Accelerated reactive transport simulations in heterogeneous porous media using Reaktoro and Firedrake

Svetlana Kyas1 · Diego Volpatto2,3 · Martin O. Saar1,4 · Allan M. M. Leal1

Received: 10 November 2020 / Accepted: 12 December 2021 / Published online: 20 January 2022 © The Author(s) 2022

Abstract
This work investigates the performance of the on-demand machine learning (ODML) algorithm introduced in Leal et al. (Transp. Porous Media 133(2), 161–204, 2020) when applied to different reactive transport problems in heterogeneous porous media. This approach was devised to accelerate the computationally expensive geochemical reaction calculations in reactive transport simulations. We demonstrate that even with a strong heterogeneity present, the ODML algorithm speeds up these calculations by one to three orders of magnitude. Such acceleration, in turn, significantly advances the entire reactive transport simulation. The performed numerical experiments are enabled by the novel coupling of two open-source software packages: Reaktoro (Leal 2015) and Firedrake (Rathgeber et al. ACM Trans. Math. Softw. 43(3), 2016). The first library provides the most recent version of the ODML approach for the chemical equilibrium calculations, whereas, the second framework includes the newly implemented conservative Discontinuous Galerkin finite element scheme for the Darcy problem, i.e., the Stabilized Dual Hybrid Mixed (SDHM) method Núñez et al. (Int. J. Model. Simul. Petroleum Industry, 6, 2012).

Keywords Reactive transport modeling · Heterogeneous porous media · Accelerated chemical equilibrium calculations · On-demand machine learning algorithm · Coupling of Reaktoro and Firedrake · Fluid flow and reactive transport coupling

1 Introduction
The challenge in coupling physical and chemical processes lies not only in deriving the working mathematical model but also in solving it numerically. There is an excessive computational load, most of which often results from the geochemical reaction calculations. At the same time, the importance of modeling such coupled processes has significantly increased over the recent years. They have become essential for interpreting the phenomena occurring in the surface or subsurface systems, as well as engineering and environmental problems. Real-life applications include a broad diversity of reactive chemical processes. Among many are rock/mineral alteration in natural diagenetic systems as a result of the carbon capture and storage (CCS), geological sequestration, and chemical stimulation of geothermal reservoirs, injection of acid gases resulting in groundwater contamination, and enhanced oil and gas recovery, transport, and storage of radiogenic and toxic waste products in geological formations (see [91, 93, 102] and references therein). The alterations in fluid and rock composition cause changes in the rock porosity and permeability, fluid density and viscosity, all of these ultimately affecting how fluids, heat, and chemical species are transported.

Depending on the formulated reactive transport model, chemical reaction calculations (i.e., chemical equilibrium
and kinetics) can be responsible for most of the computing costs in the simulation. Achieving a more balanced cost distribution remains a demanding area of research, given the nature of the computations for the complex chemical system. Simulation of the physical processes typically leads to systems of linear equations that follow from the discretized partial differential equations (PDEs) governing physical laws. To simulate chemical processes, in turn, one needs to compute thermodynamic properties of a substantial number of species employing complex equations of state and calculate chemical equilibrium/kinetic states by solving a system of non-linear differential-algebraic equations. These operations must be repeated in all discretization cells of the computational domain, at each reactive transport step.

Simulators for reactive transport simulations include the packages presented in the following list: PHREEQC [80, 81], CORE2D [86], OpenGeoSys (OGS) [10, 44–46], HYTec [98], ORCHESTRA [70], Frachem [7], TOUGHREACT [103], eSTOMP [101], The Geochemist’s Workbench [9], GEM-Selektor and GEMS3K [48, 49], HYDROGEOCHEM [106], PROOST [28], CrunchFlow [92], MIN3P [69], PFLOTRAN [61], and MODFLOW [50] among many others. An exhaustive overview of these packages’ capabilities, along with a list of applications, has been provided in [94]. For more details, the reader is referred to the webpages, user manuals, or publications cited above.

Given the inherent complexity of modeling physical and chemical processes individually, reactive transport codes comprise a combination of specialized packages (i.e., one for the physics and another for the chemistry). Several examples of such coupling are given next: HP1/HPx [40] (based on HYDRUS [89] and PHREEQC), PHT3D (merged from capabilities ofMT3DMS [107] and PHREEQC), COMSOL-PHREEQC iCP [74], COMSOL-PHREEQC [31], COMSOL-GEMS [6], CSMP++GEM [105], DuCOM-Phreeqc [26], GeoSysBRNS [17], Matlab-Phreeqc [73], OGS-Chemapp [59], OGS-GEM [47], OGS-Phreeqc [33], ReactMiCP [29], ReacTran [31], TReacLab [41]. A comprehensive overview of these packages’ advantages and limitations is presented in [28] and [22].

Even though the ODML method contrasts with classical machine learning methods, we acknowledge the body of the literature that reports efforts to apply machine learning to the simulations of porous media. This challenging application has been investigated in the different contexts, i.e., creating high-resolution data from low-resolution images, micro-macro modeling, correlation of the hydraulic conductivity field and the longitudinal macro-dispersivity [108], two-phase upscaling computations, characterization process for the flow simulations in underground fractures [8], particle transport and filtration in porous media [67], and well placement for geological carbon storage. Machine learning methods have significantly penetrated the subject of the uncertainty propagation for multiphase flow, i.e., from the point of view of data-driven approaches based on the labeled data [71, 96, 109] and by the so-called physics-informed neural networks (PINN) methods [27, 84, 110]. We also acknowledge [95] with the references therein for a systematic review of the recent development and applications of domain-aware scientific machine learning in geosystem research. When it comes to solving multiphase flow in heterogeneous porous media, [104] introduces a physics-constrained deep learning model to learn such modeling using a customized U-Net and a surrogate model for predicting well’s flow rate. The ML-based approach performs prediction with a speedup of about 1400 times compared to physics-based simulations. Finally, in [100], the so-called CCSNet was proposed as an alternative to conventional numerical simulators for CCS problems. The authors report that the CCSNet produces all the outputs that a conventional simulator typically provides but about 100 times faster.

Despite that, in all the aforementioned literature, machine learning applications for geochemical speciation and equilibrium calculations are still incipient. All the proposed techniques rely on the training stages with data set generation and configuration and aim at accelerating flow and reactive transport simulations as one. Since the chemical equilibrium is computationally intensive, machine learning methods (aiming to accelerate speciation and equilibrium calculations for reactive flow problems) can provide a substantial improvement to decrease the demanded computational time. In [58], the ODML method is proposed to fill this gap. In contrast with all the proposed techniques employed in cited references above, the ODML algorithm accelerates the geochemical calculations only. It does not require training in advance but rather learns on the fly throughout the simulation, achieving considerable speedups using learned calculations and first-order Taylor approximations.

In this work, we present a novel coupling of the packages Reaktoro [51] and Firedrake [85] for modeling reactive transport processes with the on-demand machine learning (ODML) acceleration strategy [58]. This strategy can substantially speed up the geochemical calculations in reactive transport simulations by orders of magnitude. ODML is able to learn essential chemical equilibrium states during the simulation so that subsequent similar equilibrium states can be calculated considerably faster. The on-demand learning is triggered only when the previously learned equilibrium calculations are unable to predict the result of a new calculation. By detecting key chemical equilibrium calculations, the ODML can quickly predict the vast majority of equilibrium states in a reactive transport simulation instead of fully computing...
them. However, many questions related to the ODML method’s extension to more challenging problems remained out of the scope of [58]. This paper provides a detailed study of the ODML capabilities to substantially accelerate reactive transport simulations with complex chemical systems and/or geologic features, i.e., two-dimensional (2D) porous media with heterogeneity. In particular, due to heterogeneous permeability fields, the presented examples include a substantial change in the chemical composition throughout the domain, making the learning of the ODML more demanding throughout the simulation. We also note that this paper only studies the time complexity of the ODML algorithm (and its decrease), not the spatial complexity (storage and memory).

**Note** Comprehensive evaluations conducted during the learning stage will also be referred to as conventional or full evaluations of chemical equilibrium states throughout the paper. These operations can be performed by either of conventional algorithms, i.e., the Gibbs energy minimization (GEM) approach or the law of mass action (LMA) equations [90]. In contrast, fast approximations will be alternatively referred to as smart estimations/predictions and employ only matrix-vector multiplication.

**Note** The on-demand machine learning (ODML) strategy applied in this paper does not rely on deep learning algorithms. The ODML efficiently stores the results of previous key equilibrium reaction calculations (as well as corresponding sensitivity matrix computed by automatic differentiation) and then employs these collected results for predictions, avoiding thus similar expensive computations at other spatial points and time steps. Thus, the ODML does not use neural networks (e.g., convolutional neural network) nor surrogate models. The ODML exploits, however, some machine learning strategies such as k-nearest-neighbors (kNN) and clustering algorithms, which can be viewed as non-supervised learning schemes. The on-demand clustering strategy within ODML classifies the learned calculations based on essential characteristics of the obtained chemical states (e.g., the major/primary species at that state). For more details, we refer the reader to the original work of [58] on the ODML algorithm.

**Reaktoro** is a computational framework developed in C++ and Python for modeling chemically reactive processes governed by either chemical equilibrium or/and kinetics. For chemical equilibrium calculations, Reaktoro implements numerical methods based on the GEM [52, 54, 57] or the extended LMA (xLMA) formulation [55]. For chemical kinetics calculations, assuming partial chemical equilibrium considerations, Reaktoro relies on the approach presented in [53]. It adopts an implicit time integration scheme for enhanced stability combined with an adaptive time-stepping strategy for efficient simulation of chemical kinetics. The on-demand machine learning (ODML) algorithm for faster chemical equilibrium calculations was introduced recently in Reaktoro [58].

**Firedrake** is an open-source library for solving PDEs with finite element methods (FEM). It is employed here to solve the equations that govern (solute/heat) transport processes (advection/diffusion equations) and fluid flow (i.e., the Darcy equation). The package uses a high-level Unified Form Language (UFL) embedded in Python [2], which provides symbolic representations of variational problems corresponding to PDEs that govern physical laws.

**Note:** It is worth remarking that a similar software coupling is relatively straightforward with another open-source computing platform for solving PDEs, e.g., the FEniCS Project [62, 63]. Examples using such coupling can be found in [22]. Besides Firedrake and the FEniCS Project, Reaktoro has also been coupled with OpenFOAM [42] to produce the pore-scale reactive transport simulator poroReact [79].

This study has the following structure. Section 2 introduces physical and chemical processes considered for the reactive transport simulations and corresponding governing equations. Section 3 provides an overview of the numerical methods used in Reaktoro and Firedrake. Section 4 describes two reactive transport simulations conducted with Reaktoro and Firedrake and discusses the numerical performance of the ODML approach when applying it to 2D heterogeneous problems. In Section 5, we summarize the obtained results, draw conclusions, and discuss future work.

2 Governing equations

This section presents the governing equations for the physical and chemical processes computationally simulated in Section 4. Due to the complexity of each subprocess, our presentation is organized into three parts:

- Section 2.1: single-phase fluid flow in porous media;
- Section 2.2: reactive transport of the fluid species;
- Section 2.3: chemical reactions among fluid species and rock minerals.

2.1 Single-phase fluid flow in porous media

We consider a relatively simple model for fluid flow for the purpose of assessing the potential of the ODML algorithm to accelerate chemical equilibrium calculations in reactive transport simulations. This is because our interest in this study is to evaluate how ODML behaves considering more complex chemical systems and heterogeneous porous media. In what follows, we assume the fluid is incompressible, gravity effects are negligible, and the porous medium...
is isotropic and nondeformable. Given these assumptions, we solve the coupled continuity equation and the Darcy equation below to compute the fluid pressure \( P \) and fluid Darcy velocity \( u \) throughout the medium:

\[
\nabla \cdot (\varrho u) = f \quad \text{in} \quad \Omega \times (0, t_{\text{final}}), \quad (1)
\]

\[
u = -\frac{\kappa}{\mu} \nabla P \quad \text{in} \quad \Omega \times (0, t_{\text{final}}). \quad (2)
\]

Here, \( \varrho \) and \( \mu \) are the fluid density and the dynamic viscosity, \( f \) corresponds to the fluid injection/production rate, \( \kappa = \kappa(x) \), \( x \in \Omega \), represents the (isotropic) permeability field of the porous rock, \( \Omega \) denotes the physical domain, and \( t_{\text{final}} \) indicates the final simulation time.

### 2.2 Reactive transport of the fluid species

Mass conservation applies to the chemical species in the fluid as they advect, diffuse, disperse, and react with rock minerals in a porous medium. We apply the mathematical formulation presented in Appendix 4 of [58] for the reactive transport of fluid species in terms of mass conservation of chemical elements and electric charge. This approach is a standard procedure in the literature that substantially reduces the number of PDEs to be solved for transport phenomena \([60]\). The formulation also accounts for the dissolution and precipitation of rock minerals. For convenience, we write below the governing equation:

\[
\frac{\partial (b_i^f + b_i^s)}{\partial t} + \nabla \cdot (v b_i^f - D \nabla b_i^f) = 0 \quad (i = 1, \ldots, E), \quad (3)
\]

where \( b_i^f \) and \( b_i^s \) are the bulk concentrations (mol/m³) of elements (chemical elements and charge) in the fluid and the solid partitions, respectively. Here, \( v \) is the mean percolation velocity (in each cell) related to the Darcy velocity through the porosity \( \phi \), i.e., \( v = u/\phi \), and \( D \) is the dispersion-diffusion tensor \([82]\), i.e.,

\[
D = (\alpha_{\text{mol}} + \alpha_{|u|}) I + \frac{\alpha_l - \alpha_t}{|u|} u \otimes u, \quad (4)
\]

where \( \alpha_{\text{mol}} \) is the molecular diffusion coefficient as well as \( \alpha_l \) and \( \alpha_t \) are the longitudinal and the transversal dispersion coefficients, respectively.

**Note:** We assume \( \alpha_l \) and \( \alpha_t \) be both zero so that \( D = \alpha_{\text{mol}} \), which suffices for the numerical investigations of the ODML approach performance in Section 4.

Equation 3 possesses two advantages when compared to the transport equation formulated for chemical species:

- Absence of the reaction term in the convection-diffusion equation, leaving such concerns to a separate chemical kinetic/equilibrium solver and easing the coupling procedure.

- A considerable decrease in the number of unknowns, as the number of elements \( E \) is typically considerably smaller than the number of species \( N \) (see Eq. 3).

**Note:** For more than one fluid phase (each with its velocity field) and different diffusion coefficients for the fluid species, this simplified transport formulation becomes less straightforward.

### 2.3 Chemical reactions among fluid species and rock minerals

In our simulations, homogeneous and heterogeneous chemical reactions among the species are considered. We adopt a local chemical equilibrium model so that both fluid species and rock minerals are in chemical equilibrium at any point in space and time. Because of transport processes and variations in temperature/pressure (when applicable), the chemical equilibrium states are continually altered at each point of the domain. For example, a rock mineral may gradually dissolve as the more acidic fluid passes through that point in space.

In this manner, at every discretized point in space, we solve the Gibbs energy minimization problem:

\[
\min_n G(T, P, n) \quad \text{subject to} \quad \begin{cases}
An = b \\
n \geq 0
\end{cases}, \quad (5)
\]

to compute the chemical equilibrium amounts \( n = (n_1, \ldots, n_N) \) of the species distributed among all fluid and solid phases in the chemical system. Vector \( n \) includes the aqueous species amounts (solute and solvent water) and the mineral amounts composing the porous rock. **Note:** The given input of this chemical equilibrium problem is composed of the temperature \( T \), pressure \( P \), and the amounts of chemical elements \( b = (b_1, \ldots, b_E) \) (including electric charge), whereas the constraints include the mass balance \( An = b \) and non-negativity of the species amounts \( n \geq 0 \). In (5), \( T \) is assumed to be constant throughout the medium, \( P \) computed via the solution of the continuity and Darcy equations, and \( b \) is updated over time via (3). For more information on solving the GEM problem, we refer to \([57, 58]\).

### 3 Numerical methods

In what follows, we consider the numerical methods required to solve the governing equations presented earlier. To solve the fluid flow through a heterogeneous porous medium, we use a highly mass conservative finite element method (FEM) to accurately capture the velocity field. The transport equations are solved with a suitable FEM that handles advection-dominated flow. For the multiphase...
chemical equilibrium calculations, involving the fluid and solid species, we employ the Gibbs energy minimization algorithm (GEM) [54, 57]. In addition to these numerical methods, we provide a brief review of the on-demand machine learning algorithm (ODML) presented in [58]. It is applied to accelerate several millions of time-consuming equilibrium calculations in the reactive transport simulation course. This total number of calculations results from the necessity to compute equilibrium states at each mesh point (or degree of freedom (DOF) in the finite element naming convention) during each time step.

### 3.1 Time staggered operator splitting steps

To solve the time-dependent reactive transport equations in Eq. 3, we apply the fully discrete formulation typically resulting from a combination of the finite difference approximation in time with the finite element approach in space. We use index $k$ to denote the time-step and $\Delta t = t^{k+1} - t^k$ to determine the time-step size. We assume the uniform discretization $I_M := \{0 = t_0 < t_1 < \ldots < t_K = t_{\text{final}}\}$ of the time interval $[0, t_{\text{final}}]$, where $t_{\text{final}} > 0$ is the total time. The operator splitting scheme using classical sequential non-iterative approach (SNIA) for solving these transport equations is described below:

**I.** Consider (3) assuming $\frac{\partial b_i}{\partial t} = 0$ to calculate an intermediate concentration $\tilde{b}_i^t$. We apply the backward Euler scheme in time to compute an auxiliary element concentrations approximation $\tilde{b}_i^{t,k+1} = \tilde{b}_i^t(t_{k+1})$, $i = 1, \ldots, E$, in the fluid partition:

$$\frac{\tilde{b}_i^{t,k+1} - \tilde{b}_i^{t,k}}{\Delta t} + \nabla \cdot (\rho \tilde{b}_i^{t,k+1} - D \nabla \tilde{b}_i^{t,k+1}) = 0 \quad \text{in} \quad \Omega. \tag{6}$$

We assume the flux boundary condition on the inlet face of boundary $\Gamma_{\text{inlet}} \subset \partial \Omega$, where we inject the brine,

$$-(\rho \tilde{b}_i^{t,k+1} - D \nabla \tilde{b}_i^{t,k+1}) \cdot \mathbf{n}_{\text{inlet}} = u \hat{b}_i_{\text{inlet}} \quad \text{on} \quad \Gamma_{\text{inlet}}$$

and zero flux on the top and bottom of the boundary, $\Gamma_{\text{top}} \subset \Gamma \equiv \partial \Omega$,

$$-(\rho \tilde{b}_i^{t,k+1} - D \nabla \tilde{b}_i^{t,k+1}) \cdot \mathbf{n} = 0 \quad \text{on} \quad \Gamma_{\text{top}} \cup \Gamma_{\text{bottom}}.$$

The right boundary is considered a free outflow boundary. Vector $\mathbf{n}$ denotes the outward-pointing normal to the boundary face $\Gamma$ ($\mathbf{n}_{\text{inlet}}$ is the normal on $\Gamma_{\text{inlet}}$), and $\hat{b}_i_{\text{inlet}}$ (in mol/m$^3$) represents the given concentration of the injected brine. As a space discretization solver for Eq. 6, we employ the Streamline-Upwind Petrov-Galerkin (SUPG) scheme introduced in [13] (see Section 3.3) to handle advection-dominated transport in an accurate and stable way.

The velocity in Eq. 6 is generated from the coupling to the Darcy problem, i.e., $\mathbf{u} = \mathbf{u}^k$, where $\mathbf{u}^k$ satisfies the system

$$\nabla \cdot (\rho \mathbf{u}^k) = \mathbf{j}^k \quad \text{in} \quad \Omega \times (0, t_{\text{final}}),$$

$$\mathbf{u}^k = -\frac{1}{\mu} \nabla P^k \quad \text{in} \quad \Omega \times (0, t_{\text{final}}). \tag{7}$$

A comprehensive numerical analysis, demonstrating the existence and uniqueness of the solution for the prior semi-discrete system, can be found in [65, 66].

**II.** Evaluate the total element concentrations $b_i$, using prior reconstructed $\tilde{b}_i^{t,k+1}$ and $b_i^k$ (assumed constant):

$$b_i^{k+1} = \tilde{b}_i^{t,k+1} + b_i^k.$$  

**III.** In each space discretization cell, calculate $n^{k+1}$ for given $T$, $P$, and $b^{k+1}$ using the smart chemical equilibrium algorithm accelerated with the ODML strategy (see Section 3.4).

To make sure that the Courant–Friedrichs–Lewy (CFL) condition is satisfied, we assume CFL = 0.3 and the time step is defined by

$$\Delta t = \frac{\text{CFL}}{\max \left\{ \max |v_x|/\Delta x, \max |v_y|/\Delta y \right\}}, \tag{8}$$

where $v = [v_x; v_y]^T$, and $\Delta x$ and $\Delta y$ are the lengths of the cells along the $x$ and $y$ coordinates, respectively.

**Note:** As it is mentioned above, we use the sequential non-iterative approach (SNIA) to couple transport and chemical codes. Carayrou et al. [14] provides the comparison of different coupling schemes for a reactive transport benchmark, where the SNIA proves to be an efficient scheme with a reasonable error magnitude. According to [16], the SNIA approach works well for problems with advection and has a good balance between accuracy, computational time, and transparency of implementation. These results provide us with the confidence to use the SNIA in more realistic reservoir simulation cases. Future works on extending Reaktoro and Firedrake coupling will use two-way coupling to ensure the proper handling of the mass conservation. It would be interesting to study the transport and chemical equilibrium coupling performance for the famous benchmarks. However, it lies beyond the scope of Reaktoro package, which focuses mainly on modeling realistic chemical systems with corresponding thermodynamical databases.

### 3.2 SDHM method for the fluid flow in porous media

In the following, we describe in short the relevant part of the finite element method (FEM) applied in the present...
work. We use a conservative FEM to obtain accurate velocity fields that satisfy mass conservation, a critical numerical feature for transport problems in heterogeneous media. The formulation is based on stabilized mixed finite element methods [12, 21, 68] combined with hybridization techniques [18, 19]. The resulting Stabilized Dual Hybrid Mixed (SDHM) method [76] has all Discontinuous Galerkin (DG) desirable features while requiring fewer degrees of freedom (DOF) due to the static condensation procedure. The discretized global system is solved for Lagrange multipliers only (defined on the mesh skeleton). The solution for pressure and velocity fields, in turn, is recovered by the element-wise post-processing of these multipliers’ solution.

To present the SDHM formulation used in this work, we introduce the classical $L^2$ inner products for an arbitrary regular domain $\Omega \subset \mathbb{R}^n$ ($n = 2, 3$) as:

$$(p, q)_\Omega := \int_\Omega pq \, dx, \quad (p, q)_{\partial \Omega} := \int_{\partial \Omega} pq \, ds,$$

$$(z, w)_\Omega := \int_\Omega z \cdot w \, dx, \quad (z, w)_{\partial \Omega} := \int_{\partial \Omega} z \cdot w \, ds,$$

in which $p$ and $q$ represent scalar functions and $z$ and $w$ vector-valued functions defined on $\Omega$.

To derive the weak formulation of the Darcy problem (2) in the heterogeneous medium, we follow the lines of [78]. Let $\mathcal{H}_h := \{\mathcal{K}\}$ represent a regular tessellation of the domain $\Omega$, such that $\Omega = \bigcup \mathcal{K}$. Let $\mathcal{E}_h := \{e : e \subset \partial K, \mathcal{K} \in \mathcal{H}_h\}$ denote the set of edges, where $e$ is the edge of the element $\mathcal{K}$. Let $\mathcal{E}_h^I \subset \mathcal{E}_h$ denote a subset of interior edges. Thus, we obtain the following weak form of Eq. 2:

$$(\mu \kappa^{-1} u, w)_\mathcal{K} - (p \cdot w)_\mathcal{K} + \int_{\partial \mathcal{K}} \hat{p} (w \cdot n_\mathcal{K}) \, ds = 0, \quad \forall w \in W_\mathcal{K},$$

$${-}(\nabla \cdot (\rho u), q)_\mathcal{K} + (f, q)_\mathcal{K} = 0, \quad \forall q \in Q_\mathcal{K},$$

where

$$W_\mathcal{K} := \left\{ w \in [L^2(\mathcal{K})]^2 : \nabla \cdot w \in L^2(\mathcal{K}), \forall \mathcal{K} \in \mathcal{H}_h \right\}$$

and

$$Q_\mathcal{K} := \left\{ q \in L^2(\mathcal{K}), \forall \mathcal{K} \in \mathcal{H}_h \right\}$$

stand for the local functional spaces, $n_\mathcal{K}$ denotes the outward pointing normal to $\partial \mathcal{K}$, and $\hat{p}$ is the trace of pressure on $\Omega$. We assume that $\kappa$ is at least invertible element-wise.

Then, the weak formulation on each $\mathcal{K} \in \mathcal{H}_h$ reads as follows: find $(u, p) \in W_\mathcal{K} \times Q_\mathcal{K}$, such that

$$(\mu \kappa^{-1} u, w)_\mathcal{K} - (p \cdot w)_\mathcal{K} - (\nabla \cdot (\rho u), q)_\mathcal{K}$$

$$+ \langle \hat{p}, w \cdot n_\mathcal{K} \rangle_{\partial \mathcal{K}} = -(f, q)_\mathcal{K}, \quad \forall (w, q) \in W_\mathcal{K} \times Q_\mathcal{K}.$$ 

Utilizing the notation and formulation from [4], to obtain the pressure trace we solve the global dual hybrid mixed problem: find $u \in W := \prod_{\mathcal{K}} W_\mathcal{K}, p \in Q := \prod_{\mathcal{K}} Q_\mathcal{K}$, and $\lambda \in \mathcal{M} := \{\mu \in L^2(e), \ e \in \mathcal{E}_h\}$, such that

$$\sum_{\mathcal{K} \in \mathcal{H}_h} \left[ (\mu \kappa^{-1} u, w)_{\partial \mathcal{K}} - (p \cdot w)_{\partial \mathcal{K}} + \langle \lambda, w \cdot n_\mathcal{K} \rangle_{\partial \mathcal{K}} + (p_D, w \cdot n_\mathcal{K})_{\partial \mathcal{K} \in \Gamma_D} \right] = 0, \quad \forall w \in W,$$

$$\sum_{\mathcal{K} \in \mathcal{H}_h} \left[ -(\nabla \cdot (\rho u), q)_{\partial \mathcal{K}} + (f, q)_{\partial \mathcal{K}} \right] = 0, \quad \forall q \in Q,$$

$$\sum_{\mathcal{K} \in \mathcal{H}_h} \left[ (\mu, u \cdot n_\mathcal{K})_{\partial \mathcal{K}} - \langle \mu, g_N \rangle_{\partial \mathcal{K} \in \Gamma_N} \right] = 0, \quad \forall \mu \in \mathcal{M},$$

in which $\Gamma_D \subset \partial \Omega$ and $\Gamma_N \subset \partial \Omega$ ($\Gamma_D \bigcup \Gamma_N = \partial \Omega$ and $\Gamma_D \bigcap \Gamma_N = \emptyset$) are the sub-domains for pressure and flux boundary conditions, respectively.

In Eq. 9, each Lagrange multiplier $\lambda$ corresponds to the pressure trace on $K \in \mathcal{H}_h$ satisfying the decomposition $\hat{p} = \lambda + p_D$, i.e., $\lambda$ is solved for internal edges and $p_D$ is prescribed on boundary edges [19, 75]. The third equation of Eq. 9, known as transmission condition [19] or flux continuity of the Darcy velocity field normal component. Note that flux boundary conditions are also weakly imposed through the transmission condition.

To stabilize (9) and obtain (locally) adjoint consistent formulation, we add respective stabilization terms, i.e.,

$$r_{MB}(u, w) := \delta_1 (\|\kappa^{-1}\|_\infty (\nabla \cdot (\rho u) - f), \nabla \cdot (\rho w))_{\mathcal{K}},$$

$$r_D(u, w, p) := \delta_2 (\mu \kappa^{-1} u + \nabla p, w)_{\mathcal{K}},$$

$$r_{\nabla \cdot D}(u, w) := \delta_3 (\|\kappa\|_\infty \nabla \times (\mu \kappa^{-1} u), \nabla \times \mu \kappa^{-1} w)_{\mathcal{K}},$$

$$r_\lambda(\mu, \lambda, p) := \|\kappa\|_\infty \beta (\mu, \lambda - p)_{\partial \mathcal{K}}.$$

These least square residual forms correspond the mass balance, the Darcy, and the curl of Darcy’s laws [20] together with the Lagrange multiplier introduced in [5].

Springer
The weighting terms \( \delta_i \) \((i = 1, 2, \text{ or } 3)\) are based on [21, 64], and [76]. However, we modify \( \delta_i \) to take into account local mesh-size \( h \), following the ideas of [68], when considering the mass balance residual term among stabilization mechanisms. Thus, we apply the mesh-dependent weighted Least-Squares [11] terms with \( \delta_i = -1/2 \) and \( \delta_2 = \delta_3 = h^2/2 \).

Then, the SDHM formulation reads as: find \( u \in W := \prod_{\mathcal{K}} H^1(\mathcal{K}) \times H^1(\mathcal{K}), q \in Q := \prod_{\mathcal{K}} H^1(\mathcal{K}), \) and \( \lambda \in \mathcal{M}, \) such that

\[
\begin{align*}
\sum_{\mathcal{K} \in \mathcal{T}_h} & \left[ (\mu \kappa^{-1} u, w)_{\mathcal{K}} - (p, \nabla \cdot w)_{\mathcal{K}} + (\lambda, w \cdot n)_{\partial \mathcal{K}} + (p D, w \cdot n)_{\partial \mathcal{K}} \right]_{\partial \mathcal{K} \in \Gamma_D} + r_M(u, w) - r_D(u, w, p) + r_G(u, w) \\
\sum_{\mathcal{K} \in \mathcal{T}_h} & - (\nabla \cdot (q u), q)_{\mathcal{K}} + (f, q)_{\mathcal{K}} + r_h(u, \nabla q, p) - r_\lambda(\lambda, q, p) \\
\sum_{\mathcal{K} \in \mathcal{T}_h} & (\mu, u \cdot n)_{\partial \mathcal{K}} - (\mu, g N)_{\partial \mathcal{K} \in \Gamma_N} + r_\mu(\lambda, \mu, p) \\
& = 0, \quad \forall w \in W, \\
& = 0, \quad \forall q \in Q, \\
& = 0, \quad \forall \mu \in \mathcal{M},
\end{align*}
\]

where \( \nabla \times \) is the curl operator and \( \beta := \frac{\beta_0}{h} \geq 0 \) is the stabilization factor with \( \beta_0 \in \mathbb{R} \).

The first and second equations in Eq. 10 generate the two local problems and the third the global problem. Subsequently, we can approximate \( u, p, \) and \( \lambda \) by approximations from the broken function spaces, i.e., \( u_h \in W_h^m := \left\{ w_h \in \mathcal{W} : w_h|_{\mathcal{K}} \in \mathbb{P}^m \times \mathbb{P}^m, \mathcal{K} \in \mathcal{T}_h \right\}, \) \( p_h \in Q_h^l := \left\{ q_h \in \mathcal{Q} : q_h|_{\mathcal{K}} \in \mathbb{P}^l, \forall \mathcal{K} \in \mathcal{T}_h \right\}, \) \( \lambda_h \in \mathcal{M}_h^k := \left\{ \mu_h \in \mathcal{M} : \mu_h|_e \in \mathbb{P}^k, \forall e \in \mathcal{E}_h \right\} \), where \( \mathcal{W}_h^m, \mathcal{Q}_h^l, \mathcal{M}_h^k \) are discontinuous Lagrangian finite element spaces. Here, \( \mathbb{P}^k \) is the polynomial set with \( m < k \) if \( \mathcal{K} \) is a triangular element, or \( m = k \) if \( e \) represents an edge, or \( m \leq k \) in each Cartesian variable if \( \mathcal{K} \) is represented by a quadrilateral element (here, \( k = l, m, \) or \( s \)). The SDHM method is known with its consistency and flexibility in the approximation spaces choice. It also provides optimal rates of error convergence [76], stable for any value \( \beta \), including \( \beta = 0 \) [77], and is locally conservative for \( l = m = s \). The multiplier choice as the pressure trace is essential to ensure that Eq. 10 is solvable for \( (u_h, p_h) \in W_h^m \times Q_h^l \) as functions of \( \lambda_h \in \mathcal{M}_h^k \). Moreover, it results in a computationally efficient way to solve a global system with only the trace variable as unknown due to static condensation procedure.

### 3.3 SUPG method for the semi-discrete element-based transport problem

To obtain the approximation of the semi-discrete transport problem, we use the Streamline Upwind Petrov-Galerkin (SUPG) method. Usually, it is applied to advection-dominated partial differential equations to suppress numerical oscillations present in the classical Petrov-Galerkin method for this class of a problem [13]. Discretizing (3) in time yields the implicit time-stepping scheme: for \( k = 1, \ldots, K - 1 \)

\[
\begin{align*}
\frac{b^{k+1} - b^k}{\Delta t} + \nabla \cdot (u b^{k+1} - D \nabla b^{k+1}) &= 0 \quad \text{in} \; \Omega, \\
- (u b^{k+1} - D \nabla b^{k+1}) \cdot n_{\text{inlet}} &= u b_{\text{inlet}} \quad \text{on} \; \Gamma_{\text{inlet}}, \\
- (u b^{k+1} - D \nabla b^{k+1}) \cdot n_{\text{bottom}} &= 0 \quad \text{on} \; \Gamma_{\text{top}} \cup \Gamma_{\text{bottom}}.
\end{align*}
\]

To concentrate on the SUPG approximation scheme, we omit the index \( f \) indicating the amount of elements in the fluid species and the index \( j = 1, \ldots, E \), numbering the elements. Then, the fully discrete approximation of Eq. 11 reads as: find \( b^{k+1} \in X_h^p \) satisfying the bilinear form

\[
B(b^{k+1}, \eta) = F(b^{k+1}, \eta), \quad \forall \eta \in X_h^p,
\]

where

\[
B(b^{k+1}, \eta) := (b^{k+1} + \Delta t (v \cdot \nabla b^{k+1}, \eta) + \Delta t (D \nabla b^{k+1}, \nabla \eta)_{\mathcal{K}} + S(b^{k+1}, \eta),
\]

\[
F(b^{k+1}, \eta) := (b^k, \eta) + (u b_{\text{inlet}}, \eta)_{\Gamma_{\text{inlet}}}.
\]

We omit the sub-index of domain \( \Omega \) in some of the \( L_2 \) inner-products, i.e., \( (u, v)_{\Omega} \equiv (u, v) \), and assume the approximation space \( X_h^p \) to be a continuous Lagrangian finite element space of degree \( p \geq 1 \). The stabilizing term \( S(b^{k+1}, \eta) \) to the standard Petrov-Galerkin variational formulation is defined as

\[
S(b^{k+1}, \eta) := \sum_{\mathcal{K}} \tau_{\mathcal{K}} \left( (b^{k+1} - b^k) + \Delta t u \cdot \nabla b^{k+1} - \Delta t \nabla \cdot (D \nabla b^{k+1}, \nabla \eta)_{\mathcal{K}} - \sum_{\partial \mathcal{K} \subseteq \Gamma_{\text{inlet}}} \tau_{\mathcal{K}} (u b_{\text{inlet}}, \nabla \eta)_{\partial \mathcal{K} \subseteq \Gamma_{\text{inlet}}}
\]

\[\square\]
with the stabilization parameter
\[
\tau_{\mathcal{K}} := \begin{cases} \frac{h_{\mathcal{K}}}{2\|v\|_{L^\infty(\mathcal{K})}}, & \text{Pe}_\mathcal{K} \geq 1, \\ 0, & 0 < \text{Pe}_\mathcal{K} < 1. \end{cases}
\]

It is dependent of the Peclet number \(\text{Pe}_{\mathcal{K}} := \frac{m_{\mathcal{K}}}{D_{\mathcal{K}}^+} \) with \(m_{\mathcal{K}} := \alpha_{\text{mol}} + \alpha_t \inf_{x \in \mathcal{K}} |v(x)|\), \(D_{\mathcal{K}} := (2(\alpha_{\text{mol}} + \alpha_t)\|v\|_{L^\infty(\mathcal{K})})^2 + 2(3\alpha_t - 2\alpha_{\text{mol}})^2\|v\|_{L^\infty(\mathcal{K})}^2 c_{\text{inv}} \)\(^{1/2}\), where \(c_{\text{inv}}\) stands for the inverse constant of finite element spaces.

### 3.4 ODML method for faster chemical equilibrium calculations

Finally, we briefly describe the **smart chemical equilibrium calculations method** introduced in [58], which combines a classical algorithm for chemical equilibrium with an *on-demand machine learning* (ODML) strategy that accelerates the calculations by one to three orders of magnitude. Let the process of solving (5) (i.e., the problem of computing the chemical equilibrium state) be represented in the abstract functional notation:

\[
y = f(x),
\]

where \(f\) represents the function that performs the necessary steps towards a solution, and \(x\) and \(y\) are the given and result vectors defined as

\[
x = \begin{bmatrix} T \\ P \\ b \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} n \\ \mu \end{bmatrix}.
\]

Here, \(x\) comprises given triple \((T, P, b)\) introduced in Section 2.3. Besides \(n\), the output vector \(y\) also contains the chemical potentials \(\mu = (\mu_1, \ldots, \mu_N)\) of the species. Put differently, \(y\) contains the information on the final speciation of the chemical system and thermochemical properties at that final state. Assume that \(f\) was evaluated previously with input conditions \(x\), and \((\hat{y}, \hat{y}_x)\) is stored for the future use, where \(\hat{y} = f(\hat{x})\) and \(\hat{y}_x := \partial f / \partial \hat{x}\) denote the values and Jacobian (or sensitivity) matrix of \(f\). Such sensitivities can be utilized to predict new chemical equilibrium states quickly and relatively accurately in the vicinity of \((\hat{x}, \hat{y}, \hat{y}_x)\). In particular, we use a first-order Taylor extrapolation

\[
\tilde{y} = \hat{y} + \hat{y}_x(x - \hat{x}),
\]

where \(\tilde{y}\) serves as an estimate of \(y\). In another words, we can use \(\hat{y}_x\) to estimate how the species amounts in the final state would change when small perturbations are applied to \(T\), \(P\), and \(b\). Note: Computing this sensitivity matrix efficiently and accurately is far from trivial, and we can accomplish this via the use of automatic differentiation [56].

Once the predicted output \(\tilde{y}\) is calculated, it must be verified for acceptance. For this, we introduce a function \(g(\tilde{y}, \bar{y}) < \epsilon\) that assess whether \(\tilde{y}\) provide a sufficiently accurate approximation of the exact output \(y = f(x)\) with a preselected tolerance \(\epsilon > 0\). The acceptance test function \(g(\tilde{y}, \bar{y})\) (which can vary across applications of the ODML algorithm) resolves to either an acceptance or rejection and does not require the evaluation of a computationally expensive function \(f\). For more details on the acceptance criterion and how the reference elements are stored in priority-based clusters, we refer the reader to [58].

### 4 Results

This section investigates the performance of the *on-demand machine learning* (ODML) strategy when applied to accelerate relatively complex reactive transport simulations. For a more in-depth method description, we refer to [58] and its implementation in the open-source software Reaktoro [51]. This section aims to demonstrate that the ODML enables accelerated reactive transport simulations without reducing accuracy. It also demonstrates the intrinsic capability of ODML in conserving electric charge and mass of chemical elements.

Allan et al. [58] demonstrates the algorithmic and computing features of the ODML method using relatively simple one-dimensional reactive transport simulations. In this work, we consider two reactive transport problems with more complex chemical and geologic conditions in the simulations. The first problem models the dolomitization phenomenon in a rock column, similar to the one discussed in [58]. We undertake this test deliberately to enable comparing the numerical results and obtained computation speedups. The second problem addresses the H$_2$S-scavenging of a siderite-bearing reservoir, particularly essential for the oil and gas industry.

**Activity models and thermodynamic data.** To evaluate activity coefficients of aqueous species, we use the Pitzer [32, 83], the Debye-Hückel (DH) [24], and the Helgeson-Kirkham-Flowers (HKF) [35–38] activity models. The activity of solute species CO$_2$(aq) is computed using the model of Drummond [25]. To calculate the standard chemical potentials of the species, we utilize the equations of state derived in [34, 35, 87, 97], and [88]. Finally, the density of water (with temperature and pressure derivatives) is computed with the [99] model. Two database files are used to obtain corresponding parameters for thermodynamic calculations, i.e., model parameters for evaluation of the species standard thermodynamic properties for temperatures 0-1000°C and pressures 1-5000 bar. In particular, for the dolomitization modeling discussed in Section 4.1, we use the supcrt98.xml database of Reaktoro, whereas for the scavenging example in Section 4.2, the supcrt07.xml database is utilized. Both files
were generated from the SUPCRT92 [43] database files slop98.dat and slop07.dat, respectively.

Numerical methods and other setup details. For the numerical investigations presented further in this section, the Darcy velocity and the pressure approximations in Eq. 7 are calculated employing the SDHM method [76]. Instead of updating them at each simulation step, the pair \( \langle u, p \rangle \) is reconstructed only once at the beginning of the reactive transport simulation due to insignificant porosity changes. Furthermore, this work’s primal goal is to evaluate the ODM algorithm performance under more challenging conditions. We assume zero source-term in Eq. 7, whereas \( \varrho = 1000.0 \text{ kg/m}^3 \) and \( \mu = 8.9 \cdot 10^{-4} \text{ Pa \cdot s} \). For the pressure, \( p_{\text{inlet}} = P_{\text{inlet}} \) and \( p_{\text{outlet}} = 0.9 p_{\text{inlet}} \) on the left and right boundary of the rock, respectively. The rock’s heterogeneous permeability (with preferential flow path) was obtained using the open-source Python package GeoStatTools [72].

4.1 Case I: Reactive transport of dolomitization process

Numerical model setup. The first heterogeneous reactive transport model is illustrated in Fig. 1. The vertical and horizontal lengths of the rock are set to 1.6 m and 1.0 m, respectively. By discretizing both dimensions with 100 cells (or 101 FEM nodes), we obtain 10201 degrees of freedom (DOFs). At each DOF, we monitor the system’s entire concentrations of the species, and thermochemical properties (or 101 FEM nodes), we obtain 10201 degrees of freedom (DOFs). At each DOF, we monitor the system’s entire equilibrium with calcite and quartz. The resulting pH of the initial state is 9.2. The boundary condition is defined by the rock plate having initial 10% porosity and composed with 98% vol of quartz (CaSiO

\( = \rho \cdot \mu \cdot \varphi \)) = 10\text{ bar}. The heterogeneous permeability of the rock is presented in Fig. 2a next to the pore velocity in Fig. 2b reconstructed as the numerical solution of Eq. 7. The diffusion coefficient is fixed to \( D = 10^{-9} \text{ m}^2/\text{s} \) for all fluid species. We note that the setup of this numerical example is selected for illustration purposes only. The natural process (on the geological time scale) is more complex than the thermodynamic equilibrium approach described in the paper, due to kinetic/nucleation hurdles.

To illustrate the complexity of the reaction network modeled by the considered chemical system, we list the most important reactions below:

\[
\begin{align*}
\text{NaCl(aq)} & \rightleftharpoons \text{Na}^+ + \text{Cl}^- & \log K = 1.570 \\
\text{H}_2\text{O(l)} & \rightleftharpoons \text{H}^+ + \text{OH}^- & \log K = -13.9951 \\
\text{CO}_2(\text{aq}) + \text{H}_2\text{O(l)} & \rightleftharpoons \text{H}^+ + \text{HCO}_3^- & \log K = 16.681 \\
\text{CO}_3^{2-} + \text{H}^+ & \rightleftharpoons \text{HCO}_3^- & \log K = -10.329 \\
\text{CaCO}_3(\text{aq}) & \rightleftharpoons \text{CO}_2^{aq} + \text{Ca}^{2+} & \log K = 8.48 \\
\text{MgCO}_3(\text{aq}) & \rightleftharpoons \text{CO}_2^{aq} + \text{Mg}^{2+} & \log K = 2.98 \\
\text{CaHCO}_3^- & \rightleftharpoons \text{H}^+ + \text{CO}_3^{2-} + \text{Ca}^{2+} & \log K = 11.435 \\
\text{MgHCO}_3^- & \rightleftharpoons \text{H}^+ + \text{CO}_3^{2-} + \text{Mg}^{2+} & \log K = 11.399 \\
\text{CaCO}_3(\text{s,calcite}) & \equiv \text{Ca}^{2+} + \text{CO}_3^{2-} & \log K = 8.48 \\
\text{(CaMg)}(\text{CO}_3)_2(\text{s,dolomite}) & \equiv \text{Ca}^{2+} + \text{Mg}^{2+} + 2\text{CO}_3^{2-} & \log K = 17.09
\end{align*}
\]

As one can see, the \( \log K \) constant for the system reaction ranges from -14 till 17 approximately causing the 21-order magnitude change in the equilibrium constant. Note, however, that the GEM algorithm for chemical equilibrium
Table 1 Recapitulation of the dolomitization example parameters

| Notation | Value | Specification |
|----------|-------|---------------|
| **Thermodynamic Conditions** | | |
| $T$ | 60 °C | temperature |
| $P$ | 100 bar | pressure |
| **Physical Properties** | | |
| $D$ | $10^{-9}$ m$^2$/s | diffusion coefficient |
| **Discretization Parameters** | | |
| $\Delta x$ | 1.6 cm (0.016 m) | spatial mesh-size along the x-axis |
| $\Delta y$ | 1.0 cm (0.01 m) | spatial mesh-size along the y-axis |
| $N_{dofs}$ | 10201 | number of degrees of freedom (DOFs) |
| $\Delta t$ | CFL/$\max\{|v_x|/\Delta x, \ max\{|v_y|/\Delta y\}$ | temporal step size |
| **Initial Condition** | | |
| $\phi$ | 10% | initial porosity |
| **Rock composition** | | |
| SiO$_2$ | 98% vol | quartz |
| CaCO$_3$ | 2% vol | calcite |
| **Resident brine composition (NaCl-brine)** | | |
| NaCl | 0.70 molal | sodium chloride |
| **Boundary Condition** | | |
| **Injected brine composition (NaCl-MgCl$_2$-CaCl$_2$-brine saturated with CO$_2$)** | | |
| NaCl | 0.90 molal | sodium chloride |
| MgCl$_2$ | 0.05 molal | magnesium chloride |
| CaCl$_2$ | 0.01 molal | calcium chloride |
| CO$_2$ | 0.75 molal | carbon dioxide |

computations used in this work does not need these reactions nor their equilibrium constants.

**Accuracy of generated chemical fields.** Figure 3 compares two-dimensional chemical fields generated by the reactive transport simulation employing two approaches for the chemical equilibrium calculation. On the left, we collect the results generated by the conventional approach, and the right column of plots corresponds to the ODML method. To be specific, it shows the time steps 500, 1500, and 2500 (corresponding to 0.48, 1.43, and 2.38 days of simulations) with the amounts of minerals measured in mol/m$^3$. Dissolving CaCO$_3$ releases $\text{Ca}^{2+}$, $\text{CO}_2^{3-}$, and $\text{HCO}_3^-$ ions. They, in turn, react with the ions of $\text{Mg}^{2+}$(aq), injected by the brine, and precipitate CaMg($\text{CO}_3$)$_2$. After 500 time steps of the brine injection, we observe the dissolution of calcite (in blue) and simultaneous precipitation of dolomite (in orange). Here, some parts of the rock retain pure quartz (in gray), where dolomite is gradually dissolved away due to the continuous injection of the acidic brine. Figure 3c and d, corresponding to 1500 time steps of simulations, illustrate the fields with a preferential path forming in the parts of the rock with higher permeability and, as a result, larger amplitude velocities. Eventually, after 2.38 days, practically all calcite is replaced by the dolomite (see Fig. 3e and f). The fields generated by the ODML approach (on the right) are rather close to those on the left, demonstrating high accuracy of the method even when applied to heterogeneous rocks.

Figure 4 illustrates the evolution of selected aqueous species after 27.42 minutes of simulations (or 20 time steps). We observe the local increase in all species

Table 2 Chemical system characteristics in the dolomitization example

| Elements (9 in total) | C, Ca, Cl, H, Mg, Na, O, Si, Z$^*$ |
|----------------------|----------------------------------|
| Phases (4 in total)  | Aqueous, Calcite, Dolomite, Quartz |
| Species (33 aqueous, 3 mineral) | CO$_2$(aq), CaCl$^+$ (aq), ClO$_2^-$ (aq), H$_2$O(l), HClO$_2$(aq), MgCl$^+$ (aq), O$_2$(aq), MgOH$^+$ (aq), OH$^-$ (aq), CaCO$_3$ (calcite), CaMg(CO$_3$)$_2$ (dolomite) |

$^*$ Z represents an electric charge
Fig. 2  (a) The permeability field and (b) the Darcy velocity field in the dolomitization example

Fig. 3  The amount of calcite (blue) and dolomite (orange) measured in mol/m$^3$ in the two-dimensional rock core at time steps 500, 1500, and 2500, which correspond to 0.48, 1.43, and 2.38 days of simulations. The plots on the left correspond to the reference/benchmark chemical fields obtained using full GEM for chemical equilibrium calculation. The plots on the right are the results generated during the same simulation but utilizing the ODML algorithm with $\varepsilon = 0.001$. 

(a) step 500 with the conventional algorithm  
(b) step 500 with the ODML algorithm  
(c) step 1500 with the conventional algorithm  
(d) step 1500 with the ODML algorithm  
(e) step 2500 with the conventional algorithm  
(f) step 2500 with the ODML algorithm
Fig. 4 The amount of selected aqueous species measured in molal in the two-dimensional rock core at time step 20, corresponding to 27.42 minutes of simulations. The plots on the left correspond to the reference/benchmark chemical fields obtained using full GEM for chemical equilibrium calculation. The plots on the right are the results generated during the same simulation but utilizing the ODML algorithm with $\varepsilon = 0.001$

equilibrium calculation. The plots on the right are the results generated during the same simulation but utilizing the ODML algorithm with $\varepsilon = 0.001$

concentrations as the result of the NaCl-MgCl$_2$-CaCl$_2$-brine injection. Concentrations for the aqueous species predicted by the ODML algorithm (on the right) practically coincide with the benchmark snapshots (on the left), confirming that the smart chemical equilibrium algorithm (ODML+GEM) is able to maintain accuracy of the predicted chemical equilibrium states. These chemical fields correspond to the reactive transport simulations using the ODML acceleration.
strategy with tolerance $\varepsilon = 0.001$. The relative error obtained during the latter simulation is illustrated in Fig. 17 of Appendix. The confirmation on the elemental mass conservation constraint satisfaction for each element is presented in Fig. 18 of Appendix. We see that the mass balance relative error does not exceed the order of $10^{-13}$ and lower depending on the element.

**Number of on-demand learning operations** Figure 5a illustrates the number of triggered on-demand learnings on each time step. We select different error control parameters $\varepsilon$ to study how they affect the number of full GEM calculations (i.e., the on-demand learning operations) required for the ODML algorithm to satisfy each such tolerance. Most of the triggered learnings happen in the first 3,000 time steps (see Fig. 5b). After this, only 1-2 (out of 10,201) DOFs require occasional comprehensive and costly chemical equilibrium calculations to guarantee imposed accuracy levels on the chemical equilibrium states and permit subsequent states to remain accurately predicted. Even though these chemical states still contain precipitating dolomite and dissolving calcite, the ODML algorithm can successfully estimate them with minor errors.

The legend of Fig. 5a includes the total number of training operations required during the reactive transport simulation with the ODML algorithm with different accuracy requirements imposed. Even though the number of trainings is ten times higher for the $\varepsilon = 0.001$ than for $\varepsilon = 0.01$, all the learnings’ percentage remains below 0.02%. Figure 5 considers only the first 3,000 time steps to magnify the difference between the number of triggered conventional evaluations for each $\varepsilon$. We see that the total number of learnings triggered on these time steps for $\varepsilon = 0.001$, for instance, is 99.73% (13460 out of 13496) of all required full evaluations. We note that unlike homogeneous one-dimensional numerical test considered in [58], where on-demand learnings were the highest on the first few times step and then gradually decayed as reactive transport proceeds, in the heterogeneous case, we see several spikes in the number of learnings. For example, observe the increase in the number of learnings between steps 250 and 350, or between 850 and 1000. Such a sudden increase can be explained by the dissolution and precipitation front of the chemical system arriving to different parts of the rock with increasing/decreasing permeabilities and corresponding to more significant or lower amplitudes of the velocity. Having said that, the percentage of the total number of fast and accurate chemical equilibrium predictions enabled by the ODML algorithm remains higher than ~99.9% of all chemical equilibrium problems required in the simulations (i.e., less than ~0.1% of all such problems are solved using a full and expensive chemical equilibrium calculation provided by the conventional GEM or LMA algorithm).

**Computing cost reduction using the ODML (with the HKF activity model applied)** Figure 6a presents the computational cost of (i) conventional chemical equilibrium, (ii) smart chemical equilibrium, and (iii) transport calculations in relation to time steps. These simulation runs are performed for different $\varepsilon$ assigned to the ODML algorithm. The transport costs are determined by the time required for solving the systems generated by the SUPG method. They remain 1-2 orders of magnitude lower than the conventional chemical equilibrium calculations. The figure also highlights that the ODML approach promotes considerable decrease of the CPU cost associated with chemical equilibrium calculations. More restricted $\varepsilon$ is set, the higher the frequency of triggered on-demand learnings becomes. This
Fig. 6 (a) Computing costs (measured as the CPU time in seconds) for each sub-procedure of the reactive transport simulation in relation to time steps (including the ODML algorithm run with different \( \varepsilon \) and using the HKF activity model). (b) The corresponding speedup factor in chemical equilibrium calculations achieved by the ODML application.

Equally affects the CPU time of the corresponding ODML simulations (especially on the first 3000 steps).

Figure 6b shows the ODML speedups corresponding to the CPU costs and tolerances considered in Fig. 6a. These speedups are presented step-wise as a ratio of the total time required for the conventional and smart chemical equilibrium calculations across all the DOFs. The red curve illustrates the ODML algorithm’s speedup achieved for the strictest \( \varepsilon \), and, as expected, it reaches the lowest speedup values until the time step 3000. The blue and purple curves depict the speedups of the ODML algorithm performed with tolerances \( \varepsilon = 0.01 \) and \( \varepsilon = 0.005 \), respectively, indicate a similar acceleration level. We highlight that all the lines converge to the same speedup \( 9 \times \), approximately. This demonstrates that the ODML approach with on-demand clustering and priority-based search is able to attain similar speedup levels for different choices of accuracy tolerance. A stricter accuracy tolerance will cause this common speedup to be achieved later (after more time steps), but at the benefit of more accurate predictions for the rest of the simulation.

Computing cost reduction using the ODML (with the Pitzer activity model applied) Besides the HKF activity model, we also evaluate the ODML performance with a Pitzer activity model for the aqueous phase. Figure 7 shows the computing cost (in CPU time) of the chemical equilibrium computations when using Pitzer activity model. Only 10% of 10,000 time steps are considered in this setting. We can note, compared to previous results using HKF activity model, that the speedups strongly depend on the choice of the activity model.

The Pitzer model is considerably more computationally expensive to evaluate and require more Newton iterations...
to minimize Gibbs energy than the HKF model. Because the ODML transforms a complex GEM calculation into a simpler matrix-vector multiplication (first-order Taylor prediction), the achieved speedups are higher the more expensive the GEM calculation is. Thus, when the Pitzer activity model is used, the obtained speedups are considerably higher than that when using HKF. Figure 7b confirms for two different options of accuracy tolerance, $\epsilon = 0.01$ and $\epsilon = 0.001$, that for simulations with the Pitzer model, the ODML algorithm achieves speedups 100-150x (10 times higher than for the HKF model). The list of the ODML method characteristics for various tolerances and activity models are summarized in Fig. 16a.

**Clustering during the simulation process** During the reactive transport simulation, we adopt the on-demand clustering strategy introduced in [58]. This strategy classifies chemical states on a subject of primary species associated with each of them. The set of clusters is updated every time the on-demand learning operation happens. During such learning/training, a new fully evaluated chemical equilibrium state is produced (with a zero-priority rank) and stored. Suppose the primary species of this chemical state coincide with the primary species in one of the existing clusters. In that case, this particular cluster will be enriched with a newly learned reference state. If we fully compute a new chemical equilibrium state and no clusters correspond to that state’s primary species, a new cluster is created to store it. Whenever a reference chemical equilibrium state is successfully used to predict another, its priority rank (and the priority rank of the cluster) is incremented. Within each cluster, the records of learned chemical equilibrium states are sorted to use those with higher success rates/ranks first.

Table 3 lists only those clusters that were successfully used as the reference states more than twice (with the ranks higher than one), keeping the numbering, in which they were added during the test. The first column reflects this order (in time), whereas the second one lists the primary

Table 3 The list of clusters generated by the ODML method ($\epsilon = 0.01$) throughout the entire simulation run of the dolomitization example (using the HKF activity model)

| Clusters # | Primary species | Frequency / Rank | # of Records |
|------------|----------------|-----------------|--------------|
| 1          | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Ca$^{2+}$ Dolomite O$_2$ | 1,799,389 | 2 |
| 2          | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ HCO$_3$ Ca$^{2+}$ Dolomite O$_2$ | 326 | 1 |
| 3          | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ HCO$_3$ Ca$^{2+}$ Mg$^{2+}$ O$_2$ | 54,553 | 15 |
| 4          | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ Ca$^{2+}$ HCO$_3$ Mg$^{2+}$ O$_2$ | 118,099 | 1 |
| 9          | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ Ca$^{2+}$ OH$^{-}$ O$_2$ Mg$^{2+}$ | 5 | 7 |
| 11         | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ Ca$^{2+}$ HCO$_3$ O$_2$ Mg$^{2+}$ | 19 | 2 |
| 12         | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Ca$^{2+}$ Mg$^{2+}$ O$_2$(aq) | 10,914,505 | 8 |
| 14         | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Ca$^{2+}$ H$_2$(aq) Mg$^{2+}$ | 1,448 | 23 |
| 18         | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ OH$^{-}$ Ca$^{2+}$ Mg$^{2+}$ O$_2$ | 3 | 3 |
| 19         | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ Ca$^{2+}$ OH$^{-}$ Mg$^{2+}$ O$_2$ | 2 | 1 |
| 20         | H$_2$O(l) Dolomite Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Ca$^{2+}$ HCO$_3$ O$_2$(aq) | 7,910 | 1 |
| 21         | H$_2$O(l) Dolomite Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Mg$^{2+}$ Ca$^{2+}$ O$_2$(aq) | 49,064,564 | 1 |
| 22         | H$_2$O(l) Dolomite Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Ca$^{2+}$ Mg$^{2+}$ O$_2$(aq) | 3,436 | 1 |
| 23         | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Ca$^{2+}$ H$_2$(aq) Dolomite | 123,437 | 1,225 |
| 24         | H$_2$O(l) Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Mg$^{2+}$ HCO$_3$ Ca$^{2+}$ O$_2$(aq) | 77,668 | 2 |
| 25         | H$_2$O(l) Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Mg$^{2+}$ Ca$^{2+}$ O$_2$(aq) | 11,283,955 | 32 |
| 26         | H$_2$O(l) Dolomite Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Mg$^{2+}$ HCO$_3$ H$_2$(aq) | 4 | 27 |
| 27         | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Ca$^{2+}$ O$_2$(aq) Dolomite | 66 | 18 |
| 28         | H$_2$O(l) Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Mg$^{2+}$ Ca$^{2+}$ HCO$_3$ O$_2$(aq) | 28,499,792 | 3 |
| 29         | H$_2$O(l) Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Mg$^{2+}$ H$^{+}$ Ca$^{2+}$ O$_2$(aq) | 64,339 | 13 |
| 30         | H$_2$O(l) Calcite Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Dolomite Ca$^{2+}$ H$_2$(aq) | 8 | 16 |
| 31         | H$_2$O(l) Calcite Dolomite Cl$^{-}$ Na$^{+}$ CO$_2$(aq) Ca$^{2+}$ H$_2$(aq) | 15 | 44 |
| 32         | H$_2$O(l) Calcite Cl$^{-}$ Dolomite Na$^{+}$ CO$_2$(aq) Ca$^{2+}$ H$_2$(aq) | 18 | 15 |

Clusters # reflects the order they were generated. Frequency/Rank indicates the how often a cluster was used to retrieve suitable reference equilibrium state for new prediction. Column with # of Records corresponds to the number of fully evaluated and saved chemical states every cluster contains. Clusters #21 and #28 are responsible for 48.09% and 27.93%, respectively, of all 102,010,000 fast predictions, and together, these two clusters only have 4 learned equilibrium calculations.
species. Besides, the table presents the frequency each cluster was used in Taylor extrapolation as a reference equilibrium state as well as the number of equilibrium states each cluster stores. Clusters 1–19, 23, 27, 30-33 contain the stable calcite phase, but it is unstable in some of the clusters created on later times, reflecting the equilibrium states in which this mineral becomes completely dissolved. Clusters 20–22, 26 are responsible for chemical states where dolomite is stable or precipitates. Cluster 21 has the highest rank in providing reference equilibrium states for the accurate ODML approximations. This cluster’s single reference equilibrium state was successfully used 49,064,564 times for predictions, which accounts for 48.06% of all chemical equilibrium evaluations.

4.2 Case II: Reactive transport of H$_2$S scavenging

We continue with a reactive transport modeling of hydrogen sulfide scavenging, a widely adopted practice in the oil and gas industry production and processing operations. By sulfur scavenger, we understand a mineral that can react with sulfide species (H$_2$S, HS$^-$, and S$_2$O$_3$ etc) and convert them to a more inert form. For that purpose, siderite (FeCO$_3$) is considered below. Generally, the increase of the hydrogen sulfide mass is related to activities of sulfate-reducing bacteria (SRB) as a side-effect happening in the waterflooded reservoir (also called the reservoir souring). The prediction of H$_2$S generation and production in reservoirs remains a significant phenomenon for modeling and further understanding due to several reasons. Hydrogen sulfide is not only extremely toxic for humans and animals but is highly corrosive to most metals involved in the field operations. It may cause cracking of drill or transport pipes and tubular goods as well as destroy the testing tools and wire lines. Therefore, modeling of the mineral-H$_2$S reactions is important for studying the field-specific hydrogen sulfide scavenging capacities.

Below, we provide the set of the most important reactions happening in this example together with corresponding log $K$ constants ranging from -33.65 till 12.78:

\[
\begin{align*}
\text{Fe}^{2+} + \text{Cl}^- &= \text{FeCl}^+ & \log K &= -0.14 \\
\text{Fe}^{3+} + \text{H}_2\text{O} &= \text{FeOH}^{2+} + \text{H}^+ & \log K &= 2.19 \\
\text{Fe}^{2+} + \text{H}_2\text{O} &= \text{FeOH}^{+} + \text{H}^+ & \log K &= 9.5 \\
\text{Na}^+ + \text{HCO}_3^- &= \text{NaHCO}_3 & \log K &= 0.25 \\
\text{NaCl(aq)} &= \text{Na}^+ + \text{Cl}^- & \log K &= 1.57 \\
\text{Na}^+ + \text{SO}_4^{2-} &= \text{NaSO}_4^- & \log K &= -0.7 \\
\text{K}^+ + \text{SO}_4^{2-} &= \text{KSO}_4^- & \log K &= -0.85 \\
\text{MgCO}_3(aq) &= \text{CO}_3^{2-} + \text{Mg}^{2+} & \log K &= 2.98 \\
\text{Mg}^{2+} + \text{CO}_3^{3-} + \text{H}^+ &= \text{MgHCO}_3^- & \log K &= -11.399 \\
\text{Mg}^{2+} + \text{SO}_4^{2-} &= \text{MgSO}_4 & \log K &= -2.37 \\
\text{Mg}^{2+} + \text{H}_2\text{O} &= \text{H}^+ + \text{MgOH}^+ & \log K &= 11.44 \\
\text{Ca}^{2+} + \text{CO}_3^{2-} + \text{H}^+ &= \text{CaHCO}_3^- & \log K &= -11.435 \\
\text{CaCO}_3(aq) &= \text{CO}_3^{2-} + \text{Ca}^{2+} & \log K &= 8.48 \\
\text{Ca}^{2+} + \text{SO}_4^{2-} &= \text{CaSO}_4 & \log K &= -2.25 \\
\text{Ca}^{2+} + \text{H}_2\text{O} &= \text{H}^+ + \text{CaOH}^+ & \log K &= 12.78 \\
\text{Ca}^{2+} + \text{H}_2\text{O} &= \text{H}^+ + \text{CaOH}^+ & \log K &= -1.08 \\
\text{HS}^- + \text{H}^+ &= \text{H}_2\text{S} & \log K &= -6.994 \\
\text{H}^+ + \text{SO}_4^{2-} &= \text{HSO}_4^- & \log K &= -1.988 \\
\text{H}^+ + \text{CO}_3^{2-} &= \text{HCO}_3^- & \log K &= -10.329 \\
\text{FeCO}_3(s, iron sulfate) &= \text{Fe}^{2+} + \text{CO}_3^{2-} & \log K &= 10.89 \\
\text{FeS(s, pyrrhotite) + H}^+ &= \text{Fe}^{2+} + \text{HS}^- & \log K &= 4.648
\end{align*}
\]

Numerical model setup Similar to the previous example, Fig. 8 presents a considered reactive transport numerical model. For the reservoir, the horizontal and vertical lengths are 100 and 30 meters. We consider 101 and 31 DOFs for its discretization, which results in a total of 3,131 DOFs that must be considered in each time step. We fix the temperature to 25°C and inlet pressure to 1 atm (1.01325 bar), respectively. The heterogeneous permeability is illustrated in Fig. 9a, whereas the corresponding velocity $v$ is shown in Fig. 9b. The diffusion of fluid species is neglected. The

Fig. 8 Illustration of the H$_2$S-scavenging example in the two-dimensional heterogeneous rock core
Fig. 9 (a) The permeability field and (b) the Darcy velocity in the H2S-scavenging example

initial brine composition in the siderite-bearing (FeCO3) reservoir, the content of the brine pumped on the boundary, and thermodynamic and transport discretization parameters are communicated in Table 4. A detailed description of the chemical system can be found in Table 5. **Note:** In the ideal reservoir simulations, the rock matrix must contain a different proportion of minerals such as quartz, calcite, etc. However, such a simplification is assumed for the purpose of studying the scavenging process solely.

As highlighted above, the considered numerical test conducted in this section studies heterogeneous siderite-bearing reservoir continuously perturbed by the H2S-rich brine. Being highly soluble, FeCO3 dissociates donating the iron ion Fe2+, i.e.,

\[
\text{FeCO}_3 \rightleftharpoons \text{Fe}^{2+} + \text{CO}_3^{2-}.
\]

Simultaneously, the donated ferrous ions react with the sulfides (delivered by the brine) such that iron-sulfide FeS (pyrrhotite/mackinawite) precipitates:

\[
\text{HS}^- + \text{Fe}^{2+} \rightleftharpoons \text{FeS} + \text{H}^+.
\]

**Accuracy of generated chemical fields** The dissolution of FeCO3 (siderite) and precipitation of FeS (pyrrhotite) are

| Table 4 Parameters for the H2S-scavenging example |
|-----------------------------------------------|
| **Notation** | **Value** | **Specification** |
| Thermodynamic Conditions | $T$ | 25 °C | temperature |
| | $P$ | 1 atm | pressure |
| Physical Properties | $D$ | 0 m²/s | diffusion coefficient |
| Discretization Parameters | $\Delta x$ | 1.0 m | spatial mesh-size along the x-axis |
| | $\Delta y$ | 1.0 m | spatial mesh-size along the y-axis |
| | $N_{dofs}$ | 3131 | number of degrees of freedom |
| | $\Delta t$ | CFL/ \(\max \left\{ \max \{|v_x|/\Delta x, \max \{|v_y|/\Delta y\} \right\} \) | temporal discretization step |
| Initial Condition, pH = 8.951, pE = 8.676 | \(\phi\) | 10% | initial porosity |
| Composition of the reservoir brine | FeCO₃ | 0.5 mol | siderite |
| | H₂O | 58 kg | water |
| | O₂(aq) | 58.9 g | oxygen |
| | Cl⁻ | 1122.3 g | chloride anion |
| | SO₄²⁻ | 157.18 g | sulphate ion |
| | Mg²⁺ | 74.820 g | magnesium cation |
| | HCO₃⁻ | 8.236 g | carbonate anion |
| | Ca²⁺ | 23.838 g | calcium cation |
| | Na⁺ | 624.08 g | sodium cation |
| | K⁺ | 23.142 g | potassium cation |
| Boundary Condition, pH = 5.726, pE = 8.220 | Injected fluid composition (H₂S-brine) | Reservoir brine composition + |
| | H₂S(aq) | hydrogen sulfide |
| | HS⁻ | hydrogen sulfide anion |
shown in Fig. 10. We arrange the chemical fields generated by the reactive transport simulations using the conventional chemical equilibrium solver on the left side. We also highlight the parts of the rock with a preferential path formed due to higher permeability. The two-dimensional chemical fields on the right side correspond to the same numerical test run using the ODML algorithm with \( \varepsilon = 0.01 \). The minerals’ behavior is rather accurately approximated using ODML with even such a relaxed tolerance.

The dissolution and precipitation of minerals are accompanied by the increase and decrease in the aqueous species concentrations. Figure 11 shows the iron ions behavior at the same steps as the minerals’ two-dimensional chemical fields discussed above. Throughout the plot, we observe an initial gradual increase of Fe\(^{2+}\) as a result of FeCO\(_3\) dissolution. It is followed by the considerable drop of the iron ion concentration at the point of the phase transformation from one mineral to another as it gets used by the FeS formation. The region’s width, where Fe\(^{2+}\) is increased, is also getting more significant as the reactive transport simulation proceeds.

Figure 12 compares two-dimensional time snapshots of the sulfides \( S_2^{2-} \), HS\(^-\), and H\(_2\)S(aq) at the time step 400 (~141.04 days). The profiles with elevated species amounts coincide with parts of the rock penetrated by the injected brine. They also correspond to the siderite-pyrrhotite transformation front. Figure 12 verifies that the smart chemical equilibrium algorithm use does not reduce accuracy during the simulation, as the left and right chemical fields are relatively similar. Another confirmation is provided in Fig. 19 of Appendix, presenting the relative error obtained during the simulation run with the ODML method. On top of this, the satisfaction of element mass balance conservation is automatically in-build into the reconstructed species abundances. Figure 19 of Appendix highlights that by depicting the relative error in the mass balance equation for selected elements.

**Number of on-demand learning operations** Next, we study the dependence of the number of on-demand learnings on the ODML’s error control tolerance. In Fig. 13, the number of required comprehensive evaluations can reach up to 60, 22, or 20, depending on the selected tolerances \( \varepsilon = 0.001 \), \( \varepsilon = 0.005 \), or \( \varepsilon = 0.01 \), respectively. We see that triggered learnings are solely required up until 1,300 steps as the inserted brine penetrates the reservoir content. Again, we include the total number of on-demand trainings needed for each tolerance selected for the ODML in the legend of Fig. 13. This number is followed by the percentage it accounts from the total 6,262,000 chemical equilibrium problems that must be evaluated along the whole simulation process. As expected, the highest total number of learnings corresponds to the strictest tolerance \( \varepsilon = 0.001 \) (red marker). Due to heterogeneity of the medium, the velocity changes might cause more critical perturbations during the later transport step and, as a result, distinct initial chemical compositions for the ODML algorithm. It explains the isolated increase in learnings per time step (see, e.g., time steps 900-1,100 in Fig. 13). Nevertheless, independent of the tolerance, about 99.98\% of all chemical states are approximated using smart predictions based on the priority-based clustering combined with the first-order Taylor extrapolation. For the Pitzer activity model, the total number of learnings is slightly smaller for more relaxed tolerance \( \varepsilon = 0.01 \) and higher for \( \varepsilon = 0.001 \), even though the profile of occurring training per time step looks relatively similar in Fig. 13a and b.

**Computing cost reduction using the ODML** Figure 14a shows the computational cost (measured in seconds) step-wise spent

| Species          | Aqueous, Siderite, Pyrrhotite |
|------------------|-------------------------------|
| \( \text{CO(aq)} \) | \( \text{CaSO}_4\text{(aq)} \) | \( \text{FeCl}_2\text{+} \) | \( \text{H}_2\text{O(l)} \) | \( \text{HFeO}_2\text{(aq)} \) | \( \text{K}^+ \) | \( \text{MgOH}^+ \) | \( \text{S}_2\text{O}_3\text{2-} \) |
| \( \text{CO}_2\text{(aq)} \) | \( \text{Cl}^- \text{(aq)} \) | \( \text{FeCl}_2\text{(aq)} \) | \( \text{H}_2\text{O}_2\text{(aq)} \) | \( \text{HFeO}_2\text{−} \) | \( \text{KCl(aq)} \) | \( \text{MgSO}_4\text{(aq)} \) | \( \text{S}_2\text{O}_3\text{2-} \) |
| \( \text{CO}_2\text{(aq)} \) | \( \text{Cl}^- \text{(aq)} \) | \( \text{FeCl}_2\text{(aq)} \) | \( \text{H}_2\text{O}_2\text{(aq)} \) | \( \text{HFeO}_2\text{−} \) | \( \text{KCl(aq)} \) | \( \text{MgSO}_4\text{(aq)} \) | \( \text{S}_2\text{O}_3\text{2-} \) |
| \( \text{Ca}(\text{HCO}_3\text{)}^+\text{(aq)} \) | \( \text{ClO}_2\text{(aq)} \) | \( \text{FeO}^+ \) | \( \text{H}_2\text{S}_2\text{O}_3\text{(aq)} \) | \( \text{HS}^- \) | \( \text{KHO} \text{(aq)} \) | \( \text{NaCl(aq)} \) | \( \text{FeS(pyrrhotite)} \) |
| \( \text{Ca}^2\text{(aq)} \) | \( \text{ClO}_2\text{(aq)} \) | \( \text{FeO}^2\text{−} \) | \( \text{H}_2\text{S}_2\text{O}_4\text{(aq)} \) | \( \text{HS}_2\text{O}_4\text{(aq)} \) | \( \text{KSO}_4^- \) | \( \text{NaOH(aq)} \) |
| \( \text{CaCO}_3\text{(aq)} \) | \( \text{ClO}_2\text{(aq)} \) | \( \text{FeO}^+ \) | \( \text{H}_2\text{S}_2\text{O}_3\text{(aq)} \) | \( \text{HS}^- \) | \( \text{Mg(HCO}_3\text{)}^+ \) | \( \text{NaSO}_4^- \) |
| \( \text{CaO}_2\text{(aq)} \) | \( \text{FeO}^2\text{−} \) | \( \text{H}_2\text{S}_2\text{O}_3\text{(aq)} \) | \( \text{HS}^- \) | \( \text{Mg}^2\text{+} \) | \( \text{O}_2\text{(aq)} \) |
| \( \text{CaCl}_2\text{(aq)} \) | \( \text{Fe}^2\text{+} \) | \( \text{H}^+ \) | \( \text{HClO(aq)} \) | \( \text{H}_2\text{O}_2\text{(aq)} \) | \( \text{MgCO}_3\text{(aq)} \) | \( \text{OH}^- \text{(aq)} \) |
| \( \text{CaOH}^+\text{(aq)} \) | \( \text{FeCl}^+ \) | \( \text{H}_2\text{(aq)} \) | \( \text{HClO}_2\text{(aq)} \) | \( \text{H}_2\text{O}_2\text{(aq)} \) | \( \text{MgCl}^+ \) | \( \text{S}_2\text{O}_3\text{2-} \) |

\* \( Z \) is an electric charge
The amount of siderite (blue) and pyrrhotite (orange) measured in mol/m$^3$ in the two-dimensional rock core at time steps 100, 200, 400, and 800, corresponding to 35.26, 70.52, 141.04, and 282.08 days of simulations, respectively. The plots on the left correspond to the reference/benchmark chemical fields obtained using full GEM for chemical equilibrium calculation. The plots on the right are the results generated during the same simulation but utilizing the ODML algorithm with $\varepsilon = 0.01$

Even though we consider the heterogeneous two-dimensional problem, transport costs remain over 1.5 times lower than the CPU costs of conventional chemical equilibrium calculations. The ODML algorithm manages to decrease CPU costs of the chemical simulation by about one third.
The amount of iron cation Fe\(^{2+}\) in the two-dimensional rock core at 100, 200, 400, and 800 time steps, corresponding to 35.26, 70.52, 141.04, and 282.08 days of simulations using the ODML algorithm with \(\epsilon = 0.01\). Reactive transport simulations using the smart approach with \(\epsilon = 0.001\) have the highest costs (red marker) with occasional spikes, e.g., between time steps 900 and 1100, corresponding to higher learning rates. Next to the computation costs, we present speedup that the ODML can achieve in chemical calculations compared to the conventional approach. We observe that the average speedup with the Debye-Hückel activity model is over 30x times, except for \(\epsilon = 0.001\) on the first 1,200-1,300 steps. Nevertheless, the speedup is stabilized eventually around the value 33x, independent of how much overall trainings were performed and stored on the first steps. It confirms the efficiency of the reference chemical states retrieval procedure when the priority-based queue and clustering are incorporated.

**Computing cost reduction using the ODML (with the Pitzer activity model applied)** For comparison, we perform a similar calculation using the Pitzer activity model (see Fig. 15). Figure 15a indicates that the step-wise CPU costs utilizing the classical method (gray curve) are approximately two orders of magnitude more computationally demanding than when using the simpler DH model. The gray curve is at least several thousand times more costly than the transport-related green curve. The blue and red curves, corresponding to the reactive transport simulation with the ODML method, indicate that the smart approach drastically improves the costs of the conventional calculations, yielding them comparable to transport operations. The corresponding speedup presented in Fig. 15b has initially relatively low values due to the excessive number of training operations. However, it reaches 3500x (blue line) and 2500x (red line) and finally stabilizes around 2000x value at later steps. The ODML performance summary for various tolerances and activity models is presented in Fig. 16b.

**Clustering during the simulation process** Similar to the previous section, we present the clusters generated during the reactive transport simulation with the ODML algorithm with \(\epsilon = 0.01\). Table 6 lists only those that were successfully used as the reference states more than twice, using the order they were created. The primary species of presented clusters consist of some species that maintain constant positions, like H\(_2\)O(l), Cl\(^-\), Na\(^+\), Mg\(^{2+}\), SO\(_4^{2-}\), K\(^+\), Ca\(^{2+}\), whereas some other species, like minerals siderite, pyrrhotite, carbonate-containing species HCO\(_3^-\), CO\(_2(aq)\), MgCO\(_3(aq)\), sulfides H\(_2\)S (aq), HS\(^-\), and iron-containing species FeO\(^+\), FeO\(^+\), Fe\(^{2+}\), HFeO\(_2(aq)\) are constantly altering their place. Siderite is stable in Clusters 2–10, 12, 16, 19-20, 23, 25, 28, 31, 34, but unstable in those that reflect chemical states with the completely dissolved mineral. Clusters 1, 10–11, 14, 17-18,
The amount of sulfides $S_2^{2−}$, $H_2S^{−}$, and $H_2S$ (aq) in the two-dimensional rock core at the time step 400. The chemical fields on the left are generated during the benchmark reactive transport simulation based on full GEM equilibrium calculations. The plots on the right are the results of the similar numerical test using the ODML algorithm with $ε = 0.01$.

**Fig. 12**

(a) $S_2^{2−}$ with the conventional algorithm
(b) $S_2^{2−}$ with the ODML algorithm
(c) $H_2S^{−}$ with the conventional algorithm
(d) $H_2S^{−}$ with the ODML algorithm
(e) $H_2S$ (aq) with the conventional algorithm
(f) $H_2S$ (aq) with the ODML algorithm

The number of on-demand learning operations performed by the ODML algorithm in relation to reactive transport time step. The frequency of these learnings is illustrated for different $ε$ applying (a) the Debye-Hückel and (b) the Pitzer activity model.

**Fig. 13**

(a) the Debye-Hückel activity model
(b) the Pitzer activity model

With 2,000 time steps containing 3131 DOFs each, we obtain 6,262,000 chemical equilibrium states in total to be computed. The legend includes the total number of on-demand learning operations required by each run and the percentage it accounts from the total chemical evaluations.
Fig. 14  (a) Computing costs (measured as the CPU time in seconds) for each sub-procedure of the reactive transport simulation w.r.t. time steps (including the ODML algorithm run with different $\epsilon$ and using the Debye-Hückel activity model). (b) The corresponding speedup factor in chemical equilibrium calculations achieved by the ODML application.

21-22, 29-30 correspond to chemical states where pyrrhotite is precipitated. Finally, equilibrium states, where neither of the minerals is present, are reflected by Clusters 13, 24, 26-27, and 36-46. Cluster 21 is the most often used for suitable reference equilibrium states that helped with accurate approximations of the ODML algorithm. Only two states stored in this cluster were used $3,899,826$ times in the Taylor extrapolation, which bears responsibility for $62.27\%$ of all smart predictions.

Note  Similar reactive transport setup has been tested for the chemical systems with alternative minerals being dissolved, i.e. siderite, hematite, goethite, and pyrite being precipitated. The respective reactions are listed below:

$$\text{FeCO}_3(s, \text{siderite}) \rightleftharpoons \text{Fe}^{2+} + \text{CO}_3^{2-} \quad \log K = 10.89$$

$$\text{Fe}_2\text{O}_3(s, \text{hematite}) + 6\text{H}^+ \rightleftharpoons 3\text{H}_2\text{O(l)} + 2\text{Fe}^{3+} \quad \log K = 4.008$$

$$\text{FeOOH(s, goethite)} + 3\text{H}^+ \rightleftharpoons \text{Fe}^{3+} + 2\text{H}_2\text{O(l)} \quad \log K = 1.0$$

$$\text{Fe}^{2+} + 2\text{HS}^- \rightleftharpoons \text{FeS}_2(\text{pyrite}) + 2\text{H}^+ \quad \log K = -18.479$$

Even with the presence of the non-unit stoichiometry coefficients and a rather wide range of the equilibrium...
**Case I: Dolomitization process**

| HKF activity model | Pitzer activity model |
|--------------------|-----------------------|
| $\varepsilon = 0.01$ | $\varepsilon = 0.01$ |
| # of learnings | # of learnings |
| 1271 | 1003 |
| % of smart predictions | % of smart predictions |
| 99.999% | 99.999% |
| Speedup in chemical equilibrium | Speedup in chemical equilibrium |
| 7-22x | 84-264x |
| Total speedup in reactive transport | Total speedup in reactive transport |
| 4.5x | 53.6x |

**Case II: $H_2S$-scavenging process**

| Debye-Hückel activity model | Pitzer activity model |
|-----------------------------|-----------------------|
| $\varepsilon = 0.01$ | $\varepsilon = 0.01$ |
| Total learnings | Total learnings |
| 966 | 894 |
| % of smart predictions | % of smart predictions |
| 99.985% | 99.986% |
| Speedup in chemical equilibrium | Speedup in chemical equilibrium |
| 21-50x | 58-3785x |
| Total speedup in reactive transport | Total speedup in reactive transport |
| 13x | 56x |

**Fig. 16** Recapitulation of the total number of learning operations, the percentage of the smart predictions in relation to the total number of the chemical equilibrium calculations, the range of speedups in chemical equilibrium calculations step-wise, and the overall reactive transport simulation speedups for different tolerances applying (a) the HKF and the Pitzer activity models in the dolomitization example and (b) the DH and the Pitzer activity models in the scavenging example.
Table 6  The list of clusters generated by the ODML method ($\varepsilon = 0.01$) throughout the entire simulation run of the H$_2$S example (with the DH activity model)

| Clusters # | Primary Species | Frequency / Rank | # of Records |
|------------|-----------------|-----------------|--------------|
| 1          | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Pyrrhotite HCO$_3^-$ CO$_2$(aq) Siderite | 119,835 | 1 |
| 2          | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ FeFeO$_2$(aq) MgCO$_3$(aq) | 4,444 | 3 |
| 3          | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Pyrrhotite MgCO$_3$(aq) | 116 | 1 |
| 4          | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ MgCO$_3$(aq) HFeO$_2$(aq) | 712,472 | 24 |
| 5          | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ MgCO$_3$(aq) Pyrrhotite | 256 | 1 |
| 6          | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ MgCO$_3$(aq) HS$^-$ | 12,147 | 108 |
| 7          | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ FeFeO$_2$(aq) FeO$^+$ | 5,492 | 3 |
| 8          | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ Siderite FeFeO$_2$(aq) FeO$^+$ | 1,954 | 1 |
| 9          | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ Pyrrhotite CO$_2$(aq) | 310 | 1 |
| 10         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Pyrrhotite HCO$_3^-$ CO$_2$(aq) H$_2$S(aq) | 43,065 | 8 |
| 11         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ CO$_2$(aq) FeFeO$_2$(aq) | 63,669 | 19 |
| 12         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ FeFeO$_2$(aq) O$_2$(aq) | 40 | 49 |
| 13         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Pyrrhotite H$_2$S$_2$(aq) HCO$_3^-$ | 383,052 | 25 |
| 14         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ HCO$_3^-$ FeO$^+$ HFeO$_2$(aq) Siderite | 1,466 | 1 |
| 15         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Pyrrhotite H$_2$S$_2$(aq) HCO$_3^-$ H$_2$S(aq) | 4,176 | 6 |
| 16         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ FeFeO$_2$(aq) FeO$^+$ | 3,759 | 4 |
| 17         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ CO$_2$(aq) FeO$^+$ | 825,867 | 31 |
| 18         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ CO$_2$(aq) HS$^-$ | 25,644 | 81 |
| 19         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Pyrrhotite H$_2$S$_2$(aq) HCO$_3^-$ CO$_2$(aq) Fe$^{2+}$ | 3,899,826 | 2 |
| 20         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Pyrrhotite H$_2$S$_2$(aq) HCO$_3^-$ HS$^-$ | 1207 | 4 |
| 21         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ CO$_2$(aq) H$_2$S(aq) | 67,448 | 26 |
| 22         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Pyrrhotite HCO$_3^-$ CO$_2$(aq) O$_2$(aq) | 73 | 91 |
| 23         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ CO$_2$(aq) FeO$^+$ | 80,302 | 2 |
| 24         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ CO$_2$(aq) FeO$^+$ | 744 | 294 |
| 25         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ CO$_2$(aq) FeO$^+$ | 260 | 16 |
| 26         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ CO$_2$(aq) | 15 | 59 |
| 27         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Siderite HCO$_3^-$ CO$_2$(aq) | 9 | 1 |
| 28         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Pyrrhotite H$_2$S$_2$(aq) FeFeO$_2$(aq) CO$_2$(aq) | 9 | 2 |
| 29         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Pyrrhotite H$_2$S$_2$(aq) | 9 | 3 |
| 30         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Pyrrhotite H$_2$S$_2$(aq) | 296 | 1 |
| 31         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ Pyrrhotite H$_2$S$_2$(aq) | 2,622 | 1 |
| 32         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ HCO$_3^-$ CO$_2$(aq) O$_2$(aq) FeO$^+$ | 18 | 4 |
| 33         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ HCO$_3^-$ CO$_2$(aq) FeO$^+$ | 68 | 2 |
| 34         | H$_2$O(l) Cl$^-$ Na$^+$ Mg$^{2+}$ SO$_4^{2-}$ K$^+$ Ca$^{2+}$ HCO$_3^-$ CO$_2$(aq) FeO$^+$ | 318 | 1 |

We select 38 clusters (out of the total 46) used more than twice by the ODML for the quick and smart predictions. Clusters # reflects the order they were generated. Frequency / Rank indicates how often a cluster was used to retrieve suitable reference equilibrium state for new prediction. Column with # of Records corresponds to the number of fully evaluated and saved chemical states every cluster contains.
constant, the efficiency of the ODML approach remains similar to the above-reported test cases.

5 Discussion and conclusions

Reaktoro and Firedrake’s coupling enabled investigating the performance of the new ODML algorithm on more challenging reactive transport problems. The obtained numerical results confirm that the resulting acceleration of the chemical equilibrium calculations provided by the new strategy depends on several factors. The first and most important one remains the activity models used for the aqueous species in the numerical experiment. The second is the chosen error control/accuracy tolerance (parameter $\varepsilon$) which determines how strict the acceptance criterion is. Ultimately, the obtained acceleration depends on the complexity of the numerical reactive transport experiment, e.g., chemical system size, heterogeneity, mesh dimension, etc. Having said that, the ODML algorithm enables speedups of one to three orders of magnitude in chemical equilibrium calculation and at least one order of magnitude of overall reactive transport simulations. A significant property of ODML strategy is its ability to converse mass of chemical elements (including the electric charge) up to machine precision levels, a capability not naturally found in prevailing neural network algorithms (e.g., neural network and most classes of surrogate models). This inherent feature of the algorithm is explained and demonstrated mathematically in [58].

To highlight the most critical performance characteristics of the ODML algorithm, we use Fig. 16a and b. To be specific, we recap the overall number of learning operations each simulation run required for various tolerances and activity models. Furthermore, we emphasize that the percentage of smart predictions remains greater than 99.8% in relation to the total number of chemical equilibrium evaluations throughout complete simulation run. We additionally include the lowest and highest speedups in obtained in chemical equilibrium calculations step-wise as well as the overall speedups in the reactive transport simulations enhanced by the ODML method.

In future works, we aim to apply the ODML strategy to accelerate modeling of chemical kinetics with partial chemical equilibrium assumptions, typical in geochemical systems. We equally consider further investigations with more challenging geochemical and geological models. We plan to extend Reaktoro’s functionality to model reservoirs souring as a result of the activities of sulfide-reducing bacteria, mixing of the groundwater and seawater in the oil reservoir as well as scaling effects this process results to, modeling the effects that seawater or sodium chloride cause during the cement rock attack, among many more. An extension to the three-dimensional problem with heterogeneity will require a stable numerical scheme to solve the Darcy problem. To enable full coupling of the transport and flow problems, we plan to continue developing the reactive transport simulator and present results in future articles.

As this work presents a new coupling of the software packages to attempt reactive transport modeling, we acknowledge an important and well-known MoMaS benchmark [1, 3, 15, 23, 30, 39] used to compare numerical solvers developed for reactive transport simulations in porous media. At this moment, we are limited to chemical equilibrium formulations in terms of the GEM, but are currently working on the new version of the ODML algorithm. Therefore, assessing its performance for the chemical equilibrium problems considered in the MoMaS benchmark is planned to be included in future work.

Appendix

In this appendix, we collect the plots showing the relative errors in minerals and aqueous species on the fixed reactive transport steps for both Cases I and II (see Figs. 17 and 19, respectively). Moreover, using similar graphics, we demonstrate that the balance equation $An = b$ is intrinsically satisfied in the approximations provided by the ODML algorithm (up to machine precision). In particular, Figs. 18 and 20 confirm this property by illustrating relative errors of the balance conservation for each of the elements used to generate the chemical systems in dolomitization and scavenging examples.
Fig. 17 Relative error in minerals and several aqueous species on the fixed reactive transport steps in the dolomitization example. The ODML algorithm performed with $\varepsilon = 0.001$
Fig. 18 Relative error in the mass balance equation for all the elements on the reactive step 1500 in the dolomitization example. The ODML algorithm preformed with $\varepsilon = 0.001$
Fig. 19  Relative error in minerals and several aqueous species on the fixed reactive transport steps in the H₂S-scavenging example. The ODML algorithm performed with ε = 0.001.
Fig. 20 Relative error in the mass balance equation for the selected elements on the reactive step 200 in the H$_2$S-scavenging example. The ODML algorithm performed with $\varepsilon = 0.001$. 
Acknowledgments This research project is supported by the Swiss National Science Foundation (Ambizione Grant PZ00P2-179967), the Werner Siemens Foundation (Werner Siemens-Stiftung), and Shell Global Solutions International BV.

Funding Open access funding provided by Swiss Federal Institute of Technology Zurich.

Author Contributions
- Svetlana Kyas: Conceptualization, Formal analysis, Methodology, Software, Investigation, Validation, Writing - original draft, Writing - review & editing.
- Diego Volpatto: Conceptualization, Formal analysis, Methodology, Software, Investigation, Validation, Writing - review & editing.
- Martin O. Saar: Funding acquisition, Writing - review & editing.
- Allan M. M. Leal: Funding acquisition, Conceptualization, Formal analysis, Methodology, Software, Investigation, Validation, Writing - review & editing.

Data Availability For this research, the following open-source numerical packages were used and developed:
- Reaktoro, LGPL v2.1 (https://reaktoro.org/), https://github.com/reaktoro/reaktoro.
- Firedrake, LGPL v3.1 (https://www.firedrakeproject.org/), https://github.com/firedrakeproject/firedrake.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References
1. Abubakr, E., El Ossmani, M., Id Moulay, M.: A fully implicit finite volume scheme for single phase flow with reactive transport in porous media. Math. Comput. Simul. 164, 3–23 (2019)
2. Alnæs, M.S., Logg, A., Ølgaard, K.B., Rognes, M.E., Wells, G.N.: Unified form language: a domain-specific language for weak formulations and partial differential equations. ACM Trans. Math. Softw. 40(2), Art. 9, 37 (2014)
3. Amir, L., Kern, M.: Preconditioning a coupled model for reactive transport in porous media. Int. J. Numer. Anal. Model. 16(1), 18–48 (2019)
4. Arnold, D.N., Brezzi, F.: Mixed and nonconforming finite element methods: implementation, postprocessing and error estimates. RAIRO Modél. Math. Anal. Numer. 19(1), 7–32 (1985)
5. Arnold, D.N., Brezzi, F., Cockburn, B., Marini, L.D.: Unified analysis of discontinuous Galerkin methods for elliptic problems. SIAM J. Numer. Anal. 39(5), 1749–1779 (2001/02)
6. Azad, V.J., Li, C., Verba, C., Ideker, J.H., Isgor, O.B.: A COMSOL-GMS interface for modeling coupled reactive transport geochemical processes. Comput. Geosci. 92, 79–89 (2016)
7. Bächler, D., Kohl, T.: Coupled thermal-hydraulic-chemical modelling of enhanced geothermal systems. Geophys. J. Int. 161(2), 533–548 (2005)
8. Berrone, S., Della Santa, F., Pieraccini, S., Vaccarino, F.: Machine learning for flux regression in discrete fracture networks. GEM - Int. J. Geomath. 12(9) (2021)
9. Bethke, C.M. Geochemical and Biogeochemical Reaction Modeling, 2nd edn. Cambridge University Press, New York (2007)
10. Bilke, L., Flemisch, B., Kalbach, T., Kolditz, O., Helming, R., Nagel, T.: Development of Open-Source porous media simulators: Principles and experiences. Transp. Porous Media 130, 337–361 (2019)
11. Bochev, P.B., Gunzburger, M.D.: Finite element methods of lowest-squares type. SIAM Rev. 40(4), 789–837 (1998)
12. Brezzi, F., Fortin, M.: A minimal stabilisation procedure for mixed finite element methods. Numer. Math. 89(3), 457–491 (2001)
13. Brooks, A.N., Hughes, T.J.R.: Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. Comput. Methods Appl. Mech. Engrg. 32(1-3), 199–259. FENOMECH ’81, Part I (Stuttgart 1981) (1982)
14. Carrayrou, J., Hoffmann, J., Knabner, P., Kräutle, S., de Dieulleveult, C., Erhel, J., Van der Lee, J., Lagneau, V., Mayer, K.U., MacQuarrie, K.T.B.: Comparison of numerical methods for simulating strongly nonlinear and heterogeneous reactive transport problems – the momas benchmark case. Comput. Geosci. (14), 483–502 (2010a)
15. Carrayrou, J., Kern, M., Knabner, P.: Reactive transport benchmark of MoMaS. Comput. Geosci. (14) (2010b)
16. Carrayrou, J., Mosé, R., Behra, P.: Operator-splitting procedures for reactive transport and comparison of mass balance errors. J. Contam. Hydrol. 68(3-4), 239–268 (2004)
17. Centl, F., Shao, H., De Biase, C., Park, C.H., Regnier, P., Kolditz, O., Thullner, M.: GeosysBRNS-a flexible multidimensional reactive transport model for simulating biogeochemical subsurface processes. Comput. Geosci. 36(3), 397–405 (2010)
18. Cockburn, B., Gopalakrishnan, J.: A characterization of hybridized mixed methods for second order elliptic problems. SIAM J. Numer. Analy. 42(1), 283–301 (2004)
19. Cockburn, B., Gopalakrishnan, J., Lazarov, R.: Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems. SIAM. J. Numer. Anal. 47(2), 1319–1365 (2009)
20. Correa, M.R., Loula, A.F.D.: Stabilized velocity post-processing for Darcy flow in heterogeneous porous media. Comm. Numer. Methods Engrg. 23(6), 461–489 (2007)
21. Correa, M.R., Loula, A.F.D.: Unconditionally stable mixed finite element methods for Darcy flow. Comput. Methods Appl. Mech. Engrg. 197(17-18), 1525–1540 (2008)
22. Damiani, L.H., Kosakovski, G., Glaus, M.A., Churakov, S.V.: A framework for reactive transport modeling using FEniCS-reaktoro: governing equations and benchmarking results. Comput. Geosci. 24(3), 1071–1085 (2020)
23. de Dieulleveult, C., Erhel, J.: A global approach to reactive transport: application to the MoMaS benchmark. Comput. Geosci. (14), 451–464 (2010)
24. Debye, P., Hückel, E.: The theory of electrolytes. 1. lowering of freezing point and related phenomena. Phys. Zeitsch. 24, 85–206 (1923)
25. Drummond, S.: Boiling and Mixing of Hydrothermal Fluids: Chemical Effects on Mineral Precipitation. Ph.d, Pennsylvania State University (1981)
26. Elakneswaran, Y., Ishida, T.: Development and verification of an integrated physicochemical and geochemical modelling framework for performance assessment of cement-based materials. J. Adv. Concr. Technol. 12(4), 111–126 (2014)
27. Fuks, O., Tchelepi, H.A.: Limitations of physics informed machine learning for nonlinear two-phase transport in porous media. J. Mach. Learn. Model. Comput. 1(1), 9–17 (2020)
28. Gamazo, P., Slooten, L.J., Carrera, J., Saaltink, M.W., Bea, S., Soler, J.: PROOST: Object-Oriented approach to multiphase reactive transport modeling in porous media. J. Hydroinf. 18(2), 310–328 (2016)
29. Georgot, F., Prevost, J.H., Huet, B.: A reactive transport simulator for variable porosity problems. Comput. Geosci. 21(1), 95–116 (2017a)
30. Georgot, F., Prévost, J.H., Huet, B.: A reactive transport simulator for variable porosity problems. Comput. Geosci. 21(1), 95–116 (2017b)
31. Guo, B., Hong, Y., Qiao, G., Ou, J.: A COMSOL-PHREEQC interface for modeling the multi-species transport of saturated cement-based materials. Construct. Build Mater. 187, 839–853 (2018)
32. Harvie, C.E., Møller, N., Weare, J.H.: The prediction of mineral solubilities in natural waters: The Na-K-Mg-Ca-H-CI-SO4-OH-HCO3-CO3-CO2-H2O system to high ionic strengths at 25°C. Geoch. Cosmoch. Acta 48(4), 723–751 (1984)
33. He, W., Beyer, C., Fleckenstein, J.H., Jang, E., Kolditz, O., Naumov, D., Kalbacher, T.: A parallelization scheme to simulate reactive transport in the subsurface environment with OGS#IP hreeqc 5.5.7-3.1.2. Geosci. Model Dev. 8(10), 3333–3348 (2015)
34. Helgeson, H.C., Delany, J.M., Nesbitt, H.W., Bird, D.K.: Summary and critique of the thermodynamic properties of rock-forming minerals. Amer. J. Sci. 278 A(1), 229 (1978)
35. Helgeson, H.C., Kirkham, D.H.: Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures: I. Summary of the thermodynamic/electrostatic properties of the solvent. Am. J. Sci. 274(10), 1089–1198 (1974a)
36. Helgeson, H.C., Kirkham, D.H.: Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures: II. Debye-huckel parameters for activity coefficients and relative partial molal properties. Am. J. Sci. 274(10), 1199–1261 (1974b)
37. Helgeson, H.C., Kirkham, D.H.: Theoretical prediction of the thermodynamic properties of aqueous electrolytes at high pressures and temperatures: III. Equation of state for aqueous species at infinite dilution. Am. J. Sci. 276(2), 97–240 (1976)
38. Helgeson, H.C., Kirkham, D.H., Flowers, G.C.: Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures: IV. Calculation of activity coefficients, osmotic coefficients, and apparent molal and standard and relative partial molal properties to 600 C. Am. J. Sci. 281(10), 1249–1516 (1981)
39. Hoffmann, J., Kräutle, S., Knabner, P.: A parallel global-implicit 2-D solver for reactive transport problems in porous media based on a reduction scheme and its application to the MoMaS benchmark problem. Comput. Geosci. (14), 421–433 (2010)
40. Jacques, D., Simunek, J.: User manual of the multicomponent variably-saturated flow and transport model hpl (2005)
41. Jara, D., de Dreuxy, J.R., Cochevin, B.: TReEcalb: An object-oriented implementation of non-intrusive splitting methods to couple independent transport and geochemical software. Comput. Geosci. 109, 281–294 (2017)
42. Jasakh, H.: OpenFOAM: Open source CFD in research and industry. International Journal of Naval Architecture and Ocean Engineering (2012)
43. Johnson, J.W., Oelkers, E.H., Helgeson, H.C.: SUPCRT92: A Software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species, and reactions from 1 to 5000 bar and 0 to 1000 C. Comput. Geosci. 18(7), 899–947 (1992)
44. Kolditz, O., Bauer, S., Bülke, L., Böttcher, N., Delfs, J., Fischer, T., Görke, U., Kalbacher, T., Kosakovski, G., McDermott, C., Park, C., Rad, U.F., Rink, K., Shao, H., Sun, F., Sun, Y., Singh, A., Taron, J., Walther, M., Wang, W., Watanabe, N., Wu, N., Xie, M., Xu, W., Zehner, B.: Opengeosys. OpenGeoSys (OGS) is a scientific open source project for the development of numerical methods for the simulation of thermo-hydro-mechanical-chemical (THMC) processes in porous and fractured media (2019)
45. Kolditz, O., Görke, U.-J., Shao, H., Wang, W.: Thermo-Hydro-Mechanical-Chemical Processes in Porous Media - Benchmarks and Examples, vol. 86. Springer (2012)
46. Kolditz, O., Nagel, T., Shao, H., Wang, W., Sebastian, B.: Thermo-Hydro-Mechanical-Chemical Processes in Fractured Porous Media. Modelling and Benchmarking – From Benchmarking to Tutoring. Springer International Publishing (2018)
47. Kosakovski, G., Watanabe, N.: Opengeosys-gem: A numerical tool for calculating geochemical and porosity changes in saturated and partially saturated media. Phys. Chem. Earth 70-71, 138–149 (2014)
48. Kulik, D., Berner, U.R., Curti, E.: Modeling Chemical Equilibrium Partitioning with the GEMS-PSI Code. Technical Report March Paul Scherrer Institut, Villigen (2004)
49. Kulik, D.A., Wagner, T., Dmytryeva, S.V., Kosakovski, G., Hingerl, F.F., Chudnenko, K.V., Berner, U.R.: GEMS-Selektor geochemical modeling package: revised algorithm and GEMS3k numerical kernel for coupled simulation codes. Comput. Geosci. 17(1), 1–24 (2013)
50. Langevin, C., Hughes, J., Banta, E., Provost, A., Niswonger, R., Panday, S.: MODFLOW 6 Modular Hydrologic Model version 6.1.0: U.S. Geological Survey Software Release (2019)
51. Leal, A.M.: Reaktoro: A unified framework for modeling chemically reactive systems (2015)
52. Leal, A.M., Blunt, M.J., LaForce, T.C.: Efficient chemical equilibrium calculations for geochemical speciation and reactive transport modeling. Geochim. Cosmochim. Acta 131, 301–322 (2014)
53. Leal, A.M., Blunt, M.J., LaForce, T.C.: A chemical kinetics algorithm for geochemical modelling. Appl. Geochem. 55, 46–61 (2015)
54. Leal, A.M., Kulik, D.A., Kosakovski, G.: Computational methods for reactive transport modeling: A Gibbs energy minimization approach for multiphase equilibrium calculations. Adv. Water Resour. 88, 231–240 (2016a)
55. Leal, A.M., Kulik, D.A., Kosakovski, G., Saar, M.O.: Computational methods for reactive transport modeling: An extended law of mass-action, xLMA, method for multiphase equilibrium calculations. Adv. Water Resour. 96, 405–422 (2016b)
56. Leal, A.M.M., et al.: autodiff, a modern, fast and expressive C++ library for automatic differentiation. https://autodiff.github.io (2018)
57. Leal, A.M.M., Kulik, D.A., Smith, W.R., Saar, M.O.: An overview of computational methods for chemical equilibria and kinetic calculations for geochemical and reactive transport modeling. Pure Appl. Chem. 89(5), 597–643 (2017)
58. Leal, A.M.M., Kyas, S., Kulik, D.A., Saar, M.O.: Accelerating reactive transport modeling: on-demand machine learning algorithm for chemical equilibrium calculations. Transp. Porous Media 133(2), 161–204 (2020)
59. Li, D., Bauer, S., Benisch, K., Graupner, B., Beyer, C.: Opengeosys-chemapp: A coupled simulator for reactive transport in multiphase systems and application to CO2 storage formation in Northern Germany. Acta Geotech. 9(1), 67–79 (2014)
60. Lichtner, P.C.: Continuum model for simultaneous chemical reactions and mass transport in hydrothermal systems. Geochim. Cosmochim. Acta 49(3), 779–800 (1985)
98. Van der Lee, J., De Windt, L., Lagneau, V., Goblet, P.: Module-oriented modeling of reactive transport with HYTEC. Comput. Geosci. 29(3), 265–275 (2003)
99. Wagner, W., Pruss, A.: The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. J. Phys. Chem. Ref. Data 31(2), 387 (2002)
100. Wen, G., Hay, C., Benson, S.M.: CCSNet: a deep learning modeling suite for CO2 storage. (arXiv:2104.01795 [physics.flu-dyn]) (2021)
101. White, M., Oostrom, M.: Stomp subsurface transport over multiple phases version 4.0 user’s guide. Technical report (2006)
102. Xiao, Y., Whitaker, F., Xu, T., Steefel, C.: Reactive transport modeling: Applications in subsurface energy and environmental problems (2018)
103. Xu, T., Sonnenthal, E., Spycher, N., Pruess, K.: TOUGHREACT - A Simulation program for non-isothermal multiphase reactive geochemical transport in variably saturated geologic media; Applications to geothermal injectivity and CO2 geological sequestration. Comput. Geosci. 32(2), 145–165 (2006)
104. Yan, B., Harp, D.R., Chen, B., Pawar, R.: A Physics-Constrained Deep Learning Model for Simulating Multiphase Flow in 3D Heterogeneous Porous Media. (arXiv:2105.09467 [physics.geo-ph]) (2021)
105. Yapparova, A., Gabellone, T., Whitaker, F., Kulik, D.A., Matthäi, S.K.: Reactive transport modelling of dolomitisation using the new CSMP++GEM coupled code: Governing equations, solution method and benchmarking results. Transp. Porous Media 117(3), 385–413 (2017)
106. Yeh, G., Tsai, C., Ni, C.: Hydrogeochem 6.0: A model to couple thermal-hydrology-mechanics-chemical (thmc) processes user guide. Technical report (2013)
107. Zheng, C., Wang, P.: Mt3dms: A modular three-dimensional multispecies transport model for simulation of advection, dispersion and chemical reactions of contaminants in ground water systems: documentation and user’s guide. Technical report. http://hydro.geo.ua.edu/mt3d (1999)
108. Zhou, Z., Shi, L., Zha, Y.: Seeing macro-dispersivity from hydraulic conductivity field with convolutional neural network. Adv. Water Resour. 138(103545), 421–433 (2020)
109. Zhu, Y., Zabaras, N.: Bayesian deep convolutional encoder-decoder networks for surrogate modeling and uncertainty quantification. J. Comput. Phys. 366(1), 415–447 (2018)
110. Zhu, Y., Zabaras, N., Koutsourelakis, P.S., Perdikaris, P.: Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled dataa. J. Comput. Phys. 394(1), 56–81 (2019)

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