



Geochemical equilibrium determination using an artificial neural network in compositional reservoir flow simulation

Dominique Guérillot¹ · Jérémie Bruyelle¹

Received: 13 September 2018 / Accepted: 18 July 2019 / Published online: 28 November 2019
© The Author(s) 2019

Abstract

The fluid injection in sedimentary formations may generate geochemical interactions between the fluids and the rock minerals, e.g., CO₂ storage in a depleted reservoir or a saline aquifer. To simulate such reactive transfer processes, geochemical equations (equilibrium and kinetics equations) are coupled with compositional flows in porous media in order to represent, for example, precipitation/dissolution phenomena. The aim of the decoupled approach proposed consists in replacing the geochemical equilibrium solver with a substitute method to bypass the huge consuming time required to balance the geochemical system while keeping an accurate equilibrium calculation. This paper focuses on the use of artificial neural networks (ANN) to determine the geochemical equilibrium instead of solving geochemical equations system. To illustrate the proposed workflow, a 3D case study of CO₂ storage in geological formation is presented.

Keywords Reservoir simulation · Compositional · Heterogeneity · CO₂ storage · Chemically reacting flows · Artificial neural network

Mathematics Subject Classification (2010) 80A32 · 76S05 · 68T05

1 Introduction

The transport of a geochemical species through porous media is related, in particular, to the interactions and geochemical reactions in which this species participates. These reactions occur, for example, during the re-injection of a CO₂-rich gas, CO₂ storage in a depleted reservoir or a saline aquifer, or stimulation of wells by acid injection. Understanding these phenomena is essential in order to elaborate on different scenarios of field developments, storage plans, or well treatments.

To simulate such reactive transport processes, compositional flows in porous media with complex mineralogy must be coupled with the geochemical equilibria in the fluid phases and the precipitation/dissolution reactions of the minerals [23].

Depending on the number of geochemical species, the number of geochemical equations (equilibrium and kinetics), and the dimension of the geological model, a significant part of the time consumed during reactive transport simulation is related to the resolution of the geochemical system.

Global or decoupled approaches have been developed to solve this type of problem. The work presented here is part of the decoupled approaches (sequential non-iterative approach), where species are transported without geochemical reaction, then the geochemical system is solved for each cell at each time-step. This resolution of the geochemical system represents the most important part of the calculation time consumed. To reduce the computation time, the number of species in solution is often reduced. However, such an assumption leads to a loss of accuracy of results.

Instead of simplifying the geochemical model, an innovative approach that mimics the resolution of geochemical equilibrium is proposed here. The aim of this approach is to provide a substitute method to bypass the huge consuming time required to balance the geochemical system.

The paper starts with a description of previous works. Then, the reactive compositional multi-phase flow model with geochemical reactions is given. An important part of our methodology needs the use of artificial neural network

✉ Dominique Guérillot
dominique.guerillot@qatar.tamu.edu

Jérémie Bruyelle
jeremie.bruyelle@qatar.tamu.edu

¹ Texas A&M University at Qatar, Doha, Qatar

(ANN); we recall here the definition. The methodology to replace the geochemical equations by an ANN is described. The validation of the proposed methodology is done through a case study of calcite dissolution. Finally, an application to a CO₂ storage in a geological formation demonstrates the feasibility and efficiency of the proposed approach.

2 Previous work

Several techniques aim to substitute the forward problems using proxies; these methods are part of response surface models (RSM). Among the proxy methods, the most used ones were polynomials and kriging methods. However, these proxy models are not sufficiently sophisticated to represent the non-linearities of real problems. Artificial neural networks are proxy models with the ability to reproduce complex and non-linear physical phenomena. ANN requires significant training dataset to learn and replicate the physical problem. With the increase in computing power and the amount of data produced, artificial neural networks have become an increasingly used tool in the industry.

Regarding the use of a proxy model in reactive transport modeling, Guéillot [11, 12] proposes a complete workflow to replace thermodynamic equilibria by an artificial neural network designed and trained in advance to reproduce these flashes during compositional reservoir simulations. Jatnieks et al. [19] propose to replace the geochemical model by a data-driven surrogate model. They compare 32 statistical and machine learning methods on a 1D case study corresponding to the injection of a reactive solution leading to the dissolution of calcite and the precipitation of dolomite. Their results show a good agreement with the simulation results and highlight that accuracy highly depends on the training sample. Leal et al. [20] propose a machine learning approach to calculate equilibrium states in reactive transport simulations where the learning is carried out during the actual simulation without an initial training phase. This approach leads to an impressive speed-up factor range of 60–125.

This paper fits into the framework defined by [11, 12] and [19] and focuses on the use of artificial neural network to replace the geochemical equilibrium package in a 3D fluid flow simulator.

3 Reactive compositional multi-phase flow model with geochemical reactions

A rigorous model of multi-phase flow in porous media derived from fundamental mechanical equations and principles has been described by Marle [23]. In his paper, macroscopic balance equations for components, momentum, and energy

are established for multi-phase flow with diffusion, geochemical reactions, heat transfer, and exchanges of components between phases in a porous medium.

These equations are established separately for each fluid phase, for the solid part of the medium, and for interfaces, by starting from the corresponding equations valid at the pore level and taking their mean value around each point.

From the equation derived by Marle [23], the partial differential equations that we are considering are given below. Geochemical reactions are considered with both equilibrium reactions, precipitation, and dissolution reactions.

The compositional reactive transport is based on the mass balance equation as below

$$\partial_t C_\alpha + \nabla J_\alpha + M_\alpha \sum_{r \in R} \nu_{\alpha r} \xi_r = q_\alpha \quad \alpha = 1, \dots, n \quad (1)$$

where C_α is the mass of the species α , J_α is the total flux (convective and diffusive), M_α is the molar mass of the species α , $\nu_{\alpha r}$ is the stoichiometric coefficient of species α in reaction r , ξ_r is the rate of geochemical reaction r , q_α is the source term of the species α , and n is the number of species.

$$C_\alpha = \varphi \sum_{l=1}^P \rho_l s_l c_{\alpha l} + (1 - \varphi) \sum_{s=1}^M \rho_s c_{\alpha s} \quad (2)$$

where φ is the porosity, P is the number of fluid phase, s_l is the saturation of phase l , M is the number of solid phase, ρ is the phase density, and c_α is the mass fraction of the species α in phase.

$$J_\alpha = \sum_{l=1}^P \left(\rho_l c_{\alpha l} \mathbf{v}_l - \varphi \rho_l s_l \overline{\overline{\mathbf{D}}}_{\alpha l} \nabla c_{\alpha l} \right) \quad (3)$$

where \mathbf{v}_l is the Darcy flux of phase l and $\overline{\overline{\mathbf{D}}}_{\alpha l}$ is the diffusion tensor of the species α in phase l .

$$\mathbf{v}_l = -\overline{\overline{\mathbf{k}}}_{r l} \frac{k_{r l}}{\mu_l} (\nabla p + \rho_l g \nabla z) \quad l = 1, \dots, P \quad (4)$$

where $\overline{\overline{\mathbf{k}}}_{r l}$, $k_{r l}$, and μ_l are respectively the permeability tensor, the relative permeability and the viscosity of phase l , g is the gravity constant, and z is the depth.

To close the system, thermodynamic and geochemical equilibrium equations are used.

3.1 Thermodynamic equilibrium

The phase behavior is modeled by an equation of states (EOS) and phase equilibrium relations. The phase equilibrium of a species can be defined by the equality of the fugacities. The fugacities of a species are equal in each phase.

$$f_{l1} = f_{\alpha l}, \quad l = 2, \dots, P \text{ and } \alpha = 1, \dots, n \quad (5)$$

The two most used state equations are that of Soave, Redlich, and Kwong [29] and that of Peng-Robinson [26].

However, these models have some limitations, e.g., inaccuracy of oil compressibilities and gas compressibility factors [28]. More advanced equations of state exist (e.g., CPA or PC-SAFT) but these approaches increase the computational cost. These EOS are solved using an iterative method until equality of the fugacities in phases.

3.2 Geochemical equilibrium

The geochemical reactions are modeled through equilibrium reactions and kinetic reactions [2]. The equilibrium reactions are described by non-linear algebraic equations. The kinetic reactions are described by non-linear ordinary differential equations. Several types of chemical reactions can occur, e.g., aqueous complexation, mineral dissolution and precipitation, redox reaction, and gas dissolution and exsolution. A description of the geochemical reactions commonly used is given by Merkel et al. [24].

The set of non-linear equations is usually solved with an iterative method, e.g., the Newton-Raphson method. A detailed review of the geochemical system formulation is given by Lichtner [21, 22].

For example, for a geochemical system composed of n_{aq} species in the aqueous phase, $n_{r_{aq}}$ equilibrium reactions, n_{r_m} precipitation/dissolution reactions and assuming that reaction rates for equilibrium reactions are equal to zero, the reaction system to solve is:

$$\begin{cases} \prod_{\alpha=1}^{n_{aq}} (\gamma_{\alpha} b_{\alpha}^*)^{v_{\alpha r_{aq}}} - K_{eq,r_{aq}} = 0, & r_{aq} = 1, n_{r_{aq}} \\ \xi_{r_m} = A_{r_m} k_{r_m} \left(1 - \frac{\prod_{\alpha=1}^{n_{aq}} (\gamma_{\alpha} b_{\alpha}^*)^{v_{\alpha r_m}}}{K_{eq,r_m}} \right), & r_m = 1, n_{r_m} \end{cases} \quad (6)$$

with

$$b_{\alpha}^* = b_{\alpha} + \sum_{\beta=1}^{n_r} v_{\alpha\beta} \xi_{\beta} \quad (7)$$

where γ_{α} is the activity coefficient and b_{α} is the molality of α , $v_{\alpha r_{aq}}$ is the stoichiometric coefficient of species α in reaction r_{aq} , $K_{eq,r_{aq}}$ is the chemical equilibrium constant for the reaction r_{aq} , ξ_{r_m} is the reaction rate of the reaction r_m , A_{r_m} is the reactive surface area of the mineral r_m , k_{r_m} is the reaction rate constant of the reaction r_m and K_{eq,r_m} is the chemical equilibrium constant for the reaction r_m , ξ_{β} is the reaction rate of the reaction β , and n_r is the total number of reaction (both equilibrium and kinetics reactions). Note that γ_{α} depends on the ionic strength of the solution which depends on the charge and concentration of the species.

3.3 Coupling transport and geochemical reactions

The coupling of transport with the geochemical reaction has been widely studied in the field of hydrogeology [31]. Numerous simulation codes have been developed,

e.g., CrunchFlow [32], Fereact [34], Hydrogeochem [38], Toughreact [35], and Pflotran [15].

The codes mentioned above can be divided into two different groups: (1) the global implicit approach (GIA) where the simulator calculates the whole system simultaneously (transport and chemical reactions) and (2) the operator-splitting approaches where the transport and chemical reactions are solved sequentially.

There are three commonly used operator-splitting approaches. The strang splitting approach (SSA) [33] divides the time-step in half: first, the species are transported assuming no geochemical reaction occurs, then the geochemical calculation is performed between these two time-steps. The sequential non-iterative approach (SNIA) splits the calculation into two steps: first, species are transported assuming no geochemical reaction occurs, then the geochemical system is solved for a given time-step for each cell of the geological model. The sequential iterative approach (SIA) [36, 37] splits the calculation into two steps: first, species are transported by taking a geochemical reaction source term, then the geochemical system is solved for a given time-step for each cell of the geological model. The source terms are calculated from these two steps. The process iterates until convergence of the computation of the source terms.

The classical approach used to model reactive transport in porous media consists in partitioning the geochemical system into N_c primary species and N_x secondary species. The concentration of secondary species is related to the concentration of primary species using the mass action law. This approach allows reducing the number of mass conservation equations to the number of primary species. The mathematical formulation for arbitrarily reaction systems (equilibrium and kinetics reaction) has been described by Lichtner [21, 22] and Steefel and MacQuarrie [30].

This approach allows eliminating equilibrium reaction but not the kinetics reactions which are dependent on the activities of species. However, the activities are highly dependent on the ionic strength that depends on concentrations of all species. The inaccuracy of activity calculations is a major error of geochemical calculation.

The work presented here is part of sequential non-iterative approaches, where species are transported without geochemical reaction, then the geochemical system is given for each cell at each time-step with an artificial neural network.

4 Artificial neural network design

An artificial neural network is a mathematical model composed of several computing units (neuron) connected to each other [3, 4]. Each neuron is an elementary process

that receives a number of inputs from other neurons. For each of these inputs, a weight is assigned to give it more or less importance. Each neuron processes the information collected via an activation function and passes the results to other neurons. The activation functions most typically are threshold functions, piecewise linear functions, and sigmoidal function. A highly detailed theoretical background about artificial neural network is presented in [3] and [14].

Here, we consider a classical formulation of ANN, i.e. a multilayer perceptron with one hidden layer (Fig. 1), where the activation functions for hidden layers are sigmoidal functions and the activation functions for output layers are linear functions. In fact, the universal approximation capacity of this type of neural networks has been demonstrated by several authors [9, 10, 16–18].

The backpropagation of the gradient is used as the learning algorithm [3, 27].

The number of neurons in the hidden layer is determined following a cross-validation approach on a set of neural networks [13].

In order to make the learning algorithm efficient and to prevent that certain variables have different numerical values by several orders of magnitude compared with others, it is necessary to perform a preprocessing of the data. The normalization between 0 and 1 of the logarithmic transformation of the data is applied to normalize the inputs and the outputs of the ANN.

The prediction quality of these proxy models depends on the size of learning dataset and depends on the quality of samples making up the database [7]. In fact, these samples must cover the area of the parameters of the model. An experimental design approach has been used (Box-Behnken) to generate the learning database to which random samples have been added.

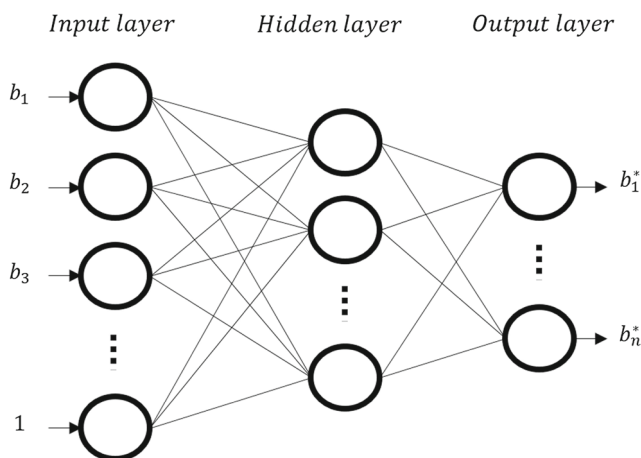


Fig. 1 An artificial neural network with one hidden layer. Each neuron receives its inputs from the previous layer and sends its output to the next layer

One of the difficulties in replacing a geochemical solver by an ANN in fluid flow simulations is due to the variable time-step. In fact, at each iteration of the IMPEC scheme, the modification of the time-steps Δt_p and Δt_c (increasing or decreasing) is given by formulas depending on the parameters. The implicit model for pressure being unconditionally stable, a large time-step can be used to compute the pressure. In practice, the variation of the time-step is limited [1]. In the explicit model for the mass fractions being limited by CFL conditions [8], many smaller time-steps are required to reach the end of the pressure time-step.

For this first application, the time-step Δt_c is fixed. At the beginning of the fluid flow simulation, while the time-step increases, the chemical reaction system is solved. Then, when the simulation reaches a given time-step, the ANN is used to replace the geochemical solver.

An ANN with the time-step as input would have the advantage to be used throughout the simulation without using the geochemical solver. The computational cost should be considerably reduced. However, including the time-step as an input of the ANN leads to drastic size increase of the learning base to cover the parameter space.

In this paper, the inputs define the geochemical system at initial condition, i.e., reactive surface area and the molality of aqueous species. The outputs are the concentration of species allowing to define the reaction rates over a specified time-step.

The accuracy of the ANN is given by the root mean square error (RMSE) between the output values given by the ANN and the output values given by the geochemical solver:

$$\text{RMSE}_{\text{ANN}} = \sqrt{\frac{1}{2dn} \sum_{i=1}^d \sum_{j=1}^n (b_{i,j}^{*S} - b_{i,j}^{*A})^2} \quad (8)$$

where d is the number of data, n is the number of output of the ANN, $b_{i,j}^{*A}$ is the molality of the i th geochemical species for the j th sample of the database calculated by the ANN, and $b_{i,j}^{*S}$ is the molality given by the geochemical solver.

The methodology is summarized in Fig. 2. Note that at each iteration, the geochemical reactions lead to a modification of the rock parameters (porosity and permeability). The permeability is updated using the new porosity value following the Kozeny-Carman relationship [5].

The resolution of the geochemical system being done by the proxy model, this approach, cannot be fully mass conserved. The error generated at each time-step grows with time. The increase of error highly depends on the accuracy of the ANN. The learning phase is also the most important step to limit the increase of error. If the accuracy of the ANN is unsatisfactory, the size of the learning database must be

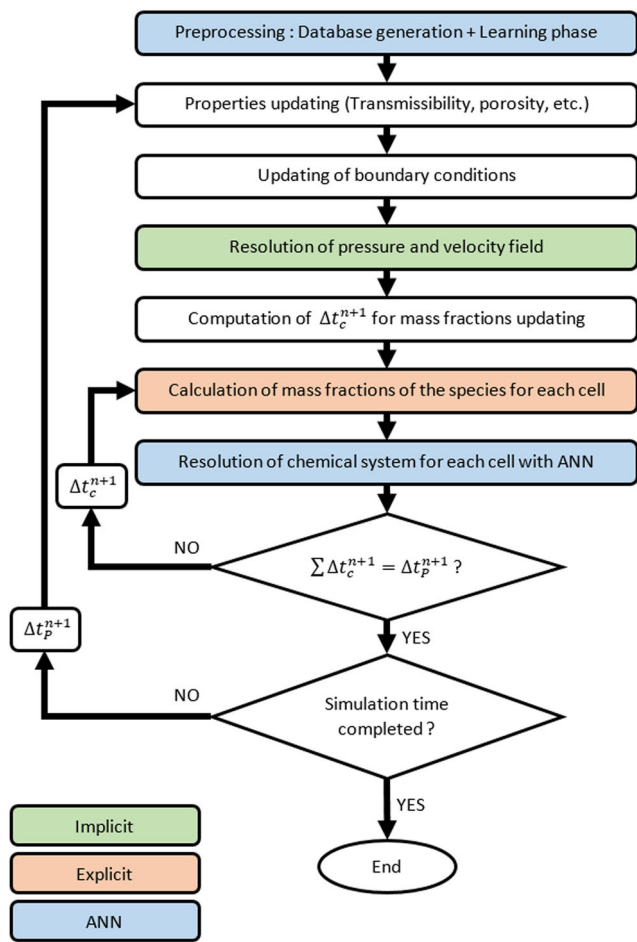
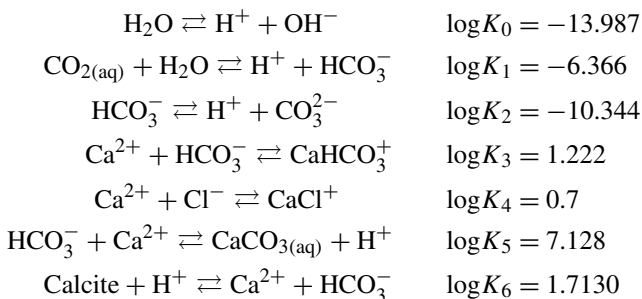


Fig. 2 Flow chart of methodology

increased by adding new samples to better cover the space of input parameters.

5 Validation of the methodology

The validation case is taken from Nghiem et al. [25]. It corresponds to a system that contains 10 g of calcite and 100 kg of H₂O. The compositional system is composed of 10 species in solution, 1 mineral component, 7 geochemical equilibrium reactions, and 1 mineral dissolution-precipitation reaction:



The kinetics of the calcite dissolution is given by the rate law with an initial reactive surface area $A_c^0 = 88 \times 10^{-3} \text{ m}^2\text{kg}^{-1}$ and a reaction rate $k_c = 1.6 \times 10^{-9} \text{ molm}^{-2}\text{s}^{-1}$.

The system to solve is:

$$\begin{cases} K_0 - [H^+][OH^-] = 0 \\ K_1 [CO_{2(aq)}] - [H^+][HCO_3^-] = 0 \\ K_2 [HCO_3^-] - [H^+][CO_3^{2-}] = 0 \\ K_3 [Ca^{2+}][HCO_3^-] - [CaHCO_3^+] = 0 \\ K_4 [Ca^{2+}][Cl^-] - [CaCl^+] = 0 \\ K_5 [HCO_3^-][Ca^{2+}] - [H^+][CaCO_{3(aq)}] = 0 \\ \frac{\xi_6}{\Delta t} - k_c A_c^0 \left(1 - \frac{[Ca^{2+}][HCO_3^-]}{[H^+]K_6} \right) = 0 \end{cases} \quad (9)$$

with

$$[\alpha] = \gamma_\alpha \left(b_\alpha + \sum_{\beta=0}^6 \nu_{\alpha\beta} \xi_\beta \right) \quad (10)$$

where α is a species, γ_α is the activity coefficient of α , b_α is the molality of α , $\nu_{\alpha\beta}$ is the stoichiometric coefficient of α in reaction β , and ξ_β is the rate of geochemical reaction β

The inputs of the solver are the reactive surface area, the molality of 10 aqueous species, and a time-step Δt . The outputs of the solver are the rates ξ of formation/consumption of species.

The artificial neural network aims to replace the geochemical equilibrium solver for a given time-step. So, the artificial neural network is composed of 11 inputs and 10 outputs (Eq. 11).

$$\text{Input} = \begin{bmatrix} A_c \\ b_{CO_{2(aq)}} \\ b_{H^+} \\ b_{OH^-} \\ b_{HCO_3^-} \\ b_{CO_3^{2-}} \\ b_{Ca^{2+}} \\ b_{CaHCO_3^+} \\ b_{Cl^-} \\ b_{CaCl^+} \\ b_{CaCO_{3(aq)}} \end{bmatrix} \quad \text{Output} = \begin{bmatrix} b_{CO_{2(aq)}}^* \\ b_{H^+}^* \\ b_{OH^-}^* \\ b_{HCO_3^-}^* \\ b_{CO_3^{2-}}^* \\ b_{Ca^{2+}}^* \\ b_{CaHCO_3^+}^* \\ b_{Cl^-}^* \\ b_{CaCl^+}^* \\ b_{CaCO_{3(aq)}}^* \end{bmatrix} \quad (11)$$

The inputs define the geochemical system at initial condition, i.e., reactive surface area and the molality of 10 aqueous species. The outputs are the molality of species over a specified time-step. For this validation case study, we have fixed Δt equal to 10 days in order to track the evolution of species for 10 time-steps. The sensitivity related to the choice of Δt has not been studied here.

The ionic strength of the solution over a specified time-step is calculated from the outputs of the neural network.

Fig. 3 Validation case study — Green, learning data; red, validation data

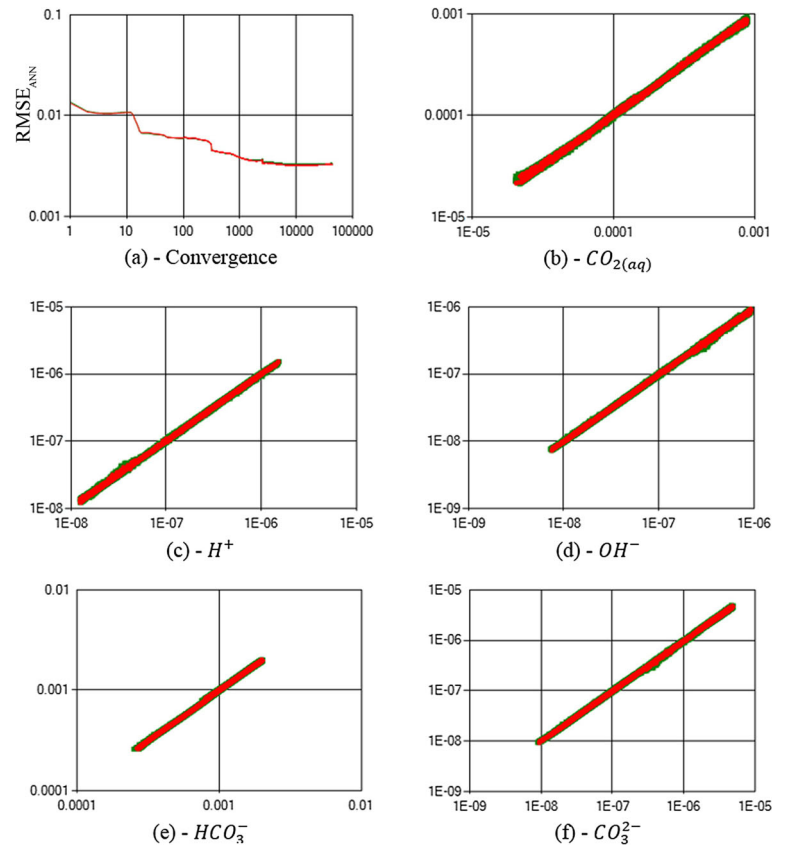
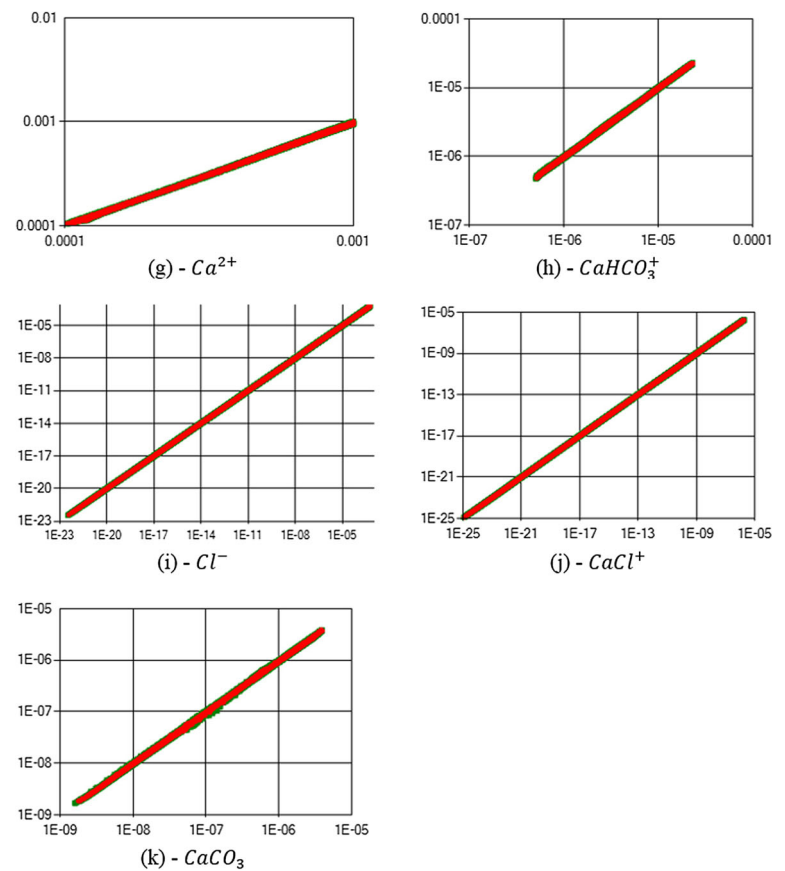


Fig. 4 Validation case study — Green, learning data; red, validation data



The activity coefficients are calculated from the ionic strength. The reaction rate ξ_6 is given by:

$$\xi_6 = \Delta t k_c A_c^0 \left(1 - \frac{\gamma_{Ca^{2+}}^* b_{Ca^{2+}}^* \gamma_{HCO_3^-}^* b_{HCO_3^-}^*}{\gamma_{H^+}^* b_{H^+}^* K_6} \right) \quad (12)$$

The mole number of mineral and the reactive surface area are deduced from the reaction rates.

$$n_{calcite}^* = n_{calcite} - \xi_6 \times m_{H_2O} \quad (13)$$

$$A_c^* = A_c \times \frac{n_{calcite}^*}{n_{calcite}} \quad (14)$$

The ANN is composed of one hidden layer with 11 hidden neurons.

The training database has been generated by solving the geochemical equilibrium of 50,000 random sample values for a given $\Delta t=864,000$ s. A validation database has been generated by solving the geochemical equilibrium of 5000 random sample values for a given $\Delta t=864,000$ s. Different size for the learning database have been tested: 20,000, 40,000, and 50,000. The ANN obtained with 20,000 and 40,000 was not able to reproduce the geochemical equilibrium of the validation database. The size of the database has been increased. While the ANN cannot reproduce the data, the learning database must be improved by adding new samples.

Figure 3a shows the convergence curves of the learning phase. The decrease of the error during the learning phase is slow, and several levels of stagnation are observed. The cross-validation plots (Fig. 4b–k) show a high capacity of the ANN to reproduce both the learning data and the validation data.

After the learning phase, the ANN is used to replace the geochemical solver. A loop is performed where outputs become inputs of the ANN for the sequence of time-steps.

Figures 5, 6, and 7 present the evolution of species versus time and show a high capacity of the ANN to reproduce the geochemistry calculations. The speed-up factor, defined as the run-time ratio of total computation time of

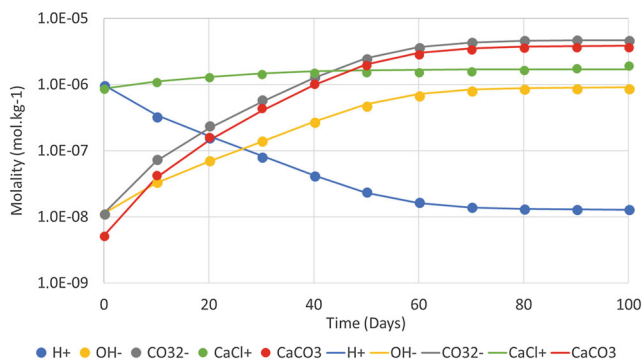


Fig. 5 Validation case study—Comparison of neural network (bullet) and geochemistry solver (dashed line)

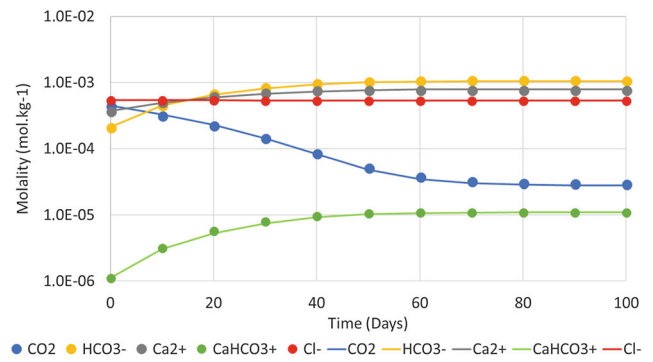


Fig. 6 Validation case study—Comparison of neural network (bullet) and geochemistry solver (dashed line)

geochemical solver and the total computation time of artificial neural network, is equal to 45.4. This measure has been done on a quadcore Intel Xeon E3-1535M v5 central processing unit (CPU) at 2.9 GHz. The speed-up factor do not take into account the CPU time in the learning step (generation of the database and learning). This overall CPU time in the learning step is less than 1 min. For the following application of CO₂ storage, we will use the same ANN for the flow simulation. For this case, 50,000 geochemical equilibrium of the learning base are equivalent to 15 time-steps Δt . For large-scale reservoir applications (a few hundred thousand cells), the learning phase should be equivalent to less than one time-step. Moreover, the speed-up factor should increase due to the number of cells.

6 Application to CO₂ storage in geological formation

The methodology is applied to a synthetic case that corresponds to a part of a layer extracted to the “SPE-10 2nd comparative solution” project [6] (cf. Fig. 8). The geological mesh of the model is composed of $30 \times 110 \times 1$ grid cells (3 300 cells) with dimensions in x, y , and z respectively equal to 6 m, 3 m, and 0.6 m. The considered model contains two

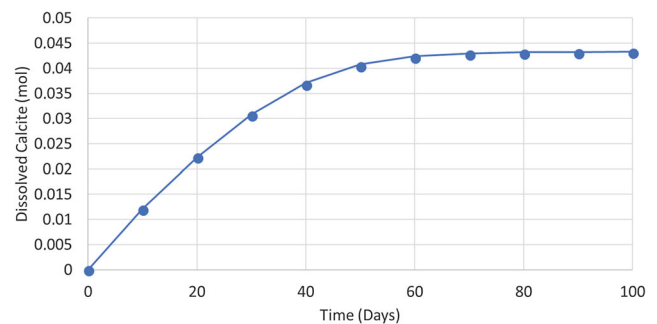


Fig. 7 Validation case study—Comparison of neural network (bullet) and geochemistry solver (dashed line)

Fig. 8 CO₂ storage case study—Porosity and permeability distribution

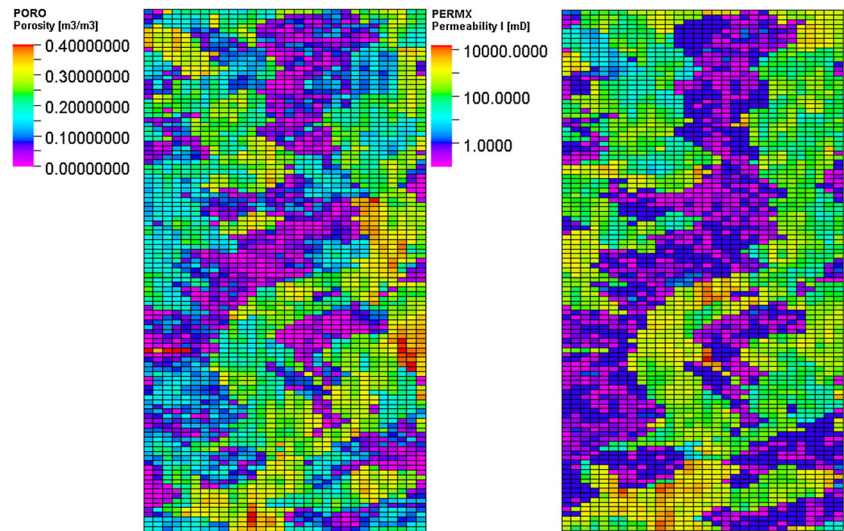


Fig. 9 CO₂ storage case study—pH plume. **a** End of injection. **b** End of simulation with geochemical solver. **c** End of simulation with ANN

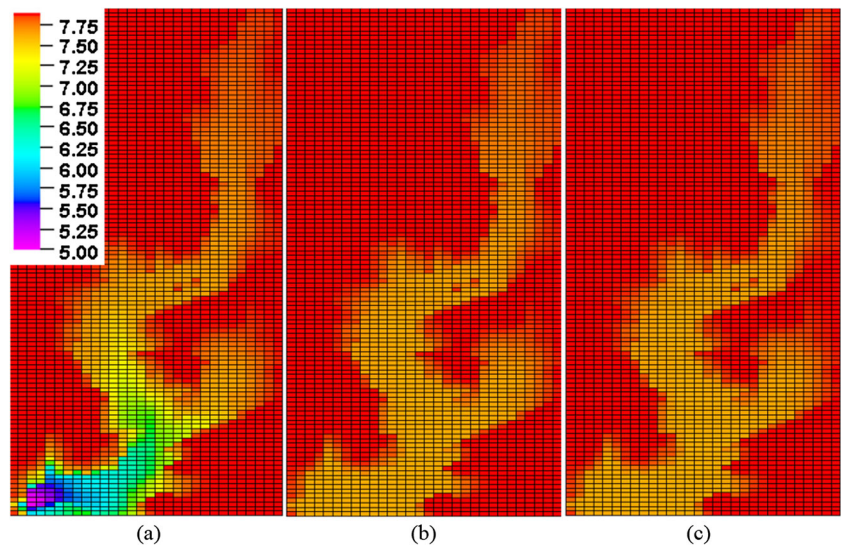
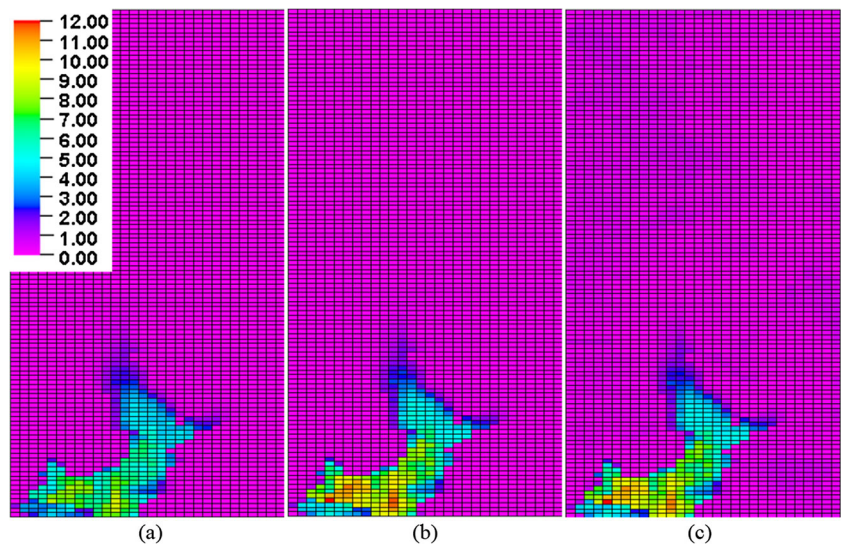


Fig. 10 CO₂ storage case study—Calcite dissolved (mol). **a** End of injection. **b** End of simulation with geochemical solver. **c** End of simulation with ANN



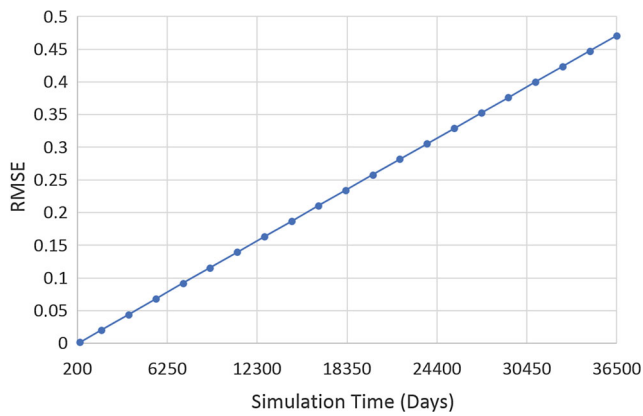


Fig. 11 Evolution of RMSE for the dissolved calcite (mol)

wells (one injector and one producer) which are vertical and completely throughout the whole reservoir formation.

The scenario corresponds to the injection of dissolved CO₂ for 6 months. Only one phase is present. The fluid is injected at 10 m³/day with a maximum bottom hole pressure equal to 690 bar. The fluid is produced at a constant bottom hole pressure of 275 bar. After the injection period, wells are shut-in and the simulation is carried out for 100 years to track the evolution of pH and the dissolved calcite.

The geochemical system is the same as described in the previous section. The geochemical solver is replaced by the artificial neural network after the injection period with a constant time-step equals to 10 days.

Figures 9 and 10 show the results obtained with geochemical solver and artificial neural network. At the end of injection, the solution around well is acidic with a pH near to 5 and dissolution reaction takes place between calcite and H⁺. At the end of the simulation, the pH value near well increases to 7.6 and the calcite has been dissolved. Results obtained with ANN are very similar to results obtained using the geochemical solver (Fig. 9). The maximum absolute error for the pH value is equal to 0.0035.

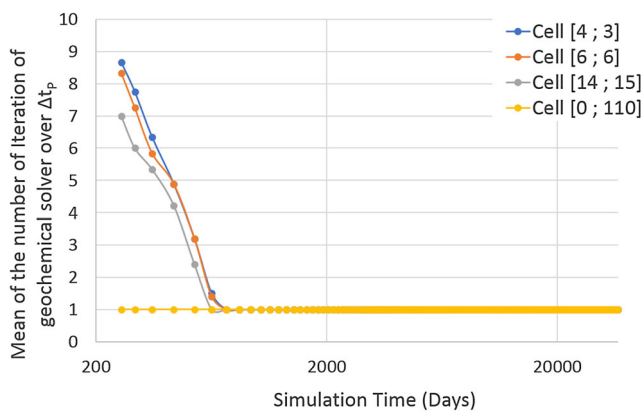


Fig. 12 Mean of the number of iteration of the geochemical solver at each time-step Δt_p for 4 cells

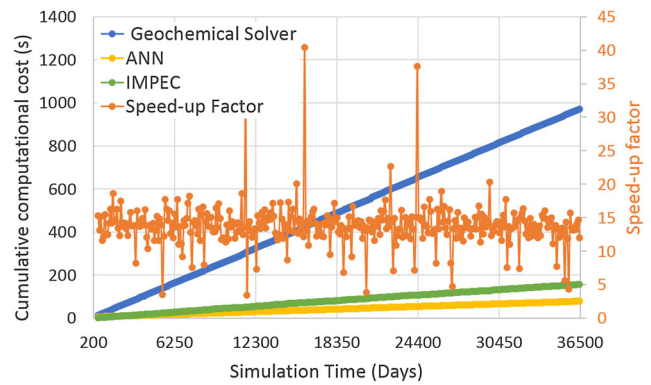


Fig. 13 Cumulative computational time of transport, geochemical equilibrium with solver, and ANN. Speed-up factor of geochemical equilibrium at each time-step Δt_p

The maximum absolute error of dissolved calcite is equal to 1.55 mol (Fig. 10). This error is due to the accumulation of error of the ANN over time. Although this error is very small for a single time-step, this error accumulates over time and becomes more important at the end of the simulation. Figure 11 shows a linear evolution of RMSE for the dissolved calcite for the entire domain.

For the cells located in an area not reached by the CO₂ plume, the geochemical solver converges in one iteration. After the injection period, many cells are not at equilibrium and requires around 8 iteration for each time-step Δt_c . As expected, the more the cells are far from the injector well, the faster the geochemical solver converges. The number of iteration decreases quickly and the geochemical solver converges in 1 iteration (Fig. 12). For these cells with a high CO₂ mass fraction, the cumulative execution time of the geochemical solver is larger than the one for the cells not reach by the CO₂.

The instantaneous speed-up factor, defined as the run-time ratio of total computation time of geochemical solver and the total computation time of artificial neural network

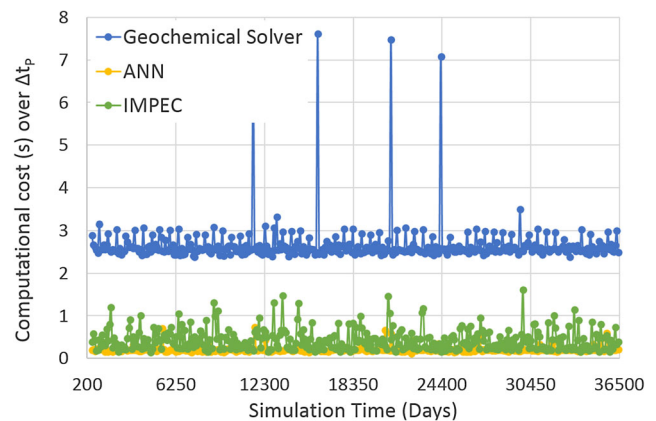


Fig. 14 Computational time of transport, geochemical equilibrium with solver, and ANN at each time-step Δt_p

for a time-step Δt_p , is constant around 14 (Fig. 13). The reduction of computation time concerns only the resolution of the geochemical system. The reduction of the computation time is less important than with the validation case because after the injection period the most part of the cells is not reached by the CO_2 , and after 7 time-steps Δt_p the geochemical solver converges in 1 iteration for all cells for each Δt_c . Even if the geochemical solver converges in 1 iteration, the call to the ANN remains faster than the solver (Fig. 14).

7 Conclusions

The simulation of CO_2 injection requires fine meshes with the most detailed geological description in order to accurately represent the phenomena generated by geochemical reactions such as changes in porosity. The classical approaches used in reservoir simulator do not allow to maintain all the detailed information needed to capture the fluid-rock interactions. The geochemical system is usually simplified.

In this paper, a complete workflow for compositional reservoir simulation using an artificial neural network to perform the geochemical calculation instead of solving equations system has been presented and validated.

Results illustrate the efficiency of the proposed approach, i.e., a decrease of computation time while maintaining the accuracy of results. Although on the presented case study the speed-up factor is important, we recall that it highly depends on many application case parameters such as the dimension of the reservoir model, the geochemical system complexity (number of species and number of equations), the geochemical solver used, the convergence criteria of solvers, and the code implementation. This speed-up factor cannot be generalized for any cases. Other investigations should be done to scale the speed-up factor according to the number of fluid species, mineral species, and reactions.

8 Remarks

The presented case study is a simple geochemical system. For a more complex geochemical system, the CPU time of the geochemical solver will increase but the execution time of the ANN will be the same (instantaneous). The learning phase should require a larger database than the one used here. However, the CPU time due to the learning phase should remain negligible compared with the number of geochemical equilibrium performed during a flow simulation.

As mentioned previously, the approach is not fully mass conservative. Further investigations should be done to solve

this issue. One of the paths to investigate is the use of the ANN results as an initial guess of the geochemical solver.

To further reduce the computation time, the next step will be to introduce the time-step as an input of the artificial neural network. Therefore, this would imply to generate a larger learning database using experimental design to better cover all of the parameter's space.

Acknowledgments Open Access funding provided by the Qatar National Library.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Aziz, K., Settari, A.: Petroleum Reservoir Simulation. Applied Science Publishers, London (1979)
2. Bethke, C.: Geochemical Reaction Modeling: Concepts and Applications. Oxford University Press, New York (1996)
3. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, Oxford (1995)
4. Bruyelle, J., Guérrillot, D.: Neural networks and their derivatives for history matching and reservoir optimization problems. *Comput. Geosci.* **18**(3–4), 549–561 (2014)
5. Carman, P.C.: Fluid flow through granular beds. *Trans. Institut. Chem. Eng. London* **15**, 150–166 (1937)
6. Christie, M.A., Blunt, M.J.: Tenth SPE comparative solution project: A comparison of upscaling techniques. Society of petroleum engineers, SPE 66599 (2001)
7. Costa, L.A.N., Maschio, C., Schiozer, D.J.: Study of the influence of training data set in artificial neural network applied to the history matching process. In: Rio Oil & Gas Expo and Conference (2010)
8. Courant, R., Friedrichs, K., Lewy, H.: Über die partiellen Differenzgleichungen der mathematischen Physik. *Math. Ann.* **100**(1), 32–74 (1928)
9. Cybenko, G.: Continuous Valued Neural Networks with Two Hidden Layers are Sufficient. Technical Report, Department of Computer Science Tufts University (1988)
10. Cybenko, G.: Approximation by superpositions of sigmoidal functions. *Mathematics of Control, Signals, and Systems* (1989)
11. Guérrillot, D.: Procédé et système de modélisation dynamique d'un écoulement de fluide polyphasique, French Patent EP 2791712 A1 (WO2013087846a1) (2011)
12. Guérrillot, D.: Method and system for dynamically modeling a multiphase fluid flow - US Patent App. 14/365,053 (2014)
13. Guérrillot, D.R., Bruyelle, J.: Uncertainty assessment in production forecast with an optimal artificial neural network. Society of Petroleum Engineers. <https://doi.org/10.2118/183921-MS> (2017)
14. Hagan, M.T., Demuth, H.B., Beale, M.H., De Jesús, O.: *Neural Network Design*, vol. 20. Pws Pub, Boston (1996)
15. Hammond, G.E., Lichtner, P.C., Lu, C., Mills, R.T.: PFLOTRAN: Reactive flow & transport code for use on laptops to leadership-class supercomputers. *Groundwater Reactive Transport Models*, 141–159 (2012)

16. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**, 359–366 (1989)
17. Hornik, K., Stinchcombe, M., White, H.: Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Netw.* **3**, 551–560 (1990)
18. Hornik, K.: Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **4**, 251–257 (1991)
19. Jatnieks, J., De Lucia, M., Dransch, D., Sips, M.: Data-driven surrogate model approach for improving the performance of reactive transport simulations. *Energy Procedia* **97**, 447–453 (2016)
20. Leal, A.M., Kulik, D.A., Saar, M.O.: Ultra-fast reactive transport simulations when chemical reactions meet machine learning: chemical equilibrium. [arXiv:1708.04825](https://arxiv.org/abs/1708.04825) (2017)
21. Lichtner, P.C.: Continuum model for simultaneous chemical reactions and mass transport in hydrothermal systems. *Geochim. Cosmochim. Acta* **49**(3), 779–800 (1985)
22. Lichtner, P.C.: Continuum formulation of multicomponent–multiphase reactive transport. *Reactive Transport in Porous Media. Rev. Mineral.* **34**, 1–81 (1996)
23. Marle, C.M.: On macroscopic equations governing multiphase flow with diffusion and chemical reactions in porous media. *Int. J. Eng. Sci.* **20**(5), 643–662 (1982)
24. Merkel, B.J., Planer-Friedrich, B., Nordstrom, D.K.: Groundwater geochemistry. A practical guide to modeling of natural and contaminated aquatic systems, 2 (2005)
25. Nghiem, L., Sammon, P., Grabenstetter, J., Ohkuma, H.: Modeling CO₂ storage in aquifers with a fully-coupled geochemical EOS compositional simulator. In: *SPE/DOE Symposium on Improved Oil Recovery*. Society of Petroleum Engineers (2004)
26. Peng, D.Y., Robinson, D.B.: A new two-constant equation of state. *Ind. Eng. Chem. Fundam.* **15**(1), 59–64 (1976)
27. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning Internal Representations by Error Backpropagation, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, pp. 318–362. MIT Press, Cambridge (1986)
28. Schou Pedersen, K., Hasdberg, C.: PC-SAFT equation of state applied to petroleum reservoir fluids. In: *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers (2007)
29. Soave, G.: Equilibrium constants from a modified Redlich-Kwong equation of state. *Chem. Eng. Sci.* **27**(6), 1197–1203 (1972)
30. Steefel, C.I., MacQuarrie, K.T.B.: Approaches to modeling reactive transport in porous media. *Reactive Transport in Porous Media. Rev. Mineral.* **34**, 83–125 (1996)
31. Steefel, C.I., Appelo, C.A.J., Arora, B., Jacques, D., Kalbacher, T., Kolditz, O., Lagneau, V., Lichtner, P.C., Mayer, K.U., Meeussen, J.C.L., Molins, S., Moulton, D., Shao, H., Šimůnek, J., Spycher, N., Yabusaki, S.B., Yeh, G.T.: Reactive transport codes for subsurface environmental simulation. *Comput. Geosci.* **19**, 445–478 (2015)
32. Steefel, C.I., Molins, S.: Crunchflow, software for modeling multicomponent reactive flow and transport, *USER'S MANUAL* (2016)
33. Strang, G.: On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.* **5**(3), 506–517 (1968)
34. Tebes-Stevens, C., Valocchi, A.J., VanBriesen, J.M., Rittmann, B.E.: Multicomponent transport with coupled geochemical and microbiological reactions: Model description and example simulations. *J. Hydrol.* **209**(1–4), 8–26 (1998)
35. Xu, T., Spycher, N., Sonnenthal, E., Zhang, G., Zheng, L., Pruess, K.: TOUGHREACT Version 2.0: A Simulator for subsurface reactive transport under non-isothermal multiphase flow conditions. *Comput. Geosci.* **37**(6), 763–774 (2011)
36. Yeh, G.T., Tripathi, V.S.: A critical evaluation of recent developments in hydrogeochemical transport models of reactive multichemical components. *Water Resour. Res.* **25**(1), 93–108 (1989)
37. Yeh, G.T., Tripathi, V.S.: A model for simulating transport of reactive multispecies components: Model development and demonstration. *Water Resour. Res.* **27**(12), 3075–3094 (1991)
38. Yeh, G.T., Sun, J., Jardine, P., Burgos, W.D., Fang, Y., Li, M.-H., Chunli, Siegel, M.D.: HYDROGEOCHEM 5.0: a three-dimensional model of coupled fluid flow, thermal transport, and HYDROGEOCHEMICAL transport through variably saturated conditions - version 5.0 (2004)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.