limited water supplies in some areas make it critical to have a good understanding of the reservoir and determining the optimum design schemes of the waterflooding operation under consideration. Poorly managed waterflooding operations result in underperforming reservoirs with reserves that are left behind, and thus, huge losses in recovered oil and associated monetary income are realized. Understanding reservoir flow dynamics to apply proper reservoir management practices is a complex problem with well data from isolated localities in an oil field, data distributed spatially across the whole reservoir and spanned tens of years of history. Current reservoir management practices highly depend on numerical flow simulation models that take months to develop and maintain, and cost millions of dollars with both significant manpower and computational power requirements.

✉ Emre Artun
artun@metu.edu

[1] Petroleum and Natural Gas Engineering Program, Middle East Technical University, Northern Cyprus Campus, Mersin 10, Kalkanlı, 99738 Güzelyurt, Turkey

Although being recognized scientifically as the most powerful approach, full field-scale models do not easily allow rapid reservoir analysis that results in right reservoir management decisions.

Over the last two decades, petroleum industry has been transformed significantly with the advancement of intelligent field technologies that are mostly based on on-site instrumentation and automation of field operations. Many technologies have been adapted to collect significant volumes of data in much shorter time frames. However, the real challenge has been to convert these data into useful information to make quick and reliable decisions that generate value. This challenge can only be overcome by utilizing proper knowledge management, data assimilation and data analysis practices. The current paradigm in the evolution of science also requires advanced data analysis to synthesize all of the earlier empirical, experimental and computational findings [1].

An efficient way of managing an hydrocarbon reservoir at any stage of development is the closed-loop reservoir management approach [2]. As shown in Fig. 1, this approach requires continuous updating of models after collecting recent data from the high-resolution and high-frequency sensors in the oil field that record measurements of time-dependent (dynamic) properties such as pressure, flow rate and temperature. In this workflow, it is very important to have a model that can respond to the following primary needs:

- A model that can be updated quickly when new data are available.
- A model that is sufficiently accurate and representative of the actual system (surface or subsurface) so that it can be used for decision-making purposes.

There are a wide variety of modeling approaches presented in the reservoir engineering literature. Each modeling approach represents various complexities, advantages and
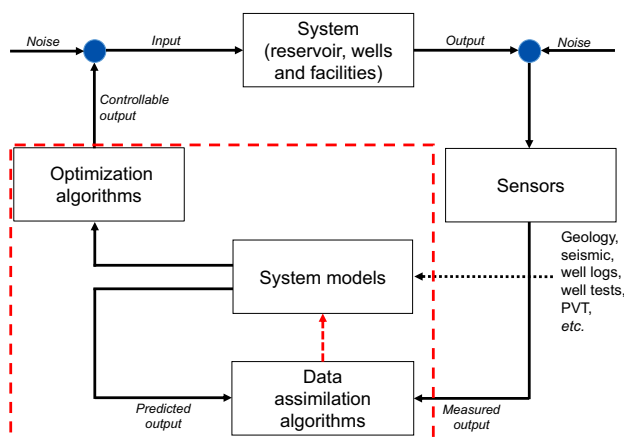

Fig. 1 Closed-loop reservoir management workflow [2]

disadvantages. In order to find a model that serves to both of the aforementioned objectives, it would be a better approach to have access to different modeling options readily available and choose the right modeling approach depending on the problem type and scope.

In this study, two conceptually different modeling approaches are investigated for the purpose of characterizing interwell connectivity in a waterflooded reservoir:

1. Fully data-driven modeling: artificial neural networks (ANN)—no functional relationship presumed.
2. Reduced-physics-driven modeling: capacitance–resistance model (CRM)—physics are incorporated with certain assumptions that simplify the problem.

Both methods have been proven to be potentially efficient tools for reservoir engineering problems based on studies presented in the literature. In this study, it is aimed to investigate the characteristics of each method including best practices and challenges associated with them to characterize interwell connectivity between injection and production wells. These would allow us to compare these methods with each other from the practical point of view and develop guidelines for the practicing engineer or asset team who is responsible for developing an optimum waterflooding plan. The primary objectives of the study can be summarized as the following:

1. Utilizing two different modeling approaches: artificial neural networks (as a data-driven modeling approach) and capacitance–resistance models (as a reduced-physics modeling approach) for quantifying interwell connectivity between injection and production wells in a waterflooded petroleum reservoir.
2. Assessing and comparing both methods (ANN and CRM) from different perspectives to determine strengths and weaknesses of each approach in terms of accuracy, data requirements, training algorithm, processing times and expertise requirements.

This study is the first attempt, to the best of our knowledge, to compare these two modeling approaches for the purpose of characterizing reservoir connectivity. This comparison provides with the necessary insight for the practicing engineer to implement either of these methods for a waterflooded petroleum reservoir. These tools have great advantages over other modeling approaches because of requiring fewer inputs, being much more computationally efficient and not being dependent on geological uncertainties. Therefore, having the necessary insights would help to decide which method is more practical for a particular problem. Based on the analysis performed during this study, the decision of choice would be affected by the expertise of the practicing engineer, availability of the data and the time frame of the study.

2 Methodology

2.1 Case study: a synthetic reservoir model

In this study, a synthetic streak case study [3] is selected to implement the aforementioned methods. It is a synthetic field that consists of 5 vertical injectors, I1 through I5, and 4 vertical producers, P1 through P4 (Fig. 2). The permeability of the reservoir is 5 md, except two high-permeability streaks:

1. Streak-1: 1000 md between I1 and P1 wells.
2. Streak-2: 500 md between I3 and P4 wells.

The porosity is constant and equal to 0.18. Total mobility of oil and water ($\lambda_o + \lambda_w$) is 0.45 and is independent of saturation. Oil, water and rock compressibilities are 5×10^{-6} psi^{-1}, 1×10^{-6} psi^{-1}, 1×10^{-6} psi^{-1}, respectively. The model is constructed with 1 layer and 31 grid blocks in each of the x and y directions with grid sizes of 80 ft $\times$ 80 ft (Δx and Δy). The thickness of the reservoir is 12 ft (Δz).

This model has been built and run using a commercial, numerical reservoir simulator [4] that utilizes black-oil formulation, which is a common formulation used as a robust approach for waterflooding problems. A variable water injection rate scenario is implemented, in which volumetric injection rates are varied significantly over a period of 100 months ($\approx$10 years). It is aimed to characterize the connectivity of the system through these rate fluctuations [3]. The bottom-hole pressure (BHP) for producing wells is fixed at 250 psia, and volumetric liquid production rates are measured (Fig. 3). The volumetric water injection rates are varied manually, while setting a
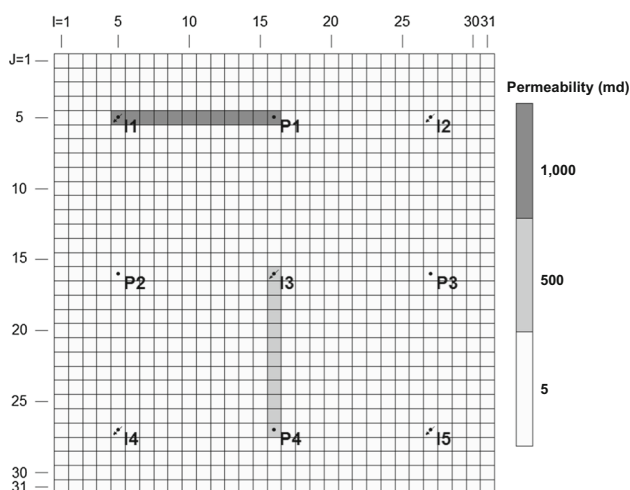
limit for maximum bottom-hole pressure of 5000 psia (Fig. 4). Both injection and production rates are measured at reservoir conditions (rbbl/day: reservoir barrels per day). One hundred months of monthly injection/production history resulted in a sample size of 100 for each well. Other descriptive statistics of volumetric injection rates for injector wells, I1 through I5, and volumetric production rates for producer wells, P1 through P4, are given in Table 1.

Although the presented synthetic case has rather simple permeability contrasts, it is a good example of a real reservoir with high-permeability streaks that must be considered during a waterflood optimization study. Existence of such streaks amplifies the importance of characterizing the connectivity between wells, since they provide a conduit in the reservoir to transport the injected water. The fact that results from a synthetic model are used should not raise any concern regarding the validity of the methods presented since both methods are proven to be successful in real reservoir cases with a number of examples in the literature. CRMs were successfully applied to a number of fields [5], and ANNs were successfully utilized for reservoir characterization problems with real-field data [6–9]. Since the main objective in this study is to perform a comparison of two methods, a synthetic case would be sufficient. However, it is anticipated that for a more complex reservoir case with more heterogeneities and more wells, more number of historical observations (more than 10 years of history) of injection and production rates might be needed for capturing the fluid flow dynamics of the reservoir.

2.2 Artificial neural networks

Intelligent systems have been applied to many different types of optimization problems in the petroleum industry. Most of these problems presented in the literature are based on development of ANN-based proxy models that can accurately mimic reservoir models within a reasonable amount of accuracy and computational efficiency. In some studies, these models are utilized to construct data-driven predictive tools and these tools are coupled with evolutionary algorithms to solve the optimization problem efficiently. Several areas of application included reservoir characterization [6–9], candidate well selection for hydraulic fracturing treatments [10], field development [11–13], well placement and trajectory optimization [14–16], scheduling of cyclic steam injection process [17], screening and optimization of secondary/enhanced oil recovery[18–23], history matching [24–26], underground gas storage management [27], reservoir monitoring and management [26, 28] and modeling of shale-gas reservoirs [29, 30].

Fig. 2 Synthetic reservoir model used in this study and its permeability distribution: a reservoir with 2 high-permeability streaks [3]

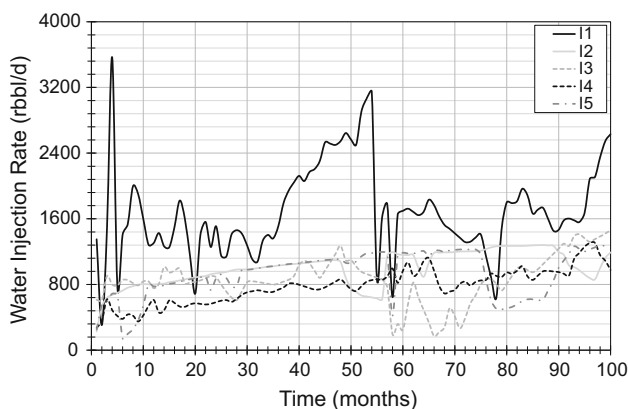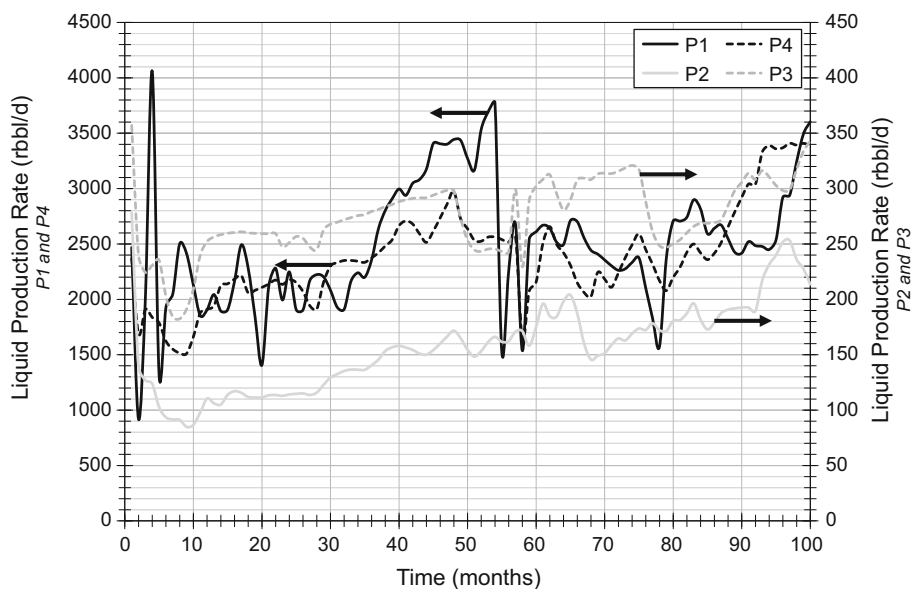Fig. 3 Production history of the synthetic reservoir model

Fig. 4 Injection history of the synthetic reservoir model

In addition to petroleum engineering and many other engineering disciplines, artificial neural networks and other data-driven modeling approaches have been used in many different kinds of applications such as spatial clustering [31], cavity-filter optimization [32], electricity load forecasting [33], control, pattern recognition, signal processing, medicine, speech recognition, speech production and business [34].

The most common training algorithm and also the one used in this study is the backpropagation algorithm. Also known as the *generalized delta rule*, backpropagation algorithm is a gradient-descent method that minimizes the total squared error of the output computed by the network. It played a major role in the re-emergence of neural networks in late 1980s. It was introduced as a training method of multilayer networks to overcome the limitations of single-layer networks [34]. Backpropagation algorithm is a supervised training technique (i.e., mapping a given set of

inputs to a specified set of target outputs) and includes three stages: (1) feedforward of the input training pattern, (2) calculation and backpropagation of the error and (3) adjustment of weights. The overall goal is to train the network such that it can [34]:

- Respond correctly to the input patterns that are used for training (memorization).
- Give reasonable responses to similar, but not identical, input patterns (generalization).

Figure 5 shows a multilayer, fully connected network with one hidden layer. There are n input neurons in the input layer, p hidden neurons in the hidden layer and m output neurons in the output layer. There are biases also shown in this figure whose activation value is constant during the training (1, in this case). While the number of inputs and outputs is based on the nature of the problem studied, the number of hidden neurons is a part of the network design process and must be optimized by the designer. A rule-of-thumb formula is presented to calculate the number of the hidden neurons, which is mostly based on experience [35]:

$$N_{\mathrm{HN}} = \frac{N_{\mathrm{I}} + N_{\mathrm{O}}}{2} + \sqrt{N_{\mathrm{TP}}} \qquad (1)$$

where N_{I} is the number of inputs, N_{O} is the number of outputs, and N_{TP} is the number of training patterns. It should be noted that this is not a theoretical formula, and this number would not necessarily be the best estimate of the number of hidden neurons. However, it can be used as a good start for the optimization process. Algorithm 1 shows a step-by-step explanation of the backpropagation algorithm.

Table 1 Descriptive statistics of volumetric injection and production rates that are output from the model and used in the study

Injection well	I1	I2	I3	I4	I5
Volumetric water injection rates (rbbl/day)					
Mean	1,683.1	999.5	860.4	779.0	931.8
Standard error	55.5	20.9	28.3	21.5	28.1
Median	1,595.2	1,018.6	854.7	792.0	1,020.1
Standard deviation	554.7	208.8	283.4	215.3	280.5
Sample variance	307,638.1	43,595.6	80,313.9	46,345.7	78,686.5
Kurtosis	1.24	0.60	0.37	−0.07	−0.05
Skewness	0.71	−0.74	−0.27	0.04	−0.86
Range	3,258.7	1,053.0	1,288.9	1,034.2	1,128.5
Minimum	305.9	227.0	172.7	272.0	145.0
Maximum	3,564.6	1,280.0	1,461.6	1,306.3	1,273.5
Sum	168,313.9	99,945.2	86,043.1	77,896.8	93,179.6
Count	100	100	100	100	100

Production well	P1	P2	P3	P4	
Volumetric liquid production rates (rbbl/day)					
Mean	2,494.9	157.1	274.3	2,386.1	
Standard error	56.8	4.0	3.3	43.8	
Median	2,458.5	158.1	272.3	2,354.8	
Standard deviation	568.4	39.7	33.2	438.4	
Sample variance	323,095.6	1,577.6	1,105.2	192,167.8	
Kurtosis	0.34	0.23	0.30	0.41	
Skewness	0.27	0.47	−0.33	0.52	
Range	3,141.9	193.4	174.5	1,899.1	
Minimum	917.5	85.0	182.5	1,509.9	
Maximum	4,059.4	278.4	357.0	3,408.9	
Sum	249,486.5	15,712.0	27,431.2	238,606.3	
Count	100	100	100	100	

Algorithm 1 Backpropagation Training Algorithm

---

initialize weights (small random values)
while stopping condition is **false do**
　　for each training pair **do**
　　　Feedforward stage:
　　　- each input unit receives signal, and broadcasts it to all units in the layer above (hidden units)
　　　- each hidden unit sums its weighted input signals, applies its activation function to calculate its output, and sends this signal to all units in the layer above (output layer)
　　　- each output unit sums its weighted input signals, and applies its activation function to calculate its output
　　　Error backpropagation stage:
　　　for both hidden units and output units **do**
　　　　compute: *error*, and *error information* term
　　　　compute: *weight correction* term
　　　end for
　　　Weight adjustment stage:
　　　update weights and biases
　　end for
　　test for stopping condition
end while

Fig. 5 Architecture of a
multilayer network
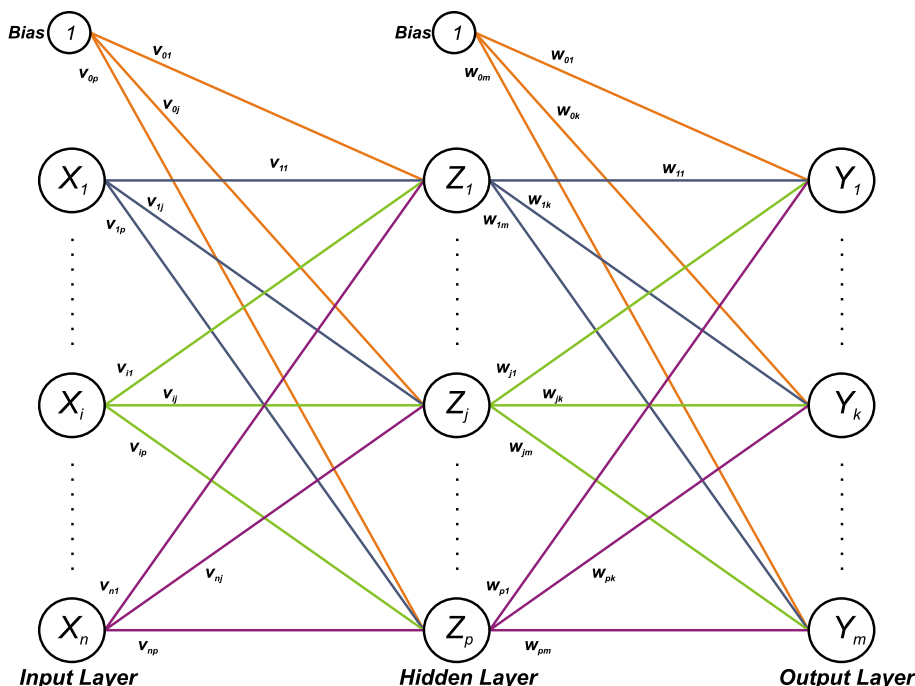
The iterative procedure for each training pair shown in Algorithm 1 is repeated for all training patterns, until a pre-specified stopping condition is achieved. Processing of each training data is known as a *training event* or *iteration*. When all training data are processed once, one *epoch* (training cycle) is completed. Each training data can be processed either by random selection, or by rotation. After each training event, average mean-squared error is calculated. Achieving minimum mean-squared error of outputs and maximum number of epochs is among most common stopping conditions. Once the defined stopping criteria are satisfied, weights on connection links achieve their optimum states. Provided that the training performance is satisfactory, the trained network with optimum weights can be used as a predictive model.

In this study, mapping input–output relationships is achieved with the inputs of injection rates from the water injectors and the output of the liquid production rate of a given producer. By analyzing the weights on connection links of the trained neural network, interwell connectivity is quantified. The optimized value of the weight on a given connection link indicates the degree of influence of the given input parameter on the output parameter. Therefore, we propose that once the training is completed, the relative values of connection links that connect each injector signal to the producer can be used to quantify the connectivity. There are individual neural networks for each producer well in the field. A neural network for a given producer would provide the connectivity of each injector to that producer. Once all neural network models are trained, all

connectivities between all injector–producer pairs would be quantified. This would provide insights about the waterflood dynamics in the reservoir and help to understand the overall reservoir connectivity to be used for further optimization studies.

A feedforward artificial neural network is constructed for each producing well in the reservoir. The training algorithm used is the Levenberg–Marquardt backpropagation algorithm [36, 37], and due to the low number of total input/output parameters (5 injectors and 1 producer: 6 parameters), only 1 hidden layer is used with 12 neurons. The schematic of the architecture of the neural network is shown in Fig. 6. Eighty percentage of the historical production/injection are used for training, 10 % are used for validation during the training to prevent over-training, and 10 % are used for blind-case testing.

2.3 Capacitance–resistance models

It was suggested that the development of an electrical model offers the promise of rapid evaluation for non-mathematical analysis of complex reservoir problems including understanding of the waterflood performance [3, 38]. An analogy between the flow behaviors of electricity in electric units and fluid in reservoir units was made through an experimental study [38]. This analogy implies that the electrical unit acts as a device which stores the electrical charge just as reservoir rock is acting as the storage of reservoir fluids [38]. After considering that the current may be equivalent to fluid flow, and the pressure is

Fig. 6 Schematic of the architecture of the neural network constructed with 5 inputs, 1 output and 12 hidden neurons
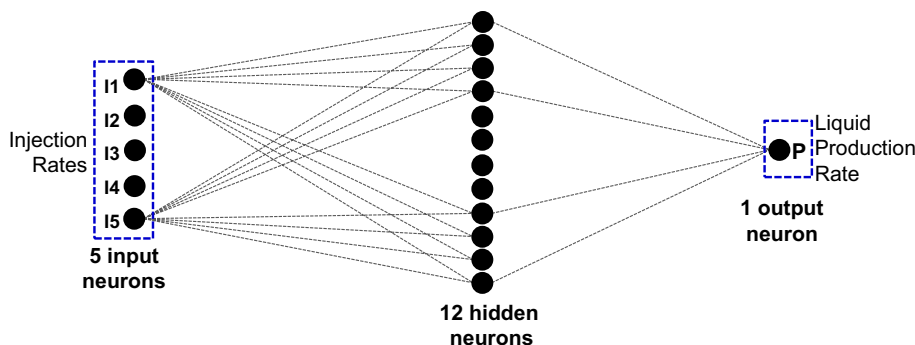
equivalent to the electrical potential, oil reservoir, as a porous continuum, is divided into small blocks so that the material balance can be used assuming the reservoir fluid is flowing in at one face of the block and out at the opposite face [38]. Role of such models for rapid estimation of waterflood performance and optimization was investigated by calling them capacitance–resistance models (CRMs) [3]. The base data necessary to run this model are production/injection data and well bottom-hole pressure (BHP) to calibrate the model against a specific reservoir. CRMs were primarily used for the characterization of interwell connectivity between injection and production wells rapidly without needing a geological model [39, 40]. After characterizing the connectivity, they are then used to optimize injection allocation and well locations in waterflooded reservoirs. An integrated capacitance–resistance model (ICRM) was presented that uses cumulative water injection and cumulative total production instead of water injection rate and total production rate while investigating the advantages of a linear reservoir model over the nonlinear capacitance–resistance model [41, 42].

The main advantage of CRM is that it requires very few inputs as little as the production/injection history. It is based on the main assumption that reservoir properties can be drawn only from production/injection history of wells. Also it requires that no significant changes in the field are observed during the analysis period. The primary application area is for the fields that are observing a secondary recovery period with water or gas injection. Its applications for primary and tertiary recovery periods are still being developed.

In these applications, the method enables to quantify connectivity between injector–producer pairs and aquifer strength, through history matching the production history by adjusting model parameters. After the capacitance models were introduced to understand interwell connectivity [43], CRMs for dynamic evaluation of waterfloods were presented [5]. Being a simple and user-friendly tool, the methodology proved to be very powerful in field applications, especially by quantifying interactions between injector and producer wells [44]. CRMs for three different control volumes are presented with semi-analytical formulations, with each of them having different level of complexities [3]:

1. One producer's control volume,
2. An injector–producer pair's control volume,
3. A field's control volume.

In this study, a producer-based control volume is considered to focus on production wells and how they are connected to different injection wells. Considering N_i number of injectors and N_p number of producers, an in situ volumetric balance over the effective pore volume of the producer is defined by the following differential equation [45]:

$$\frac{dq_j(t)}{dt} + \frac{1}{\tau_j} q_j(t) = \frac{1}{\tau_j} \sum_{k=1}^{N_i} f_{ij} i_i(t) - J_j \frac{dp_{wf,j}}{dt} \tag{2}$$

where τ_j is the time constant for producer j and defined as a function of total compressibility, c_t, pore volume, V_p, and productivity index, J, of the producer for its effective area:

$$\tau_j = \left(\frac{c_t V_p}{J} \right)_j \tag{3}$$

and, f_{ij} is defined as the fraction injection rate of injector, i, toward producer, j:

$$f_{ij} = \frac{q_{ij}(t)}{i_i(t)} \tag{4}$$

The solution of this equation, including a variation in the producer's bottom-hole pressure (BHP), is the following [43]:

$$q_j(t_n) = \overbrace{q_j(t_0) \left(e^{-\left(\frac{t_n - t_0}{\tau_j} \right)} \right)}^{\text{Primary depletion}}$$
$$+ \underbrace{e^{\left(\frac{-t}{\tau_j} \right)} \int_{t_o}^{t} e^{\left(\frac{t}{\tau_j} \right)} \frac{1}{\tau_j} \sum_{i=1}^{N_i} f_{ij} i_i(\xi) dt}_{\text{Injection input signal}} \tag{5}$$
$$- \overbrace{e^{\left(\frac{-t}{\tau_j} \right)} \int_{t_o}^{t} e^{\left(\frac{\xi}{\tau_j} \right)} J_j \frac{dp_{wf,j}}{d\xi} d\xi}^{\text{BHP variation}}$$

which includes three components for representing the production rate signal $q(t)$ at any given time on the right-hand side of the equation:

1. Primary depletion
2. Injection input signal
3. Variation in the producer's bottom-hole pressure (BHP)

By applying numerical integration, the integrals in the above solution were evaluated by proposing two approaches [45], which includes linear variation of BHP during the consecutive time intervals, and either stepwise variation in the injection rate (constant injection rate during a timestep), or linearly varying injection rate during a timestep. In this study, injection rates are kept constant during a timestep; therefore, the former solution is utilized. For the case of fixed injection rate of $i(\Delta t_k) = I_i^{(k)}$, and a linear BHP variation during time intervals Δt_k, $(k = 1, 2, \ldots, n)$, by assuming a constant productivity index at any given time, t_n, total production rate of producer j can be written as:

$$
\begin{aligned}
q_j(t_n) = q_j(t_0) \left(e^{-\left(\frac{t_n - t_0}{\tau_j}\right)} \right) + \sum_{k=1}^{n} \left\{ e^{-\left(\frac{t_n - t_k}{\tau_j}\right)} \left(1 - e^{-\left(\frac{\Delta t_k}{\tau_j}\right)} \right) \right. \\
\left. \left[\sum_{i=1}^{N_i} \left[f_{ij} I_i^{(k)} \right] - J_j \tau_j \frac{\Delta p_{wf,j}^{(k)}}{\Delta t_k} \right] \right\}
\end{aligned}
\tag{6}
$$

where $I_i^{(k)}$ and $\Delta p_{wf,j}^{(k)}$ represent injection rate of injector, i, and changes in the BHP of the producer, j, during time interval, t_{k-1} to t_k, respectively. The stepwise variation in injection rates is consistent with the discrete nature of field data that are typically reported in monthly averages [3]. If the bottom-hole pressure for producing wells does not change with time, the equation becomes:

$$
\begin{aligned}
q_j(t_n) = q_j(t_0) \left(e^{-\left(\frac{t_n - t_0}{\tau_j}\right)} \right) \\
+ \sum_{k=1}^{n} \left\{ e^{-\left(\frac{t_n - t_k}{\tau_j}\right)} \left(1 - e^{-\left(\frac{\Delta t_k}{\tau_j}\right)} \right) \left[\sum_{i=1}^{N_i} \left[f_{ij} I_i^{(k)} \right] \right] \right\}
\end{aligned}
\tag{7}
$$

The history-matching process is achieved by inputting observed production and injection rates for liquids and by changing the unknown parameters:

- Initial production rates, $q_j(t_0)$,
- Time constant for each producer, j, τ_j,
- Fraction injection rate of injector, i, toward producer, j, f_{ij} (i.e., the connectivity between injector, i, and producer, j.

Through an optimization routine, these parameters are changed until the average error between observed and calculated production rates is minimized. This error is defined as:

$$
\text{MSE} = \frac{\sum_{n=1}^{N_{\text{data}}} (q_{\text{obs}} - q_{\text{est}})^2}{N_{\text{data}}}
\tag{8}
$$

where N_{data} is the number of observations (sample size), q_{obs} is the observed flow rates, and q_{est} is the flow rate estimated by the model. The routine is initialized by assuming values for the time constant and initial flow rates and calculating the initial fractional flow parameter using the inverse-distance method [46]:

$$
f_{ij} = \frac{\frac{1}{d_{ij}}}{\sum_{j=1}^{N_{\text{pro}}} \frac{1}{d_{ij}}}
\tag{9}
$$

where d_{ij} represents the distance between each injector and producer. After this initialization, the *trust-region reflective* search algorithm [47–51] is used to search for the combination of parameters that provides the lowest range of error between the observed and estimated rates. After a pre-specified stopping criteria are met, then the solution is accepted as the optimum solution. An error tolerance of $1e-07$ and a maximum number of function evaluations of 10,000 are used as the convergence criteria. Then, the system parameters (e.g., fractional flow) are used to characterize the reservoir.

3 Results and discussion

Proposed methods, namely ANNs and CRMs, are applied to the case study presented in the previous section. The primary objective was to quantify the connectivity between injector/producer pairs using both methods. This is achieved by a contribution parameter derived from the trained neural network weights, w, in the case of ANNs and by the fractional flow parameter, f, in the case of CRMs. In the following subsections, results obtained using these two methods are presented and discussed.

3.1 Artificial neural networks

The history-matching results are shown in Fig. 7. These figures show that the training was successful in matching the historical rates observed. Once the training is completed, it is expected that the neural network would capture the dynamics of the reservoir system from observed data. Since no presumed physical laws are introduced, we call such models data-driven models. The model, during the training, captures the physics of the process through the neural network training, which is an iterative procedure. After the training is completed by satisfying certain stopping criteria, weights remain in their optimum state, at which the neural network can predict the output (production rate) within high levels of accuracy. In that case, the optimum set of weights would

represent the contribution of each injection well to the producing well's production, which can be used as a proxy to the connectivity between injection and production wells. Higher quantity of weights indicate stronger contribution and thus stronger connectivity, and lower quantity weights indicate weaker contributions and weaker connectivity. Contribution factors calculated in this case are given in Table 2. After calculating these contribution factors, a *connectivity map* is drawn which represents the strength of connection between each producing and injector well (Fig. 8). This figure is very similar to the actual reservoir grid in which the permeabilities are shown (Fig. 2). Since the neural network was able to

capture the high-permeability streaks in the reservoir system, this gives us the confidence that artificial neural networks can be used to characterize the connectivity of an oil reservoir system which goes through water injection.

3.2 Capacitance–resistance models

The history-matching results are shown in Fig. 9. These figures show that the training was successful in matching the historical rates observed. Based on the CRM formulation, the search process includes searching for the optimum combination of initial production rates, fraction of flow and
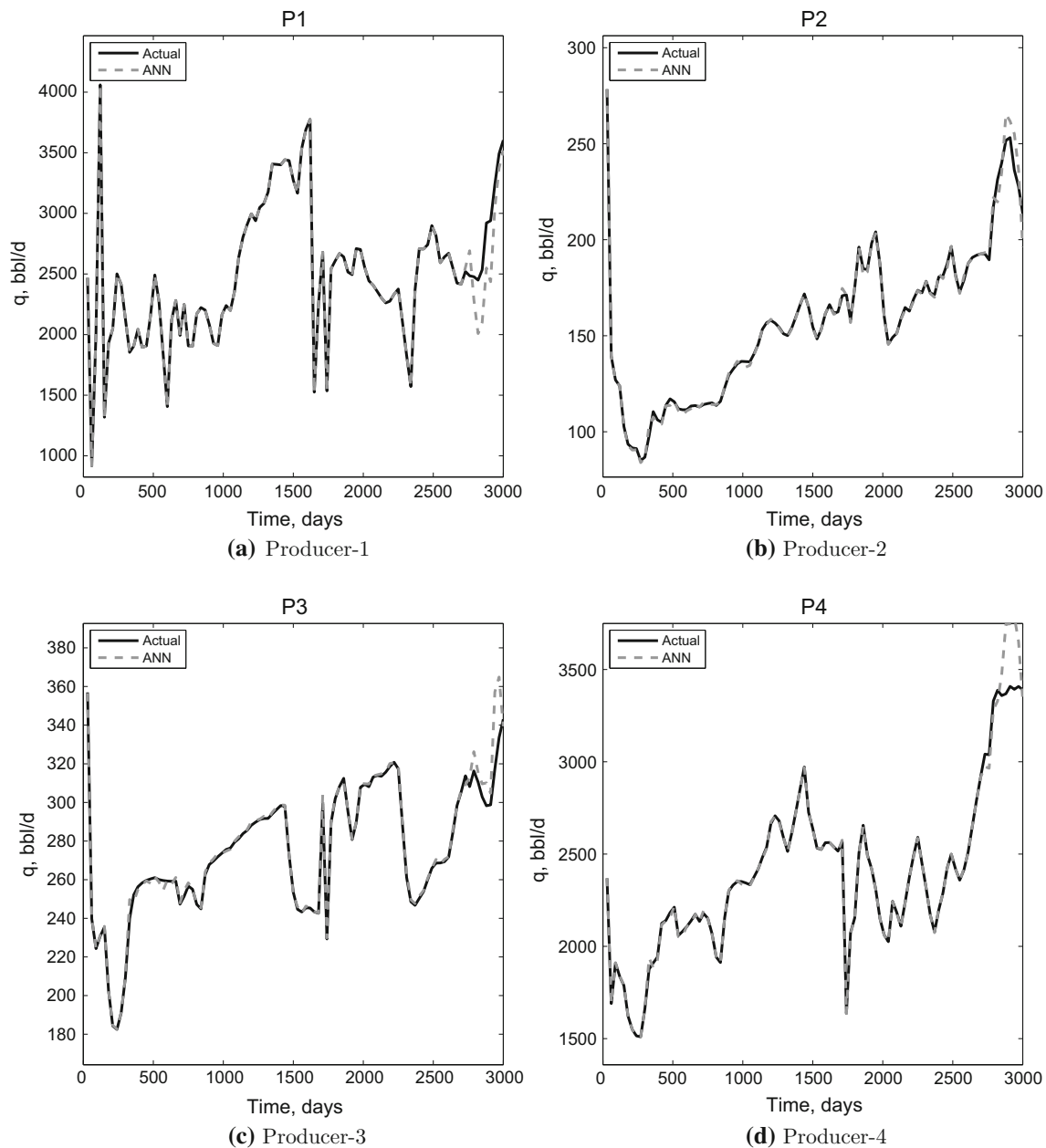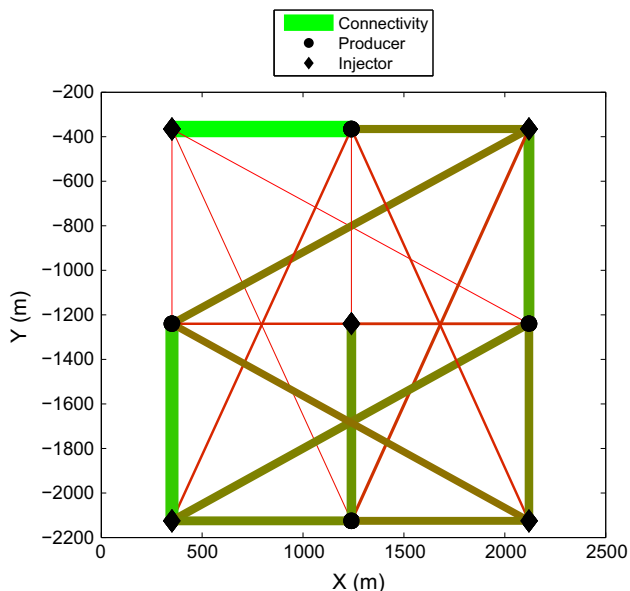

(a) Producer-1

(b) Producer-2

(c) Producer-3

(d) Producer-4

Fig. 7 History matching of the producing wells during ANN training

Table 2 Contribution values obtained from weights, w, after ANN training which is a proxy to the interwell connectivity

f	P1	P2	P3	P4
I1	1.000	0.250	0.293	0.204
I2	0.117	0.177	0.521	0.165
I3	0.206	0.205	0.184	0.561
I4	0.239	0.903	0.373	0.438
I5	0.213	0.239	0.402	0.405

Fig. 8 Connectivity map of injectors/producers using the contribution values derived from the trained neural network

time constant parameters. The fraction of flow parameter is used as a proxy to the interwell connectivity between each injector and producer. The values obtained after the optimization routine is completed are given in Table 3. Using these values, similar to the ANN case, a *connectivity map* is drawn which represents the strength of connection between each producing and injector well (Fig. 10). As in the case with ANN, CRM was able to capture both high-permeability streaks in the reservoir model which are shown in the reservoir grid (Fig. 2). Therefore, we also conclude that in a similar fashion with artificial neural networks, capacitance–resistance models can also be used to characterize the connectivity of an oil reservoir system which goes through water injection.

3.3 Comparison of data-driven and reduced-physics modeling approaches

One of the main objectives of this study is to provide a comparison of these two methods for the practicing

reservoir engineer or asset team with respect to a number of aspects. Table 4 shows a comparison of these two methods in different aspects, and a discussion of each aspect is presented as the following:

Accuracy Prediction capabilities of each method can be analyzed by comparing each method's ability to identify high-connectivity zones in the reservoir. When all of the 20 interwell connectivities between each of the five injectors and four producers are ranked, upper 10 values can be classified in the high-connectivity category, while lower 10 values can be classified in the low-connectivity category. One can approximate the connectivity values for the numerical simulation model, by utilizing the average permeability between two wells and normalizing the permeability by the distance between two wells (if two wells are close to each other and have a high-permeability streak between them, their connectivity would be the highest). These values are given in Table 5. After sorting the connectivity values in the numerical model together with the predictions of data-driven and the reduced-physics models, it is seen that the data-driven and the reduced-physics models were able to correctly estimate 80 and 70 % of each connectivity category, respectively. These acceptable accuracy levels indicate that both methods have similar prediction capabilities, while the data-driven model has a slightly better performance.

Data requirements Data-driven modeling approach (artificial neural networks) is better since it does not have any limitation regarding the types of input/output data set used. In this study, only injection and production rates are used. This can be considered as a minimum required set of data, since a set of signals would be needed to be able to relate the connectivity between wells, and production/injection rates are the most commonly available data set that can be incorporated. Meanwhile, the data included can be expanded, by including well locations, known reservoir properties and/or pressure data. Artificial neural networks are very advantageous in terms of flexibility and can be modified or restructured depending on the reservoir, wells, or any other known aspects of the field that are studied. On the other hand, capacitance–resistance models have a certain formulation and require the data in that formulation (production/injection rates, bottom-hole flowing pressures, well locations). These are common type of data available in any oil field; therefore, this does not create a significant problem in implementing the method. However, it does not have the flexibility to incorporate more data, if needed. In that case, the formulation needs to be modified and redeveloped, which is not practical to do for a quick application of the method but can be done in
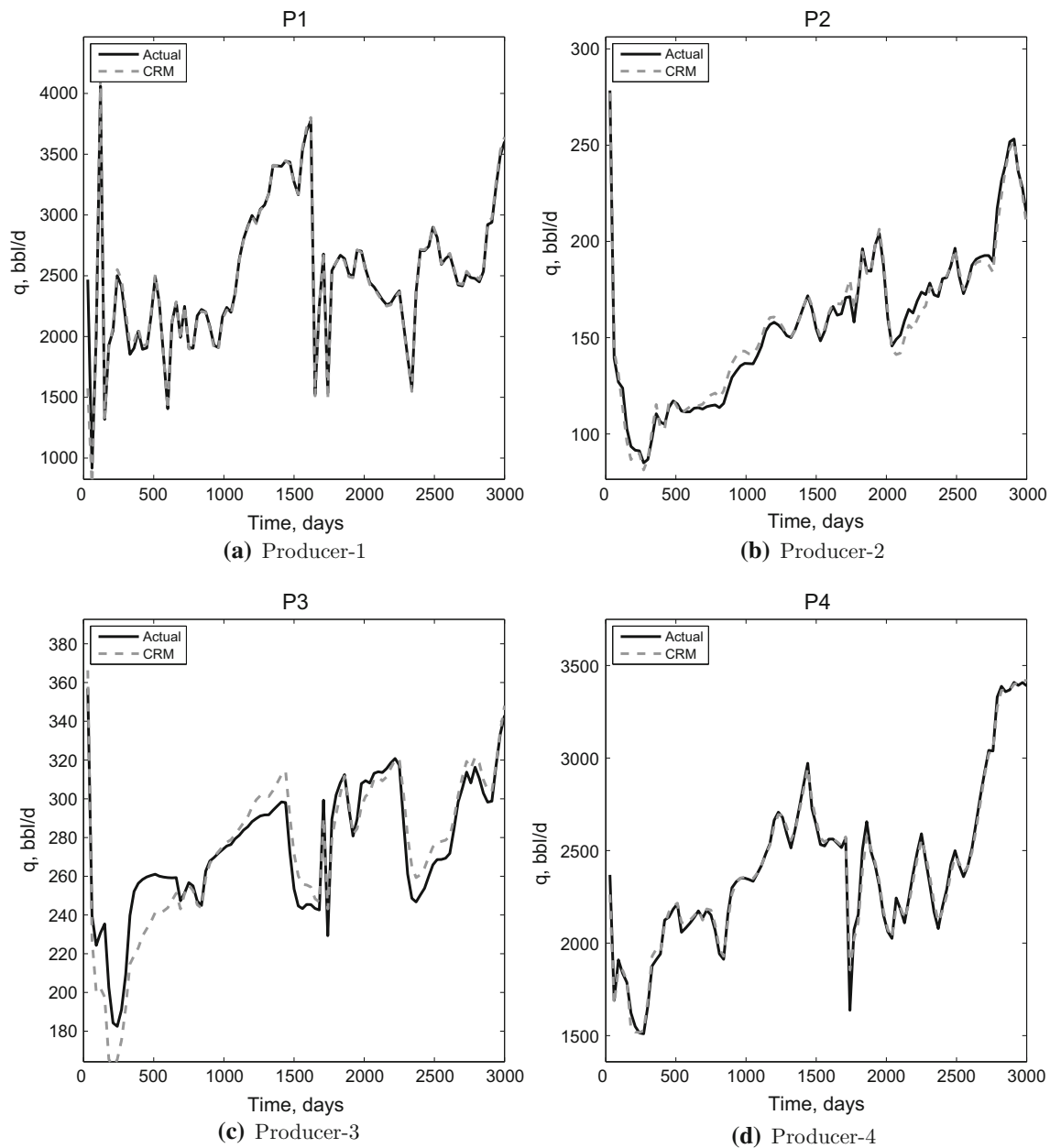
Fig. 9 History matching of the producing wells during CRM training

Table 3 Fraction of flow, f, values obtained after CRM training which is a proxy to the interwell connectivity

f	P1	P2	P3	P4
I1	1.000	0.002	0.003	0.005
I2	0.548	0.003	0.137	0.303
I3	0.057	0.010	0.033	0.974
I4	0.117	0.178	0.001	0.662
I5	0.126	0.001	0.106	0.766

the long term to have a suite of CRMs. This also can be explained with the modeling approach utilized in each method. ANNs are purely data-driven, and CRMs are reduced-physics-driven as explained earlier.

Training algorithm ANNs offer a number of different training algorithms, and all of them are purely data-driven algorithms. The choice of the algorithm requires subject-matter expertise in ANNs. ANNs do not limit the use of any training algorithm and provide the flexibility in choosing from the available options. With CRMs, the
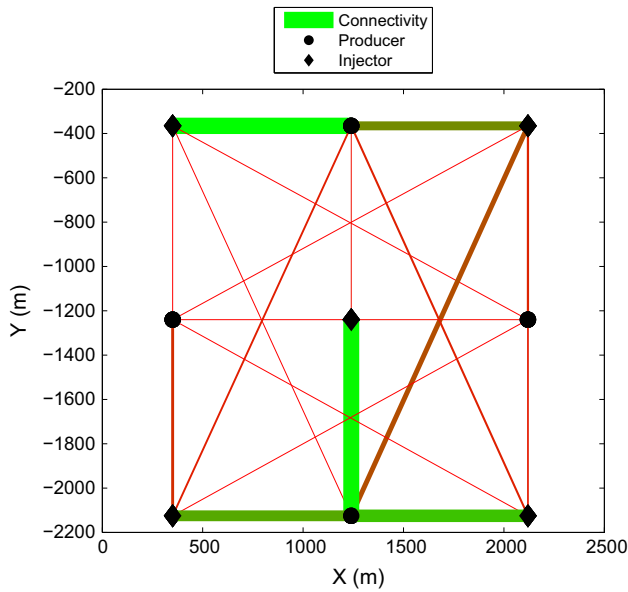
Fig. 10 Connectivity map of injectors/producers using the contribution values derived from the trained capacitance–resistance model

training is basically an optimization process, in which the optimization algorithm and parameters can be modified. Therefore, CRMs are also can be trained with a different number of optimization-related options.

Training speed Both methods are promising as the training of both models took <30 s of CPU time. Even for more complex reservoir systems, it is fair to expect training times not more than a few minutes, which is a reasonable amount of time for decision-making purposes.

Expertise requirements Both methods are considered to be moderate in terms of expertise requirements. In developing a tool that utilizes either of the method, both methods require significant expertise in certain subjects. ANNs require to be familiar with the ANN theory and terms; CRMs require the knowledge of CRM formulation and reservoir engineering concepts. To train a readily available tool using available data, ANNs still require to be familiar with the related theory to

determine the neural network architecture and training parameters. For training of CRMs, although not necessary, knowledge of optimization algorithms and parameters would help to find the best optimization approach.

4 Summary and conclusions

In this study, two methods of different modeling approaches, ANNs and CRMs, are studied to quantify the interwell connectivity between water-injecting wells and oil-producing wells for a petroleum reservoir that has gone through a reasonably long period of waterflooding. The methods were tested on a synthetic reservoir case, in which there are two high-permeability streaks (500 and 1000 md) in a reservoir having a permeability of 5 md, elsewhere. After calculating the connectivities from the model parameters, two methods are compared with each other considering various aspects from the practical points of view.

Only liquid production and injection rates as well as well locations are used as the data input to the two methods. Among these methods, artificial neural networks are purely data-driven, with no assumption regarding the governing laws of physics made. The other method, capacitance–resistance models, can be defined as reduced-physics-driven. The reason is that the physical laws included in the formulation require a large number of assumptions regarding the fluid flow. Both methods require a training process, in which the developed model *learns* from observed data, to capture the actual dynamics in the reservoir. Therefore, success of both methods depends on data quality and quantity. It was observed that for a known reservoir scenario with 2 high-permeability streaks, both methods were able to capture these streaks within reasonable ranges of accuracy. Since both methods are practical and easy to implement, a recommendation can be made that both methods are applicable for such applications.

The key conclusions obtained from this study can be summarized as follows:

Table 4 Characteristics of data-driven and reduced-physics modeling approaches in various aspects

	Data-driven modeling (ANN)	Reduced-physics modeling (CRM)
Accuracy (based on the specific example presented in this study)	80 %	70 %
Data requirements	Flexible	Fixed
Modeling approach	Flexible data-driven model	Physics-based, include assumptions
Training algorithm	Flexible	Flexible
Training speed	Fast	Fast
Expertise requirements	Moderate	Moderate
To develop the tool	ANN background needed	Reservoir eng. background needed
To train	ANN background needed	Optimization knowledge (for fine-tuning)

Table 5 List of connectivity values from the numerical, data-driven and reduced-physics models

Well pair			Connectivity		
No.	Injector	Producer	Numerical	Data-driven	Reduced-physics
1	I1	P1	1.000	1.000	1.000
2	I2	P1	0.233	0.548	0.117
3	I3	P1	0.169	0.057	0.206
4	I4	P1	0.104	0.117	0.239
5	I5	P1	0.104	0.126	0.213
6	I1	P2	0.233	0.002	0.250
7	I2	P2	0.104	0.003	0.177
8	I3	P2	0.169	0.010	0.205
9	I4	P2	0.233	0.178	0.903
10	I5	P2	0.104	0.001	0.239
11	I1	P3	0.104	0.003	0.293
12	I2	P3	0.233	0.137	0.521
13	I3	P3	0.169	0.033	0.184
14	I4	P3	0.104	0.001	0.373
15	I5	P3	0.233	0.106	0.402
16	I1	P4	0.104	0.005	0.204
17	I2	P4	0.104	0.303	0.165
18	I3	P4	0.651	0.974	0.561
19	I4	P4	0.233	0.662	0.438
20	I5	P4	0.233	0.766	0.405

1. Both ANNs and CRMs can be used to quickly estimate the interwell connectivity between injection wells and production wells in a reservoir. These tools have great advantages over numerical modeling because of requiring fewer inputs and being much more computationally efficient, while also providing the ability to utilize available historical data and not being dependent on geological uncertainties.

2. In the example presented here, 10 years of production/injection history was sufficient for training and achieving an accurate history matching. For more complicated reservoir cases (with more number of wells), higher duration of production and injection histories might be needed.

3. Both methods are efficient in terms of CPU time requirements with training times <30 s reported for the example used in this study. Even for more complex reservoir systems, this time is not expected to be greater than a few minutes.

4. In terms of data requirements and modeling approach, ANNs are more flexible than CRMs, since ANNs are purely data-driven and do not require any presumed functional relationship between process variables. Instead, it derives the relationships through training of observed data.

5. In both methods, some degree of knowledge is needed for fine-tuning of results during the training process which involves optimization of weights or the fractional flow parameter.

Compliance with ethical standards

Conflict of interest We declare that there is no any conflict of interest to report.

References

1. Hey T, Tansley S, Tole K (2009) The fourth paradigm: data-intensive scientific discovery. Microsoft Research, Redmond
2. Jansen J, Brouwer R, Douma S (2009) Closed loop reservoir management. In: SPE reservoir simulation symposium proceedings. SPE 119098. 2–4 February, The Woodlands, Texas
3. Sayarpour M, Zuluaga E, Kabir C, Lake L (2007) The use of capacitance-resistive models for rapid estimation of waterflood performance and optimization. In: SPE annual technical conference and exhibition proceedings. SPE 110081. 11–14 November. Anaheim, California
4. CMG (2013) CMG IMEX reservoir simulation software, version 2013. Computer Modeling Group, Ltd, Calgary
5. Sayarpour M, Kabir C, Lake L (2009) Field applications of capacitance–resistance models in waterfloods. SPE Res Eval Eng 12(6):853–864

6. Artun E, Mohaghegh S (2011) Intelligent seismic inversion workflow for high-resolution reservoir characterization. Comput Geosci 37(2):143–157

7. Raeesi M, Moradzadeh A, Ardejani FD, Rahimi M (2012) Classification and identification of hydrocarbon reservoir lithofacies and their heterogeneity using seismic attributes, logs data and artificial neural networks. J Pet Sci Eng 82–83:151–165

8. Das V, Chatterjee R (2012) Prediction of coal-bed permeability using artificial neural network. In: Proceedings of 9th biennial international conference and exposition on petroleum geophysics, Paper ID: 108, 16–18 February, Hyderabad, India

9. Alizadeh B, Najjari S, Ali KI (2012) Artificial neural network modelling and cluster analysis for organic facies and burial history estimation using well log data: a case study of the South Pars Gas Field, Persian Gulf, Iran. Comput Geosci 45:261–269

10. Mohaghegh S, Balan B, McVey D, Ameri S (1996) A hybrid neuro-genetic approach to hydraulic fracture treatment design and optimization. In: SPE annual technical conference and exhibition proceedings. SPE 36602. 6–9 October. Denver, Colorado

11. Centilmen A, Ertekin T, Grader A (1999) Applications of neural networks in multiwell field development. In: SPE annual technical conference and exhibition proceedings. SPE 56433. 3–6 October. Houston, Texas

12. Doraisamy H, Ertekin T, Grader A (2000) Field development studies by neuro-simulation: an effective coupling of soft and hard computing protocols. Comput Geosci 26(8):963–973

13. Mohaghegh S, Modavi A, Hafez H, Haajizadeh M, Kenawy M, Guruswamy S (2006) Development of surrogate reservoir models (SRM) for fast-track analysis of complex reservoirs. In: SPE intelligent energy conference and exhibition proceedings. SPE 99667. 11–13 April. Amsterdam, The Netherlands

14. Johnson V, Rogers L (2001) Applying soft computing methods to improve the computational tractability of a subsurface simulation-optimization problem. J Pet Sci Eng 29(3–4):153–175

15. Guyaguler B (2002) Optimization of well placement and assessment of uncertainty. Ph.D. dissertation, Stanford University, Stanford, California

16. Yeten B, Durlofsky L, Aziz K (2003) Optimization of nonconventional well type, location, and trajectory. SPE J 8(3):200–210

17. Patel AN, Davis D, Guthrie CF, Tuk D, Nguyen T, Williams J (2005) Optimizing cyclic steam oil production with genetic algorithms. In: SPE western regional meeting proceedings. SPE 93906. 30 March–1 April, Irvine, California

18. Ayala L, Ertekin T (2005) Analysis of gas-cycling performance in gas/condensate reservoirs using neuro-simulation. In: SPE annual technical conference and exhibition proceedings. SPE 95655. 9–12 October. Dallas, Texas

19. Artun E, Ertekin T, Watson R, Miller B (2010) Development and testing of proxy models for screening cyclic pressure pulsing process in a depleted, naturally fractured reservoir. J Pet Sci Eng 73(1):73–85

20. Artun E, Ertekin T, Watson R, Al-Wadhahi M (2011) Development of universal proxy models for screening and optimization of cyclic pressure pulsing in naturally fractured reservoirs. J Nat Gas Sci Eng 3(6):667–686

21. Artun E, Ertekin T, Watson R, Miller B (2012) Designing cyclic pressure pulsing in naturally fractured reservoirs using an inverse-looking recurrent neural network. Comput Geosci 38(1):68–79

22. Parada CH, Ertekin T (2012) A new screening tool for improved oil recovery methods using artificial neural networks. In: SPE western regional meeting proceedings. SPE 153321. 19–23 March, Bakersfield, California

23. Durzman PJ, Leung J, Zanon SD, Amirian E (2013) Data-driven modeling approach for recovery performance prediction in SAGD operations. In: Proceedings of the SPE heavy oil conference. SPE 165557. 11–13 June, Calgary, Alberta, Canada

24. Cullick A, Johnson D, Shi G (2006) Improved and more rapid history matching with a nonlinear proxy and global optimization. In: SPE annual technical conference and exhibition proceedings. SPE 101933. 24–27 September. San Antonio, Texas

25. Silva P, Clio M, Schiozer D (2007) Use of neuro-simulation techniques as proxies to reservoir simulator: application in production history matching. J Pet Sci Eng 57(3–4):273–280

26. Zhao H, Kang Z, Zhang X, Sun H, Cao L, Reynolds A (2015) INSIM: a data-driven model for history matching and prediction for waterflooding monitoring and management with a field application. In: Proceedings of the SPE reservoir simulation symposium. SPE 173213. 23–25 February, Houston, Texas, USA

27. Zangl G, Giovannoli M, Stundner M (2006) Application of artificial intelligence in gas storage management. In: SPE Europec/EAGE annual conference and exhibition proceedings. SPE 100133. 12–15 June. Vienna, Austria

28. Mohaghegh S, Al-Mehairi Y, Gaskari R, Maysami M, Khazaeni Y, Gashut M, Al-Hammadi AE, Kumar S (2014) Data-driven reservoir management of a giant mature oilfield in the Middle East. In: Proceedings of the SPE annual technical conference and exhibition. SPE 170660. 27–29 October. Amsterdam, The Netherlands

29. Esmaili S, Mohaghegh S (2015) Full field reservoir modeling of shale assets using advanced data-driven analytics. Geosci Front. doi:10.1016/j.gsf.2014.12.006

30. Kalantari-Dhaghi A, Mohaghegh S, Esmaili S (2015) Data-driven proxy at hydraulic fracture cluster level: a technique for efficient CO_2-enhanced gas recovery and storage assessment in shale reservoir. J Nat Gas Sci Eng. doi:10.1016/j.jngse.2015.06.039

31. Giraldo R, Delicado P, Mateu J (2012) Hierarchical clustering of spatially correlated functional data. Stat Neerl 66(4):403–421

32. Zhou J, Duan B, Huang J, Hongjun C (2014) Data-driven modeling and optimization for cavity filters using linear programming support vector regression. Neural Comput Appl 24(7–8):1771–1783

33. Dong J, Zheng C, Kan G, Zhao M, Wen J, Yu J (2015) Applying the ensemble artificial neural network-based hybrid data-driven model to daily total load forecasting. Neural Comput Appl 26(3):603–611

34. Fausett L (1994) Fundamentals of neural networks: architectures, algorithms, and applications. Prentice-Hall, Englewood Cliffs

35. Neuroshell (1998) Neuroshell 2 tutorial. Ward Systems Inc, Frederick

36. Demuth H, Beale M (2002) Neural network toolbox for use with MATLAB. Mathworks Inc, Natick

37. MATLAB (2013) MATLAB neural network toolbox, version 2013a. Mathworks Inc, Natick

38. Bruce W (1943) An electrical device for analyzing oil-reservoir behavior. Trans AIME 151(1):112–124 SPE-943112-G

39. Weber D. (2008) The use of capacitance–resistance models to optimize injection allocation and well location in waterfloods. Ph.D. dissertation, The University of Texas at Austin, Austin, Texas

40. Biswas D (2011) Techniques enhance reservoir studies. The American Oil and Gas Reporter 54(7). Derby, Kansas

41. Nguyen A, Kim J, Lake L, Edgar T, Haynes B (2011) Integrated capacitance-resistive model for reservoir characterization in primary and secondary recovery. In: SPE annual technical conference and exhibition proceedings. SPE 147344. 30 October–2 November. Denver, Colorado

42. Kim J, Lake L, Edgar T (2012) Integrated capacitance–resistance model for characterizing waterflooded reservoirs. In: IFAC workshop on automatic control in offshore oil and gas production proceedings. Norwegian University of Science and Technology, Trondheim, Norway

43. Yousef A, Jensen J, Lake L (2006) A capacitance model to infer interwell connectivity from production and injection rate fluctuations. SPE Res Eval Eng 9(6):630–646

44. Sayarpour M, Zuluaga E, Kabir C, Lake L (2009) The use of capacitance-resistive models for rapid estimation of waterflood performance and optimization. J Pet Sci Eng 69(3–4):227–238

45. Liang X, Weber B, Edgar T, Lake L, Sayarpour M, Al-Yousef A(2007) Optimization of oil production based on a capacitance model of production and injection rates. In: SPE hydrocarbon economics and evaluation symposium proceedings. SPE 107713. 1–3 April. Dallas, Texas

46. Shephard D (1968) A two-dimensional interpolation for irregularly-spaced data. In: 23rd ACM national conference proceedings, New York, pp 517–524

47. Moré JJ, Sorensen DC (1983) Computing a trust region step. SIAM J Sci Stat Comput 3:553–572

48. Steihaug T (1983) The conjugate gradient method and trust regions in large scale optimization. SIAM J Numer Anal 20:626–637

49. Byrd RH, Schnabel RB, Shultz GA (1988) Approximate solution of the trust region problem by minimization over two-dimensional subspaces. Math Program 40:247–263

50. Branch MA, Coleman TF, Li Y (1999) A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. SIAM J Sci Comput 21(1):1–23

51. MATLAB (2013) MATLAB optimization toolbox, version 2013a. Mathworks Inc, Natick