An emulation-based approach for interrogating reactive transport models

Authors: Angus Fotherby[1*], Harold J. Bradbury[1,2], Jennifer L. Druhan[3], Alexandra V. Turchyn[1]

1: Department of Earth Sciences, University of Cambridge, Cambridge, UK

2: Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, Canada

3: Department of Geology, University of Illinois at Urbana Champaign, Urbana, IL

*corresponding author: af606@cam.ac.uk

Keywords: Environmental remediation, Reactive transport, Machine learning, Sensitivity analysis

Abstract

We present an emulation-based approach to understand the interactions among different chemical and biological processes modelled in environmental reactive transport models (RTMs) and explore how the parameterisation of these processes influences the results of multi-component RTMs. We utilize a previously published RTM consisting of 20 primary species, 20 secondary complexes, 17 mineral reactions and 2 biologically-mediated reactions which describes bio-stimulation using sediment from a contaminated aquifer. We choose a subset of the input parameters to vary over a range of values. The result is the construction of a new dataset that describes the model behaviour over a range of environmental conditions. Using this dataset to train a statistical model creates an emulator of the underlying RTM. This is a condensed representation of the original RTM that facilitates rapid exploration of a broad range of environmental conditions and sensitivities. As an

illustration of this approach, we use the emulator to explore how varying the boundary conditions in the RTM describing the aquifer impacts the rates and volumes of mineral precipitation. A key result of this work is the recognition of an unanticipated dependency of pyrite precipitation on pCO_2 in the injection fluid due to the stoichiometry of the microbially-mediated sulfate reduction reaction. This complex relationship was made apparent by the emulator, while the underlying RTM was not specifically constructed to create such a feedback. We argue that this emulation approach to sensitivity analysis for RTMs may be useful in discovering such new coupled sensitives in geochemical systems and for designing experiments to optimise environmental remediation. Finally, we demonstrate that this approach can maximise specific mineral precipitation or dissolution reactions by using the emulator to find local maxima, which can be widely applied in environmental systems.

Synopsis

This study explores key factors regulating mineralization reactions in near surface environments revealed by a machine learning approach to reactive transport modelling.

1 Introduction

Reactive transport modelling has been extensively applied across a wide variety of environmental systems, providing a powerful means of quantifying, and even predicting, processes across Earth's (near) surface environments (Richter and DePaolo, 1987; Bain et al., 2000; Johnson et al., 2004; van Breukelen et al., 2004; Gaus et al., 2005; Torres et al., 2015; Li et al., 2017; Arora et al., 2020; Molins and Knabner, 2020; Rolle and Borgne, 2020; Druhan et al., 2020; Cama et al., 2020). Reactive transport models (RTMs) are constructed by combining multiple physical, chemical, and biological processes to simulate the behaviour of environmental systems. As applications and software have concurrently expanded (Steefel et al., 2015; Li et al. 2017; Maher & Mayer, 2019;

Druhan & Tournassatt, 2019), it is becoming increasingly common to explicitly calculate the rates of production and consumption for a variety of coexisting chemical species, as well as their equilibria with mineral phases, and their transport as they evolve in time and space. This type of multi-phase, multi-component RTM is a type of forward modelling where the results of the simulation emerge from a complex suite of interacting pathways, and hence the causes of observed behaviour are not always obvious.

RTMs are often designed to describe the behaviour of specific field sites and systems. Due to their process-based nature, designing RTMs requires selection of a suite of chemical reactions and transport mechanisms which are thought to dominate the geochemistry of the system over the scales of interest. However, the parameterisation of various selected processes is often not unique and can impact system behaviour (Williams et al., 2011; Martinez et al., 2014; Seigneur et al., 2021; Steefel et al., 2005a). To assess the impact of the choice of parameterisation and the values chosen for different parameters on model predictions, sensitivity analyses are generally performed (Malaguerra et al., 2013; Gatel et al., 2019). However, as RTMs become increasingly sophisticated, they incorporate disparate processes that can interact with each other in complex ways (Dwivedi et al., 2018; Hubbard et al., 2018, 2019; Maavara et al., 2021a, b; Dwivedi et al., 2017).

The sensitivity analysis of an RTM in application to a specific environmental system can elucidate the relative importance of specific interactions; for example, testing the solubility of mineral phases relative to changes in the solution chemistry. However, results might emerge that were not anticipated. These results might represent a real, but unexpected, interactions in which case the sensitivity analysis has yielded new insight into the system being modelled. Equally, the result might represent an incorrect interaction between two different processes that are known to act independently of each other, in which case the RTM can be improved. Unfortunately, due to the computational expense of many modern multi-component RTMs (e.g. Abd and Abushaikha, 2021;

66 Seigneur et al., 2021; Gharasoo et al., 2022), it is normally impractical to perform sensitivity analyses
67 in more than a few dimensions and it is up to the investigator to use their knowledge of the system to
68 choose which sensitivity analyses are necessary to explore (Steefel et al., 2005b). Ideally, we would
69 be able to systematically perform sensitivity analyses over many model parameters, considering how
70 model outputs vary as a function of multiple input parameters simultaneously (i.e. in a multivariate
71 way), while also lightening the computational burden that commonly occurs when using inverse
72 modelling approaches implemented by codes like PEST and iTOUGH2 (Doherty, 2004; Finsterle et
73 al., 2017). Such a capacity could direct future laboratory-based investigations to test whether these
74 model results are real-world phenomena, ultimately offering improved parameterisation of critical
75 components within the reaction network.

76 Here, we demonstrate a method for exploring a wide variety of potential model parameters by
77 adopting an emulator-based approach. Ours is not the first work to apply emulators to RTM
78 simulations. Notably, a rich vein of research based around replacing the geochemical solver in RTMs
79 with an emulator has emerged over the past few years (see Laloy and Jacques (2021) and Kyas et al.
80 (2022), among others). However, the work presented here is less concerned with speeding up
81 individual RTM simulations as it is with developing new methods to explore geochemical parameter
82 spaces. We also investigate the effect of changing geochemical parameters on the overall outcome of
83 RTM simulations, with an eye towards predicting system outcomes in real world scenarios. This is
84 similar in nature to recent work conducted by Ahmmed et al. (2021), which explores the ability of
85 different machine learning methods to predict the degree of mixing and the progression of a
86 simplified, generic reaction ($A + B \rightarrow C$) in a finite element simulation, and we extend the idea of
87 predicting the final state of a simulation to published RTMs describing real world systems.

88 Such emulation approaches in predicting the outcomes of physical systems have a long history,
89 including applications in physics-based animation (Grzeszczuk et al., 1998),complex multi-physics

simulators (Lu et al., 2021; Bianchi et al., 2016), climate models (Beucler et al., 2019; Krasnopolsky et al., 2005; Castruccio et al., 2014; Kashinath et al., 2021), and emulating fluid flow through Dolomite using a neural network (Li et al., 2022). In an emulator approach, the underlying physical system is approximated by a statistical model (the emulator) which can be evaluated more quickly than a conventional forward model. How this emulator is constructed varies by implementation and may encode assumptions about the underlying system to be modelled (e.g. conservation of energy (Beucler et al., 2019)). In this study, we are primarily interested in exploring and emulating the geochemical behaviour of RTMs; therefore we focus less on transport effects and restrict ourselves to emulating one dimensional RTMs.

In our implementation the emulator is built by training a Gradient Boosted Trees (GBT) regressor (Chen and He, 2015) on a synthetic dataset generated from the original RTM. By training such a GBT model on the synthetic dataset generated by the original RTM, we create an emulator of the original system. This emulation approach is general and can be applied to a wide range of RTMs, using "off the shelf" statistical libraries, requiring no special construction of the statistical model beyond the choice of some training parameters. This approach can identify the critical processes and parameters within RTMs and address the requirement for comprehensive, multivariate sensitivity analyses.

We first present a tool that automates creation of synthetic datasets: a Python wrapper for the RTM software CrunchTope (Druhan et al., 2013; Steefel et al., 2015), which we have named Omphalos. Omphalos edits and runs CrunchTope input files in an automated fashion, systematically changing model parameters according to user specification. It then records the output data, along with the corresponding model input parameters for later analysis. We then apply a machine learning method (Gradient Boosted Trees) to these recorded inputs and outputs to create a predicative model that can

reproduce RTM outputs based on the input variables, which we term a Reactivate Transport Emulator (RTE).

We suggest that our contribution to the development of Reactive Transport Emulators could be used to direct new experimental investigation to identify and corroborate predicted dependencies; providing multivariate analysis of RTMs and helping to identify effects that can, in the future, be considered explicitly when developing new RTMs. In pursuit of this goal, we demonstrate our emulator approach in application to an RTM built for biostimulation of a contaminated aquifer. We also show an additional application of this approach to efficiently predict the condition which maximises an RTM-predicted time-integrated rate over the set of chosen parameters. We also present, in the Supplement, another example in application to a deep-sea sediment column.

2 Description of the Case Study

2.1 Old Rifle Site, Colorado

The Old Rifle site is located near Rifle, Colorado, USA. The location historically hosted a vanadium and uranium ore processing facility, and the groundwater at the site remains high in aqueous uranium. Oxidised uranium (U(VI)) is fluid-mobile and highly toxic, while reduced uranium (U(IV)) is much less soluble and forms stable precipitates such as uraninite (UO_2) (Anderson et al., 2003; Wu et al., 2006; Dullies et al., 2010; Williams et al., 2011; Long et al., 2015). Thus, uranium reduction has been suggested as a means for remediating uranium contamination in groundwater. It has been shown that iron sulfide minerals ($FeS_{2(s)}$) aid the reduction of soluble U(VI) to insoluble U(IV) precipitates even after active remediation has ceased (Komlos et al., 2008; Moon et al., 2010; Bargar et al., 2013; Long et al., 2015; Bone et al., 2017).

134 The RTM published for Old Rifle, upon which the RTE is based, was originally created as a
135 comprehensive model of microbial sulfate reduction and sulfide precipitation in Old Rifle sediment
136 during stimulation of microbial activity by amendment with $C_2H_3O_2^-$ (Druhan et al., 2014) (for a
137 schematic illustration of this RTM, see Fig. S3). In this context, we choose to vary the influent
138 boundary condition chemistry, representing changes to the chemical composition of the artificial
139 groundwater injectate. The original experiment was designed to model microbial sulfate and iron
140 reduction in the sediment; therefore, we use net amorphous iron (II) sulfide ($FeS_{(am)}$), and pyrite
141 ($FeS_{2(s)}$) precipitation (both hereafter referred to simply as 'pyrite') as an observable that will record
142 the sensitivities of the model predictions to changes in the injection fluid. We also demonstrate the
143 utility of the emulator in predicting the chemical composition of the injection fluid that will
144 maximise the volume of pyrite precipitated in the sediments when amended with a labile organic
145 carbon source via injection wells.

146 **3 Methodology**

147 **3.1 General Strategy**

148 To explore the dependence of the RTM on the chosen environmental variables, we begin with a
149 Monte Carlo approach; we draw random values for each parameter and record the model output
150 under that randomised condition. We then fit a model to this Monte-Carlo-generated dataset using a
151 GBT regressor. This fitting results in a model (our emulator—RTE) that reproduces the complex
152 interdependencies of chemical species that are encoded in the original, underlying, RTM. This
153 emulator can be interrogated to examine the dependence of the RTM outputs on the originally chosen
154 environmental variables in an efficient, multivariate way. This way of performing sensitivity analyses
155 has the potential to give insight into trends and relationships that would not be apparent otherwise
156 and ultimately allows us to investigate the sensitivity of the model outputs with respect to the RTM's

original parameterisation. First, we will describe how we use the Monte Carlo approach to generate data and then how we fit a model to this data. The overall workflow is shown in Figure 1.

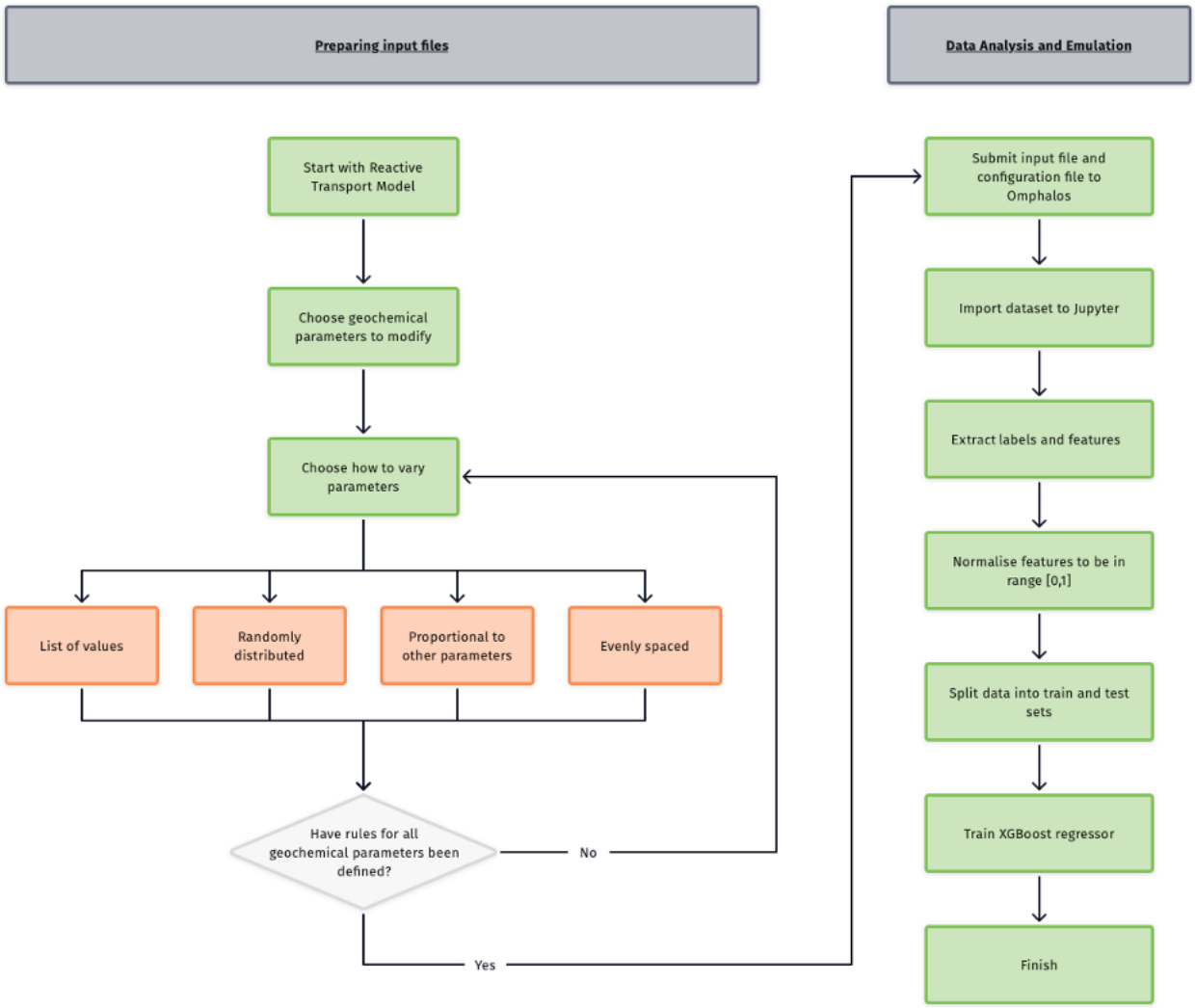

Figure 1: Flowchart describing the overall reactive transport emulation workflow developed in this study. It is divided into two key sections: preparation of the input reactive transport model for submission to Omphalos, and the analysis and emulation of the resultant data.

3.2 Generating Data

We use the open-source software CrunchTope as the reactive transport framework for the models in this study. To generate the synthetic datasets necessary for our approach, and given the time-consuming nature of generating a single point (requiring a complete run of the RTM, along with

modified boundary conditions), we developed a software package in Python to automate this process. This software package can manage the automatic generation and submission of unique input files to CrunchTope, as well as recording the output of each run, storing it in a manageable data structure for future use. Use of the software package is straight-forward, requiring the configuration of a single file listing which species/parameters are to be varied, and how they should be varied.

We have named this software package Omphalos (available for download—Sect. 6.1). Omphalos can be run on clusters using Simple Linux Utility for Resource Management (Yoo et al., 2003) to execute input files in parallel, or run locally with CrunchTope simulations on individual CPUs, which considerably reduces the time required to generate large datasets. Omphalos works by taking random values which are drawn from uniform distributions (other statistical distributions are possible) of the chosen variables, sampling the space evenly. This provides a complete dataset for training the emulator.

While the underlying principle of training emulators on synthetic data can be applied to any reactive transport code, currently the software used to implement the approach is only compatible with CrunchTope, because the input file reading and writing must be in a specific format. The approach is readily generalized, however, and the methodology could be applied to any RTM software (e.g. Geochemist's Workbench, TOUGHREACT), provided that the string input/output code is adapted for compatibility.

3.3 Application to Contaminated Aquifer Case Study

We begin by applying the emulation methodology to our case study. To create the dataset for training the emulator, we collected the results of 10,927 unique CrunchTope simulations based on the original RTM describing Old Rifle using Omphalos, drawing random concentrations for five chosen species (NH_4^+, SO_4^{2-}, Ca^{2+}, $C_2H_3O_2^-$, and pCO_2) in the boundary condition. Of these 10,927 runs, 9416

provide useable data because some runs fail to converge within the specified timeframe, or the geochemical condition generated cannot be charge balanced. Excluding these runs helps ensure that our dataset is kept realistic, because our RTM is built on a mechanistic understanding and implementation of the physical processes at work in the system that have been validated in some way. Therefore, runs that take excessively long to run are failing to converge in the simulation scheme of CrunchTope, and hence likely to be unphysical in some way. Similarly, runs that fail to speciate or charge balance indicate some kind of extreme physical condition that is unlikely to be realistic and so are excluded. The concentrations for NH_4^+, SO_4^{2-}, Ca^{2+}, and $C_2H_3O_2^-$ are varied between 0–30 mM. The pCO_2 is varied between 0–10 bar. We acknowledge that these ranges of concentrations are somewhat higher than those that occur in natural systems, but we extend the range to observe RTM behaviours at limiting concentrations. Related to this, it is possible for the dominant reaction mechanism in a system to change under differing conditions (e.g. the change in calcite dissolution mechanism as a function of pH (Dolgaleva et al., 2005)) and any such behaviour should be explicitly encoded into the RTM, otherwise the emulator may give invalid predictions under conditions that are far from the original model run. We have assumed in this study that the mechanisms governing the precipitation of pyrite do not change under very low or very high concentrations of these species.

The injection fluid was constrained at pH 7.2. This constraint, in conjunction with the concentration of various species iterated in Omphalos, speciates according to CrunchTope's internal speciation calculation. Therefore, for example, although the total amount of SO_4^{2-} in the injection will be iterated in, and dictated by, Omphalos, the amount that speciates into other aqueous complexes (i.e. secondary species) like HSO_4^- or $H_2SO_{4(aq)}$ is controlled by CrunchTope. For the sake of simplicity, we will report the input concentration, not the concentration after speciation.

The RTM describing Old Rifle has 100 grid cells with a size of 1 cm. Each run of the RTM took approximately 90 seconds, so the total time to generate the dataset was roughly four hours when run on a remote machine with 200 CPUs. The number of runs was chosen as a balance between what was computationally tractable and the ability of the emulator to achieve a good fit. We have intentionally chosen to vary some chemical species in the influent boundary condition that do not play an obvious role in the mineral precipitation process we are particularly interested in, namely, the precipitation of pyrite in Old Rifle sediments (e.g., NH_4^+ or Ca^{2+}, respectively). We did this to see if we can use the emulator to detect behaviour in the RTM beyond what we might initially hypothesise.

3.4 Fitting the emulator

We implement the GBT regressor using XGBoost (Chen and He, 2015) in Python. The code for fitting the models is available in the Supplement. For a precis on GBT models, see the supplement Sect. S1.2.

3.4.1 Data Strategy

Data generated by Omphalos was imported into a Jupyter notebook environment from the .pkl output file. There are 9416 different input file runs in this data file, having excluded 1511 runs on the grounds of them being unrealistic, as discussed previously. The relevant data was indexed out of the data structure; in our case this meant the concentrations of NH_4^+, SO_4^{2-}, Ca^{2+}, and $C_2H_3O_2^-$ in the boundary condition, as well as value of pCO_2. This results in a 5×9416 array of floating-point numbers for the features. Each feature was then normalised to be in the range 0 to 1 for training. For example, values of SO_4^{2-} concentration in the simulations were drawn randomly between 0 and 30 mM, so all SO_4^{2-} concentrations were divided through by 30 to have values in the range 0–1. We did this to improve the training performance of the GBT model over different datasets (i.e. so that the same GBT model can be applied to both the Old Rifle case study, and our supplementary case study of ODP Site 1086 (see Supplement, Sect. 3).

Similarly, the relevant data was also extracted from the data file: for each cell in the gridded RTM, we calculated the net pyrite precipitation over the course of the simulation, and then summed this value over the column to get the net pyrite precipitated across the domain. This results in a 1×9416 array of floating-point labels to be predicted from the feature array. We scale this feature array by a factor of 1×10^4 to avoid issues with small floating-point numbers in XGBoost.

We prepared these data for training the GBT regressor with a hold-out strategy using the scikitlearn.train_test_split method, keeping 10% of the dataset back for validating the model. Data was split randomly within the dataset. This means that 8474 randomly selected data points were used to train the model and 942 randomly selected data points were used to test it by using the model to predict a value based on the held back data and comparing the prediction to the true value.

3.4.2 Training Strategy

We use the data points generated by Omphalos to train an XGBoost regressor using squared error as the loss function to predict the amount of pyrite precipitated in the column as a function of varied species concentrations in the boundary condition. Squared log loss, and pseudo-Huber error were also tried but squared loss performed best overall. Training curves showing the testing and training loss as training progressions are given the supplement, Fig. S2.

Hyperparameter choices for the model are explained and given in the supplement, Sect. S1.3, Table S1. The choice of hyperparameters is the same for each emulator model, and we are able to achieve high quality fits using the default XGBoost regularisations, only changing a few settings relating to tree growth policy. While it is a known problem in machine learning that the choice of optimal hyperparameter is dependent on the data being modelled (Claesen and De Moor, 2015), it appears that in the context of these RTEs, the hyperparameters chosen give a good fit for both Old Rifle and our supplementary case study of ODP Site 1086: datasets describing very different natural

environments, with different length and time scales. This makes the workflow applicable across a wide variety of reactive transport modelling domains.

It is possible that with more complex hyperparameter tuning, better emulator fits may be achieved, but for the purposes outline in this paper, we suggest that this automated optimisation of a subset of the available hyperparameters is sufficient, and represents a balance between emulator fit, generalisability across differing RTMs, and time spent by the user.

3.4.3 Model metrics

We report our model goodness-of-fit to the underlying dataset as the R^2 value for the model, using the built in XGBRegressor.score() method from XGBoosts Scikit-Learn API, on both the original training dataset, as well as the 10% of the dataset held back for validation, show in Table 1. We also report the root-mean-square error (RMSE) over both the training and validation datasets, calculated using the Booster.eval() method. Training curves are shown in the supplementary, Figure S2. We report the same metrics for our second model in the supplementary, Table S4.

Dataset	R^2	RMSE
Training	0.99996	1.95012
Validation	0.99964	5.65848

Table 1: Training and validation metrics for the XGBoost regressor model fit to the Old Rifle dataset. R^2 represent a normalized measure of the fit quality, with the best possible score being 1. RMSE is the root-mean-squared error in the predictions, where we recall that in data pre-processing that the values to be predicted were multiplied by a factor of 10^4, and so the RMSE should be divided by that factor when assessing the average error on data presented in Figures 2, 3, and 4.

4 Results and discussion

4.1 Application to the Old Rifle Site

The synthetic data generated using Omphalos to interrogate the underlying RTM are shown in Fig. 2, colour mapped by the pCO_2 with which the injectate solution is in equilibrium. The colour mapping helps visualise how variability in the precipitated volume of pyrite over the 43-day RTM simulation might be considered in conjunction with other model parameters. Ultimately, pyrite forms because aqueous hydrogen sulfide, produced through microbial sulfate reduction, reacts with reduced ferrous iron (Fe(II)) to form pyrite. Thus, we aim to explore the interdependencies between the mechanisms driving microbial sulfate reduction and the subsequent precipitation of pyrite, as they emerge due to variations in injectate chemical composition.

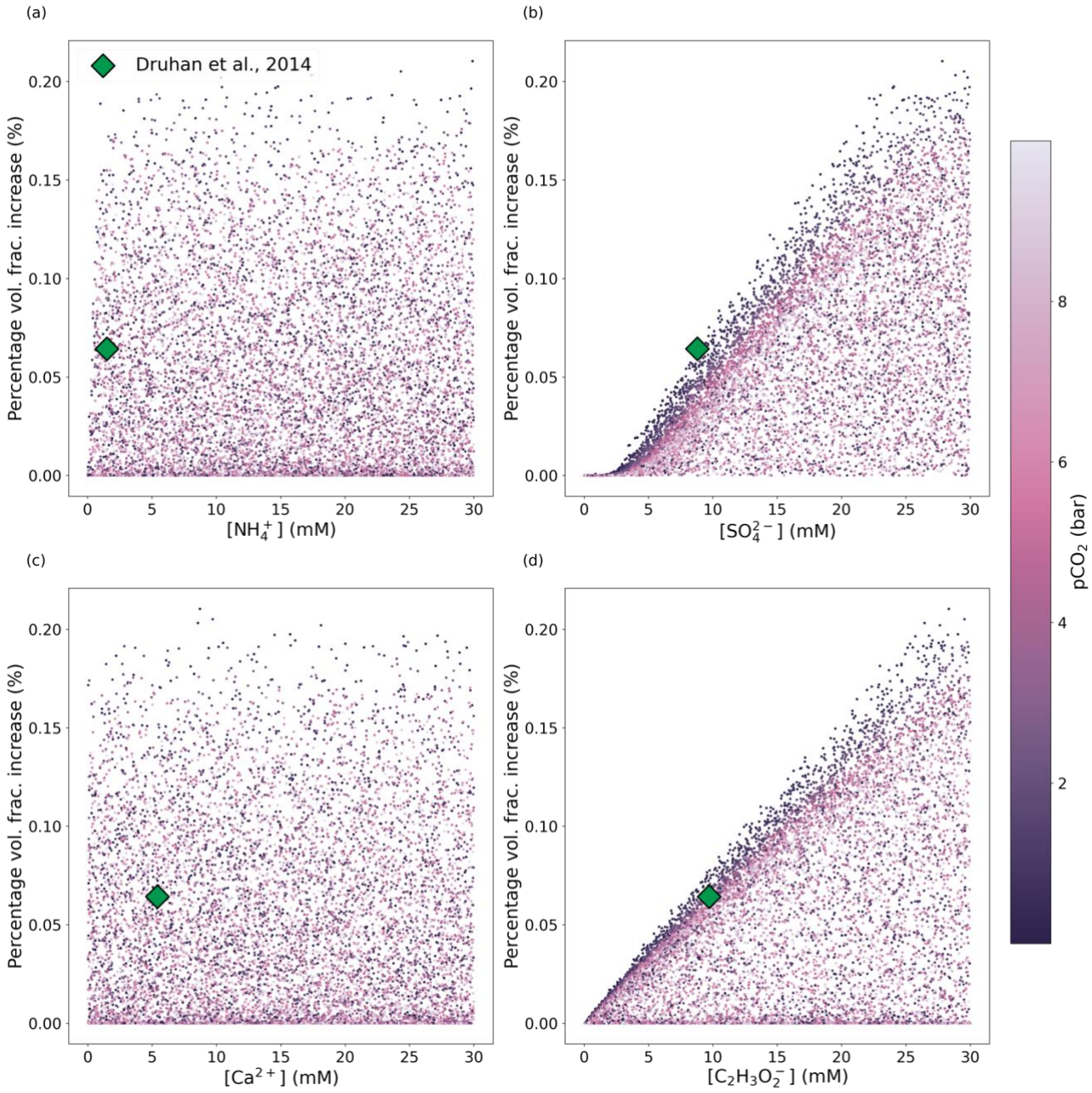

Figure 2: Scatter plots of chemical concentrations in the fluid injectate (influent boundary condition) for an RTM adapted to Old Rifle sediments colour-mapped by the pCO_2 with which the inlet boundary condition is in equilibrium. The dataset comprises 9416 points generated by drawing concentrations for all five species independently from uniform random distributions, with the corresponding net increase in pyrite volume fraction precipitated (y-axis) calculated by running the Old Rifle RTM designed by Druhan et al. (2014) with the randomised influent

boundary condition. The green diamond indicates the net pyrite volume fraction generated from the original boundary condition used in Druhan et al. (2014).

We then train the emulator on this synthetic dataset. Fitting a GBT regressor to the data in Fig. 2 means Fig. 3 can be generated by the emulator. This figure shows how the emulator predicts the change in pyrite volume fraction as the concentration of each of the species in the injection fluid is varied (other species in the RTM not defined as variables in this study are held constant at values reported by Druhan et al. (2014)). We stress that the RTM results shown in Fig. 3 are not part of the training dataset, and that the emulator has not been exposed to these exact values. This demonstrates the capability of the emulator to reproduce the underlying RTM itself. For example, Fig. 2A suggests visually that the concentration of NH_4^+ in the system is uncorrelated with net pyrite precipitation at the Old Rifle Site. Fig. 3A confirms this lack of dependence on NH_4^+, capturing the correct trend (with some noise) although being slightly offset. This slight offset also applies to Fig. 3C in the fit of the Ca^{2+} dependence. We suggest that these slight offsets to the fits in the cases of the weakly or uncorrelated variables is due to the emulator preferentially capturing stronger dependencies and slightly drawing down the predicated variable on average.

In contrast to the minimal impact that changing NH_4^+ concentration has on pyrite precipitation, $C_2H_3O_2^-$ and SO_4^{2-} concentrations correlate strongly with net pyrite precipitation. This is as expected in a system where $C_2H_3O_2^-$, which is the electron donor for microbial sulfate reduction, enables sulfate to be reduced to sulfide and thus drive pyrite precipitation in the presence of Fe(II). Approximately 20 days after $C_2H_3O_2^-$ amendment, microbial sulfate reduction takes over from dissimilatory iron reduction as the dominant process consuming $C_2H_3O_2^-$. As microbial sulfate reduction requires eight-times the number of electrons per mole of SO_4^{2-} reduced than dissimilatory iron reduction requires (per mole of iron reduced), the electron donor ($C_2H_3O_2^-$) begins to be rapidly

consumed, whereas during dissimilatory iron reduction it was effectively in excess. As a result of this new scarcity of $C_2H_3O_2^-$, the rate of dissimilatory iron reduction drops and so does the concentration of Fe(II). However, dissimilatory iron reduction is still active in the column, releasing a small—but non-zero—flux of aqueous Fe(II) that allows for continued pyrite precipitation. The emulator interprets this as Fe(II) being 'always' available in this system, and thus predicts that pyrite precipitation can scale linearly with SO_4^{2-} and $C_2H_3O_2^-$, as shown in Fig, 4A. The sediment itself would need to contain abundant ferrihydrite, goethite, or another bioavailable ferri(hydr)oxide for this reduction to continue indefinitely; this may not be the case. This highlights the need for the range of parameters sampled when training the emulator to be sufficiently wide to capture all the RTM behaviour, otherwise it may extrapolate and "learn" incorrect assumptions about the system: in this case that bioavailable iron never limits dissimilatory iron reduction. One solution would be to expand the range over which concentrations are drawn to reach the limit where iron-bearing mineral volume fraction becomes a limiting factor so that the model can learn what happens when this occurs.

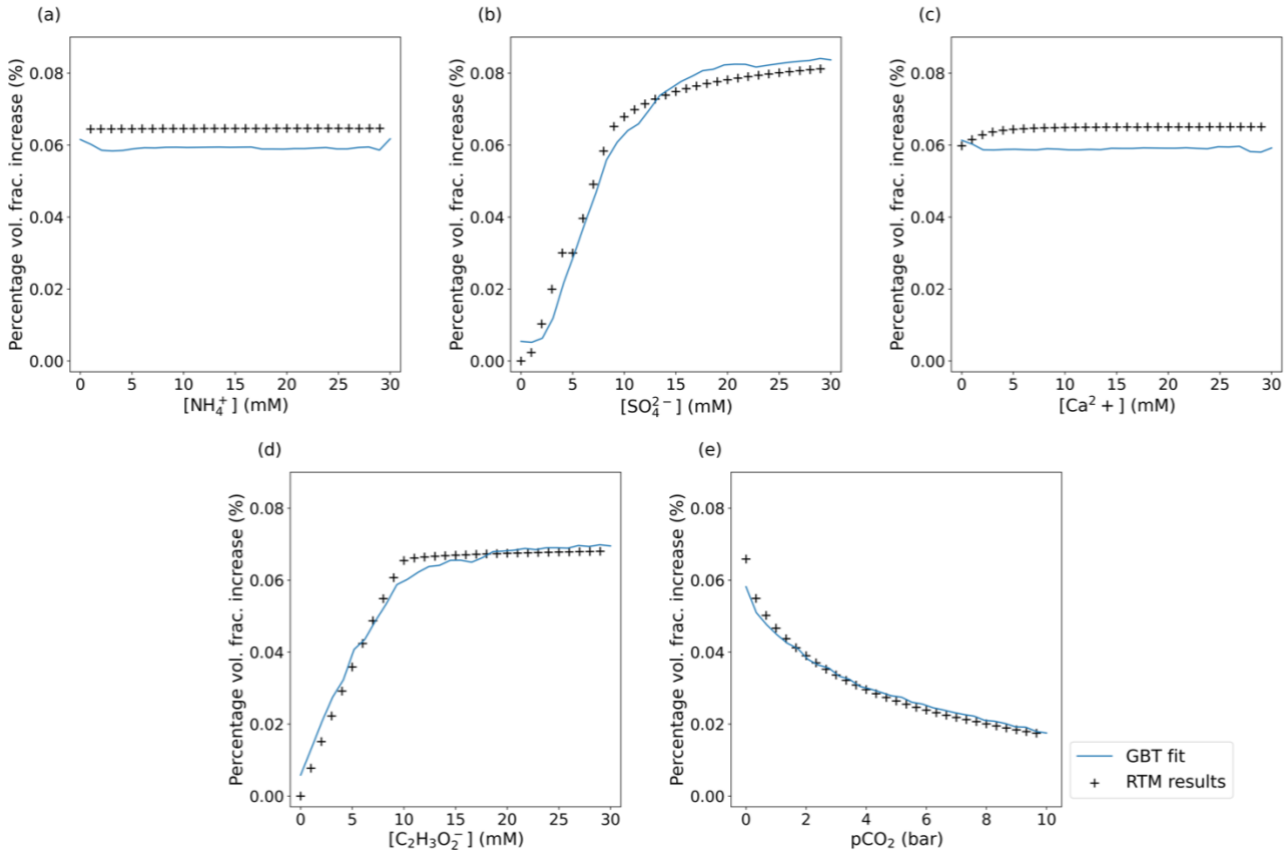

Figure 3: Plots of the GBT model fit (blue line) plotted over the results from the underlying RTM (black + symbols) when interrogated with the same input parameters (which are taken as ground truth). Each plot shows the net volume fraction due to pyrite precipitation as a percentage of the initial volume fraction of the sediment as each parameter is varied while all other parameters are held at the values used in the original experiment by Druhan et al. (2014). The emulator (blue line) captures the overall trends in the data. The lack of smoothness in the emulator predications arises from the inability to encode this as a condition in XGBoost and the discreet nature of the decision tree algorithm. The RTM results compared to here are not part of the training dataset, and so the emulator has not been exposed to those exact values.

We also note that our emulator suggests that increasing pCO_2 leads to decreased pyrite precipitation (Figure 4E), a relationship that may not have been apparent in a single run of the RTM. Three-dimensional visualisation of the data confirms that the pyrite-volume-fraction-change varies as a

345 function of pCO_2 net pyrite precipitated decreasing as pCO_2 increases (Fig. 4B and Fig. 4C). This
346 three-dimensional visualisation allows us to see that the gradient of the pyrite-volume-fraction-
347 change with respect to SO_4^{2-} and $C_2H_3O_2^-$ is itself a function of pCO_2 and flattens as pCO_2 increases.
348 To understand why the gradient changes, we must first understand why pCO_2 affects the amount of
349 pyrite precipitated in the first place.

350 Sediment samples from Old Rifle are initially poised for dissimilatory iron reduction and there is a
351 sizeable community of iron-reducing bacteria naturally present in the system. The background
352 sulfate-reducing microbial community is initially relatively small and thus, for microbial sulfate
353 reduction to proceed at significant rates, the mass of sulfate-reducing bacteria must first increase. In
354 the original experiment by Druhan et al. (2014), the sulfate-reducing biomass begins reaching a size
355 where it can start consuming large quantities of $C_2H_3O_2^-$ around day 20 of the experiment. This
356 biomass growth is modelled in CrunchTope using a Monod-biomass rate law (Jin and Bethke, 2005),
357 which has both an anabolic and catabolic component. In the formulation of this Monod-Biomass rate
358 law as implemented in CrunchTope, the thermodynamic term (Gibbs free energy of the reaction) is
359 calculated exclusively using the catabolic pathway. The catabolic pathway for this reaction (in terms
360 of the exchange of one electron) is given below in Equation (4.1), and the form of the Gibbs free
361 energy is this context is given in Equation (4.2) (we take the phosphorylation potential to be 0, and
362 the average stoichiometric number to be 1, see derivation in Jin and Bethke (2005) for further
363 details).

364 $$\frac{1}{8}SO_4^{2-}+\frac{1}{8}C_2H_3O_2^-+\frac{3}{8}H^+\rightarrow\frac{1}{8}H_2S_{(aq)}+\frac{1}{4}CO_{2(aq)}+\frac{1}{4}H_2O \quad (4.1)$$

365 $$\Delta G=\mathcal{R}\mathcal{T}\ln\left(\frac{[CO_{2(aq)}]^{\frac{1}{4}}[H_2S_{(aq)}]^{\frac{1}{8}}}{\left[SO_4^{2-}\right]^{\frac{1}{8}}[C_2H_3O_2^-]^{\frac{1}{8}}[H^+]^{\frac{3}{8}}}\right) \quad (4.2)$$

366 Taking this form for the Gibbs free energy of the reaction and substituting it into the thermodynamic
367 term of the reaction rate calculation as implemented in CrunchTope (Steefel et al., 2015) gives
368 Equation (4.3) below describing the rate of microbial sulfate reduction in the Rifle RTM.

369 $$R_{MB}=k_{max}B\frac{[C_2H_3O_2^-]}{[C_2H_3O_2^-]+K_{half[\mathrm{Ace}]}}\frac{\left[SO_4^{2-}\right]}{\left[SO_4^{2-}\right]+K_{half\left[SO_4^{2-}\right]}}F_T \quad (4.3)$$

370 where

371 $$\boldsymbol{F_T}=\left(\mathbf{1}-\frac{[CO_{2(aq)}]^{\frac{1}{4}}[H_2S_{(aq)}]^{\frac{1}{8}}}{\left[SO_4^{2-}\right]^{\frac{1}{8}}[C_2H_3O_2^-]^{\frac{1}{8}}[H^+]^{\frac{3}{8}}}\right) \quad (4.4)$$

372 R_{MB} is the overall rate of microbial sulfate reduction, k_{max} the rate constant for microbial sulfate
373 reduction, B is the biomass concentration, and $K_{half[\mathrm{X}]}$ is a half-saturation constant. The two Monod
374 kinetic factors for the electron donor ($C_2H_3O_2^-$) and the electron acceptor (SO_4^{2-}) are referred to as F_D
375 and F_A respectively (Jin and Bethke, 2003, 2005, 2007). Equation (4.4) illustrates the underlying
376 relationship between pCO_2 in the injectate solution and the resulting accumulation of pyrite. As the
377 pCO_2 of the injectate increases, the F_T term becomes smaller, inhibiting the overall rate of microbial
378 sulfate reduction (Fig. S5). Consequently, biomass growth is also inhibited, and the rate of microbial
379 sulfate reduction is never high enough to produce the concentration of $H_2S_{(aq)}$ required for significant
380 pyrite precipitation. This explains why the model suggests that the gradient of the pyrite volume
381 precipitated with respect to both $C_2H_3O_2^-$ and SO_4^{2-} varies as a function of pCO_2 in the injectate.

When pCO_2 is low and both SO_4^{2-} and $C_2H_3O_2^-$ are large with respect to their half saturation constants (Equation (4.4)), the overall Monod-biomass rate law will approach Bk_{max}.

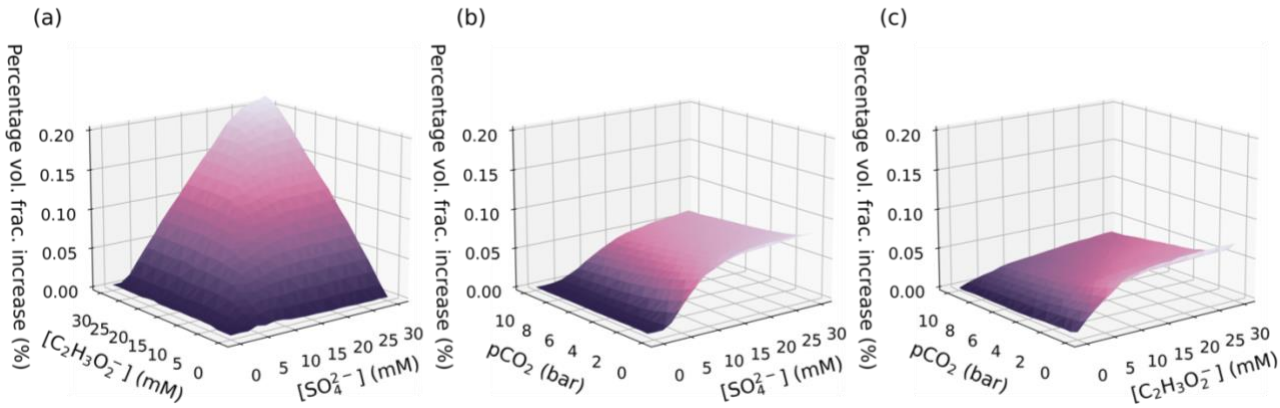

Figure 4: A selection of the GBT model predictions of the percentage volume fraction increase due to pyrite precipitation as a result of varying two parameters simultaneously for selected pairs of variables. Other model parameters are held at the values used in Druhan et al. (2014). The remaining variable-pair plots are provided in Fig. S4.

This dependence emerged somewhat unexpectedly from the emulator, as one would not inherently expect a relationship between injectate pCO_2 and SO_4^{2-} reduction rates, yet it agrees with results previously reported by Jin and Kirk (2016, 2018) as well as Paper et al. (2021). These studies related the influence of pCO_2 and pH to the rate of microbial reactions, both *in vitro*, *in situ*, and *in silico.* We suggest that our type of analysis could be used to direct future lab and field work to test hypotheses suggested by the results generated by running the emulator.

This analysis also explains some of the features observed in Fig. 4A: the gradients of $C_2H_3O_2^-$ and SO_4^{2-} are coupled in such a way as to indicate that if one is in excess, then the other becomes limiting in the production of $H_2S_{(aq)}$ and hence the precipitation of pyrite. However, the limiting behaviour when both are in excess seems to indicate that given enough SO_4^{2-} and $C_2H_3O_2^-$, pyrite precipitation can continue indefinitely assuming suitably low pCO_2. Given this prediction, it is sensible to check

whether, at such high levels of SO_4^{2-} and $C_2H_3O_2^-$ as the model suggests for maximum pyrite precipitation, there is indeed enough Fe(II) available in the system to precipitate pyrite: this is a second potential dependence as mentioned above.

Lastly, the model can be interrogated in all 5 dimensions and the amendment fluid composition that corresponds to the largest net pyrite precipitation over the modelled interval can be determined. We do this simple optimisation by evaluating the emulator at regular intervals across all 5 dimensions at intervals of ~2 mM (intervals of ~0.67 mM for pCO_2) in the range that the emulator was trained on (0–30 mM, except for pCO_2 which has a range of 0–10 mM). This corresponds to checking 759375 different boundary conditions to find which boundary condition results in the most net pyrite precipitation and takes ~7 minutes. This amendment composition is shown in Table S1. The total change in volume fraction due to pyrite precipitation predicted by the emulator is 0.143 and the actual RTM modelled precipitation when this boundary condition is used is 0.150. There is a 4.7% absolute error on the net pyrite volume fractions change predicted by the emulator when compared to the actual net pyrite precipitation calculated by the RTM. This error is inherent in statistical learning techniques but can be further mitigated with larger training datasets, in conjunction with different emulator training hyperparameterisations: an area for future improvement to the methodology. These optimised conditions represent an almost four-fold increase in the amount of pyrite precipitated in the original RTM for Old Rifle (Druhan et al., 2014).

4.2 Advantages and drawbacks of the emulation approach

In this study, 9416 individual RTM simulations were used to train a GBT regression model to predict a specific model output, in this case net pyrite precipitation. This emulator is a reduced representation of the complex system of equations in the underlying RTM, having a faster computational time but introducing some prediction errors. We now discuss the key advantages and drawbacks of this emulation approach.

4.3 Advantages of the emulation approach

9416 RTM runs were used to train the emulator (the data shown in Fig. 2). This number of runs could instead be used to perform a sensitivity analysis in all five variables at a spacing of ~4.8 mM between points by directly interrogating the simulator. What then, is the advantage of the emulation approach, if the same information can be visualised from discreet runs of the original RTM without having to go to the extra effort to train the model, which introduces prediction errors? The key advantages are outlined below.

4.3.1 Advantages over directly interrogating the simulator

The first and most obvious advantage is the lack of a need for an explicit interpolation scheme. Correlations generated by directly plotting simulator results lead to data points lying on a grid of finite resolution. If intermediate values on this grid were to be determined, an explicit interpolation scheme would have to be applied, which would introduce errors of its own that would then need to be quantified. Furthermore, an improvement in the interpolation scheme would come at the expense of adding one extra point to the grid in each dimension: in the context of Old Rifle this is an extra 9031 data points ($7^5 - 6^5 = 9031$ going from a 5D grid of 6 points in all directions to 7) roughly doubling the dataset size. In contrast, since any number of points can be submitted to the emulator for inference, concerns relating dataset size to sampling resolution are assuaged. Beyond that, the errors in the model fit are already quantified during training.

More broadly, to explore the dataspace, emulators are extremely fast compared to simulators. The time for a single query of the emulator is on the order of milliseconds rather than the seconds/minutes/hours for a single forward RTM simulation. This allows the emulator to be used as a tool for efficiently exploring the simulator by rapidly developing intuition for the space itself and how the simulator behaves in different circumstances. Furthermore, emulator models are easy to

distribute and share with collaborators. Model weights can be published directly or distributed as standalone files. This means that a well-trained emulator can be made once and then the encoded data shared.

Lastly, performing a direct interrogation of the simulator requires choices of parameters and ranges, and results in a grid of points over the region of interest at limited resolution. A similar procedure must be undertaken when creating a dataset to train the emulator, in so far as ranges and parameters of interest must be chosen. However, the dataset can always be further added to in a straightforward manner, further drawing from the random distribution to increase the size of the dataset and thus improve model performance. With both approaches, using Omphalos means that the data generation process can be parallelised and using high-performance computing facilities can reduce the computational expense of interrogating the simulator. This means that all the computational expense is upfront in both cases since the emulator need only be fit once.

The advantages we outline make the case for the emulator as a tool to be used in conjunction with the RTM, rather than a replacement for it. The alacrity with which the emulator can be interrogated means that it is an invaluable tool for investigating RTM behaviour in multiple dimensions. Further to this, the ability to evaluate the state of a system after a fixed period of time makes the emulator approach ideally suited for modelling more complex time-series models with time varying boundary conditions: instead of having to run the RTM forward each time the system changes boundary conditions, the emulator can be interrogated for the expected result given the system's current state from the previous regime.

4.3.2 Using emulators to identify new feedbacks

As modern RTMs grow in sophistication and complexity, they increasingly draw on large suites of chemical and mineralogical information from vast databases, which constitute large sets of non-linear

equations all coupled through transport and fluid chemistry. While it is true that for a sufficiently simple model, coupled geochemical behavior could be deduced by reasoning about the governing equation of the systems, for a large, modern RTM it is inevitable that during development some feedbacks will be overlooked.

Emulation makes sensitivity analysis for RTMs simple and allows us to identify correlations and interactions among parameters that would otherwise be difficult to anticipate by allowing an investigator to quickly test a wide variety of hypotheses. We demonstrate this in the case of Old Rifle by identifying the CO_2 dependency of microbially mediated reactions (Bethke et al., 2011; Jin and Kirk, 2016, 2018; Paper et al., 2021). This ability to elucidate unexpected but key model dependencies and sensitivities could prove invaluable in helping direct RTM development.

4.3.3 Application to Bayesian optimization

A critical advantage of the technique proposed here is that emulation is an essential part of Bayesian optimization. Bayesian optimisation is an approach for finding global maxima and minima in systems whose objective function is expensive to evaluate and does not return the gradients of that function (of which RTMs are an example) (Frazier, 2018). Bayesian optimisation works by applying an acquisition function that calculates the point that will give the most information about the function that requires optimisation. An emulator is then fit using these data points selected by the acquisition function and the emulator is updated with a new point each iteration. In this way, the optimiser balances exploitation of known optima, and exploration of unevaluated regions of the function. Such an approach can find the global maximum with relatively few evaluations of the RTM.

This study lays the groundwork for future application of Bayesian optimization to highly dimensioned RTMs, potentially allowing for effective optimization over many different (twenty or more) parameters at once. By demonstrating that broad (but local) fits to the RTM with an emulator

493 are possible, we have demonstrated that a GBT regressor can be used as an emulator informing a
494 Bayesian optimization algorithm in this context. This allows for a constellation of local fits in a
495 highly dimensioned space as the algorithm searches for the global optimum in problems that would
496 otherwise be computationally intractable. Bayesian optimisation could even be applied, with a
497 suitable loss function, to optimise for multiple objectives at once (subject to trade-offs among
498 objectives).

499 **4.4 Disadvantages of the emulation approach**

500 This emulation approach relies on the relative computational inexpensiveness of the RTM. In
501 situations where the underlying model is expensive or time-consuming to evaluate, and
502 computational resources are limited, this modelling approach becomes unfeasible. However, this
503 issue of computational expense can be allayed by the parallelised generation of data alluded to earlier
504 and only the most expensive RTMs would be intractable for a full emulator fit if this technique was
505 deployed correctly, and even in this extreme case, Bayesian optimisation would still be possible.

506 Additionally, caution is needed when choosing the ranges over which the parameters will be drawn
507 from the uniform random distributions. Key considerations include the number of points being
508 generated relative to the size of the space being covered—a denser cluster of training data will result
509 in a tighter fit, at the expense of range. Conversely, too small of a range and the emulator will not
510 capture key behaviour, or be unable to learn about simulator edge cases, as discussed above with
511 respect to the bioavailable iron in the Old Rifle RTM.

512 **4.5 Choice of learning algorithm**

513 Gradient boosted trees outperformed other machine-learning methods that we tested while building
514 the emulators, such as Gaussian process regression. The downsides of GBT include the lack of ability
515 to encode smoothness to preclude sharp discontinuities in the concentration-precipitation space or

other such prior assumptions. Furthermore, a low root mean squared error over the entire model fit region does not necessarily imply a good fit globally; it may be that there are some regions of good fit and other regions of poor fit which make up an acceptable root mean square error over the whole space.

4.6 The effect of scale on emulator predictions

Our case study relies on the capacity of CrunchTope to predict changes in mineral volume fraction. Therefore, the errors in the predictions, and hence the utility of the approach, ultimately depend on the scale of the system being modelled and thus the sensitivity to what could be very small changes in mineral volume fraction.

When analysing the emulator to investigate how different processes in the underlying RTM affect each other, we are primarily considering an issue of whether the emulator can correctly learn the underlying model behaviour. We are also considering whether the emulator can capture the behaviour in the output variables with respect to a changing subset of RTM parameters (some of which we may not have expected at the outset). In this use-case, the emulator is largely concerned with trends and gradients; Figs. 3, 4, S4, S8, and S9 show that this is accurately reported in all case studies. Comparing the case study considered in this paper to the additional case study presented in the Supplement we see that they are discretised at different scales (2 m and 1 cm for the deep-sea sediment column and Old Rifle respectively). However, the emulator for each RTM has root mean squared error over the dataset (and hence absolute error in prediction) of the same order of magnitude. This implies that the error in absolute volume precipitated that each model predicts is different. However, the analysis of the trends and interactions emerging from both RTMs is equally valid in both cases.

When concerned with the optimisation capabilities of the emulator, the absolute value of the optimised quantity and hence the model scale must be considered. In large-scale systems, such as weathering of the critical zone, the error in the volume fraction change (5.5×10^{-5} for pyrite) is below the resolution of measurement techniques for mineral abundance (e.g. XRD and SEM—(Gu et al., 2020)). However, in smaller-scale systems where the microscale environment becomes increasingly important, these errors in volume fraction become much harder to ignore. For example, in the RTM experiments exploring the effects of scale on simulating mineral dissolution in porous media described by Jung and Navarre-Sitchler (2018), significant errors in changes in predicted volume fraction would propagate into calculated dissolution/precipitation rates, losing sensitivity in the results.

4.7 Extension to multiple outputs

Multiple output regression (the prediction of a vector of outputs, rather than a single label) is experimentally available in XGBoost and supported by other machine learning implementations that we explored, including GPFlow for Gaussian process regression. Given that our approach is currently limited to the prediction of one label-per-emulator trained, the availability of regressors that can predict more than one label 'off the shelf' will greatly improve the utility of reactive transport emulation. The prediction of multiple outputs simultaneously will expand the scope of analysis to investigate the interaction of modelled processes in multiple outputs at once. In the context of optimisation problems, one possible application of the emulator like this could be to maximise mineral precipitation in one region of a system while trying to maximise dissolution in another region.

4.8 Improvements to the model

This proof-of-concept model demonstrates the fitting of an emulator over a relatively small range of environmental parameters. Future work will involve expanding the scope of the emulators both in terms of the number of parameters being varied, but also the range over which they are varied, so the whole behaviour of the underlying model can be captured with more accuracy. There is also scope for adding time dependency to the GBT modelling approach, to predict a time series of intermediate RTM states during the evolution of geochemical systems.

4.9 Potential applications

Our emulator approach is flexible; any quantity recorded by an RTM can be used as a target variable, and so the behaviour of any RTM output can be explored in detail to evaluate the model formulation. The behaviour of the system in response to the variation of any parameter under any other set of conditions can be projected out of the model and plotted in a straight-forward manner. This approach can be extended to two or even three dimensions and time series thereof and ultimately the emulator can be interrogated for local maxima and minima to solve optimisation problems. This approach has potential applications in industry and in environmental remediation where the chemical composition of amendments can be predicted using an underlying reactive transport simulation, provided that that system is well understood.

Omphalos also has utility outside of generating datasets for emulation; its automated submission of CrunchTope input files means it can be used to systematically explore sets of input variables in an easy way, simply by editing the Omphalos configuration file.

5 Conclusions

We have presented an emulator based approach for interrogating and understanding multi-component RTMs. By building an emulator of an RTM that captures the multidimensional nature of the underlying model we have demonstrated that such an approach can be used as a tool for performing

global sensitivity analyses on RTMs. This allows us to investigate behaviour arising from the interaction among the many disparate processes that comprise RTMs. For example, we investigated how the Monod-biomass parameterisation of microbial sulfate reduction interacted with the mechanism of pyrite precipitation. In this example, pyrite precipitation was inhibited when there was an excess of CO_2 in the column because the catabolic pathway was partially dependent on CO_2 concentration. This prevented the growth of sulfate reducing biomass, ultimately curtailing the production of hydrogen sulfide required for pyrite precipitation. This behaviour reproduced results previously reported by Jin and Kirk (2016, 2018), and suggest that emulation approaches have utility in discovering unexpected, but nonetheless real, model behaviours, potentially directing future lab and field work.

The methodology we have laid out is flexible; any quantity recorded by an RTM can be used as a target variable, and so the behaviour of any RTM output can be explored in detail to evaluate the model formulation. The behaviour of the system in response to the variation of any parameter under any other set of conditions can be projected out of the model and plot in a straight-forward manner. Emulator approaches can be extended to two or even three dimensions and ultimately the emulator can be interrogated for local maxima and minima to solve optimisation problems. We suggest that emulator based approaches to exploring RTMs have potential applications in industry and in environmental remediation where the chemical composition of amendments can be predicted using an underlying reactive transport simulation, provided that that system is well understood. The presentation of this optimisation process to Old Rifle (and to ODP Site 1086, see supplementary) represents a proof of concept.

6 Code availability

6.1 Omphalos

Omphalos is available on GitHub and Zenodo. Please note you must provide your own CrunchTope executable.

https://github.com/a-fotherby/Omphalos

https://doi.org/10.5281/zenodo.7113298

6.2 GBT Models

Jupyter notebooks for fitting the GBT models and plotting the figures are available on GitHub, and a permanent record is available on Zenodo.

https://github.com/a-fotherby/dissertation_xgboost.

https://doi.org/10.5281/zenodo.7113323

7 Data availability

The data used is available on GitHub and Zenodo.

https://github.com/a-fotherby/GMD_2022

https://doi.org/10.5281/zenodo.7113379

8 Supplement

Codebase for Omphalos. Model fitting code. Schematic figures of decision tree and the Old Rifle RTM. Table of predicted optimal values for precipitating pyrite at Old Rifle. Convergence behaviour of the GBT regressors. Additional co-dependency plots for Old Rifle. Figure showing the effect of rate law choice on CO_2 dependency in the Old Rifle RTM. Supplementary Case Study detailing application to a deep-sea sediment column. Description of XGBoost implementation.

9 Author contribution

AF and HJB conceived of the study. AF wrote the codebase and conducted the experiments. AF prepared the manuscript with contributions from all co-authors.

10 Competing interests

The authors declare that they have no conflict of interest.

11 Acknowledgements

The work was supported by NERC NE/R013519/1 to HJB and by a call for International Emerging Actions granted by the CNRS (TELEMAART: Trace ELEments and inverse Models: Advancing Applications of Reactive Transport models) to JLD. This work was also funded by ICA\R1\1801227 from the Royal Society to AVT.

12 Bibliography

Abd, A. S. and Abushaikha, A. S.: Reactive transport in porous media: a review of recent mathematical efforts in modeling geochemical reactions in petroleum subsurface reservoirs, SN Appl. Sci., 3, 401, https://doi.org/10.1007/s42452-021-04396-9, 2021.

Ahmmed, B., Mudunuru, M. K., Karra, S., James, S. C., and Vesselinov, V. V.: A comparative study of machine learning models for predicting the state of reactive mixing, Journal of Computational Physics, 432, 110147, https://doi.org/10.1016/j.jcp.2021.110147, 2021.

Anderson, R. T., Vrionis, H. A., Ortiz-Bernad, I., Resch, C. T., Long, P. E., Dayvault, R., Karp, K., Marutzky, S., Metzler, D. R., Peacock, A., White, D. C., Lowe, M., and Lovley, D. R.: Stimulating the In Situ Activity of Geobacter Species To Remove Uranium from the Groundwater of a Uranium-Contaminated Aquifer, AEM, 69, 5884–5891, https://doi.org/10.1128/AEM.69.10.5884-5891.2003, 2003.

Arora, B., Dwivedi, D., Faybishenko, B., Wainwright, H. M., and Jana, R. B.: 10. Understanding and Predicting Vadose Zone Processes, in: 10. Understanding and Predicting Vadose Zone Processes, De Gruyter, 303–328, https://doi.org/10.1515/9781501512001-011, 2020.

Bain, J. G., Blowes, D. W., Robertson, W. D., and Frind, E. O.: Modelling of sulfide oxidation with reactive transport at a mine drainage site, Journal of Contaminant Hydrology, 41, 23–47, https://doi.org/10.1016/S0169-7722(99)00069-8, 2000.

Bargar, J. R., Williams, K. H., Campbell, K. M., Long, P. E., Stubbs, J. E., Suvorova, E. I., Lezama-Pacheco, J. S., Alessi, D. S., Stylo, M., Webb, S. M., Davis, J. A., Giammar, D. E., Blue, L. Y., and Bernier-Latmani, R.: Uranium redox transition pathways in acetate-amended sediments, Proceedings of the National Academy of Sciences, 110, 4506–4511, https://doi.org/10.1073/pnas.1219198110, 2013.

Bethke, C. M., Sanford, R. A., Kirk, M. F., Jin, Q., and Flynn, T. M.: The thermodynamic ladder in geomicrobiology, American Journal of Science, 311, 183–210, https://doi.org/10.2475/03.2011.01, 2011.

Beucler, T., Rasp, S., Pritchard, M., and Gentine, P.: Achieving Conservation of Energy in Neural Network Emulators for Climate Modeling, https://doi.org/10.48550/ARXIV.1906.06622, 2019.

Bianchi, M., Zheng, L., and Birkholzer, J. T.: Combining multiple lower-fidelity models for emulating complex model responses for CCS environmental risk assessment, International Journal of Greenhouse Gas Control, 46, 248–258, https://doi.org/10.1016/j.ijggc.2016.01.009, 2016.

Bone, S. E., Dynes, J. J., Cliff, J., and Bargar, J. R.: Uranium(IV) adsorption by natural organic matter in anoxic sediments, PNAS, 114, 711–716, https://doi.org/10.1073/pnas.1611918114, 2017.

van Breukelen, B. M., Griffioen, J., Röling, W. F. M., and van Verseveld, H. W.: Reactive transport modelling of biogeochemical processes and carbon isotope geochemistry inside a landfill leachate plume, Journal of Contaminant Hydrology, 70, 249–269, https://doi.org/10.1016/j.jconhyd.2003.09.003, 2004.

Cama, J., Soler, J. M., and Ayora, C.: 15. Acid Water–Rock–Cement Interaction and Multicomponent Reactive Transport Modeling, in: 15. Acid Water–Rock–Cement Interaction and Multicomponent Reactive Transport Modeling, De Gruyter, 459–498, https://doi.org/10.1515/9781501512001-016, 2020.

Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L., and Moyer, E. J.: Statistical Emulation of Climate Model Projections Based on Precomputed GCM Runs*, Journal of Climate, 27, 1829–1844, https://doi.org/10.1175/JCLI-D-13-00099.1, 2014.

Chen, T. and He, T.: xgboost: eXtreme Gradient Boosting, 4, 2015.

Claesen, M. and De Moor, B.: Hyperparameter Search in Machine Learning, https://doi.org/10.48550/ARXIV.1502.02127, 2015.

Doherty, J.: PEST model-independent parameter estimation user manual, Watermark Numerical Computing, Brisbane, Australia, 3338, 3349, 2004.

Dolgaleva, I. V., Gorichev, I. G., Izotov, A. D., and Stepanov, V. M.: Modeling of the Effect of pH on the Calcite Dissolution Kinetics, Theor Found Chem Eng, 39, 614–621, https://doi.org/10.1007/s11236-005-0125-1, 2005.

Druhan, J. L., Steefel, C. I., Williams, K. H., and DePaolo, D. J.: Calcium isotope fractionation in groundwater: Molecular scale processes influencing field scale behavior, Geochimica et Cosmochimica Acta, 119, 93–116, https://doi.org/10.1016/j.gca.2013.05.022, 2013.

690 Druhan, J. L., Steefel, C. I., Conrad, M. E., and DePaolo, D. J.: A large column analog experiment of
691 stable isotope variations during reactive transport: I. A comprehensive model of sulfur cycling and
692 δ34S fractionation, Geochimica et Cosmochimica Acta, 124, 366–393,
693 https://doi.org/10.1016/j.gca.2013.08.037, 2014.

694 Druhan, J. L., Winnick, M. J., and Thullner, M.: 8. Stable Isotope Fractionation by Transport and
695 Transformation, in: 8. Stable Isotope Fractionation by Transport and Transformation, De Gruyter,
696 239–264, https://doi.org/10.1515/9781501512001-009, 2020.

697 Dullies, F., Lutze, W., Gong, W., and Nuttall, H. E.: Biological reduction of uranium—From the
698 laboratory to the field, Science of The Total Environment, 408, 6260–6271,
699 https://doi.org/10.1016/j.scitotenv.2010.08.018, 2010.

700 Dwivedi, D., Steefel, I. C., Arora, B., and Bisht, G.: Impact of Intra-meander Hyporheic Flow on
701 Nitrogen Cycling, Procedia Earth and Planetary Science, 17, 404–407,
702 https://doi.org/10.1016/j.proeps.2016.12.102, 2017.

703 Dwivedi, D., Steefel, C. I., Arora, B., Newcomer, M., Moulton, J. D., Dafflon, B., Faybishenko, B.,
704 Fox, P., Nico, P., Spycher, N., Carroll, R., and Williams, K. H.: Geochemical Exports to River From
705 the Intrameander Hyporheic Zone Under Transient Hydrologic Conditions: East River Mountainous
706 Watershed, Colorado, Water Resources Research, 54, 8456–8477,
707 https://doi.org/10.1029/2018WR023377, 2018.

708 Finsterle, S., Commer, M., Edmiston, J. K., Jung, Y., Kowalsky, M. B., Pau, G. S. H., Wainwright,
709 H. M., and Zhang, Y.: iTOUGH2: A multiphysics simulation-optimization framework for analyzing
710 subsurface systems, Computers & Geosciences, 108, 8–20,
711 https://doi.org/10.1016/j.cageo.2016.09.005, 2017.

712 Frazier, P. I.: A Tutorial on Bayesian Optimization, arXiv:1807.02811 [cs, math, stat], 2018.

713 Gatel, L., Lauvernet, C., Carluer, N., Weill, S., Tournebize, J., and Paniconi, C.: Global evaluation
714 and sensitivity analysis of a physically based flow and reactive transport model on a laboratory
715 experiment, Environmental Modelling & Software, 113, 73–83,
716 https://doi.org/10.1016/j.envsoft.2018.12.006, 2019.

717 Gaus, I., Azaroual, M., and Czernichowski-Lauriol, I.: Reactive transport modelling of the impact of
718 CO2 injection on the clayey cap rock at Sleipner (North Sea), Chemical Geology, 217, 319–337,
719 https://doi.org/10.1016/j.chemgeo.2004.12.016, 2005.

720 Gharasoo, M., Elsner, M., Van Cappellen, P., and Thullner, M.: Pore-Scale Heterogeneities Improve
721 the Degradation of a Self-Inhibiting Substrate: Insights from Reactive Transport Modeling, Environ.
722 Sci. Technol., 56, 13008–13018, https://doi.org/10.1021/acs.est.2c01433, 2022.

723 Grzeszczuk, R., Terzopoulos, D., and Hinton, G.: NeuroAnimator: fast neural network emulation and
724 control of physics-based models, in: Proceedings of the 25th annual conference on Computer
725 graphics and interactive techniques - SIGGRAPH '98, the 25th annual conference, Not Known, 9–
726 20, https://doi.org/10.1145/280814.280816, 1998.

Gu, X., Rempe, D. M., Dietrich, W. E., West, A. J., Lin, T.-C., Jin, L., and Brantley, S. L.: Chemical reactions, porosity, and microfracturing in shale during weathering: The effect of erosion rate, Geochimica et Cosmochimica Acta, 269, 63–100, https://doi.org/10.1016/j.gca.2019.09.044, 2020.

Hubbard, S. S., Williams, K. H., Agarwal, D., Banfield, J., Beller, H., Bouskill, N., Brodie, E., Carroll, R., Dafflon, B., Dwivedi, D., Falco, N., Faybishenko, B., Maxwell, R., Nico, P., Steefel, C., Steltzer, H., Tokunaga, T., Tran, P. A., Wainwright, H., and Varadharajan, C.: The East River, Colorado, Watershed: A Mountainous Community Testbed for Improving Predictive Understanding of Multiscale Hydrological–Biogeochemical Dynamics, Vadose Zone Journal, 17, 180061, https://doi.org/10.2136/vzj2018.03.0061, 2018.

Hubbard, S. S., Agarwal, D., Arora, B., Banfield, J. F., Bouskill, N., Brodie, E., Carroll, R. W. H., Dwivedi, D., Gilbert, B., Maavara, T., Maxwell, R. M., Newcomer, M. E., Nico, P. S., Sorensen, P., Steefel, C. I., Steltzer, H., Tokunaga, T. K., Varadharajan, C., Wainwright, H. M., Wan, J., and Williams, K. H.: Key Controls on Water and Nitrogen Exports occurring across Lifezones, Compartments and Interfaces of the Mountainous East River Watershed, CO, 2019, H23B-01, 2019.

Jin, Q. and Bethke, C. M.: A New Rate Law Describing Microbial Respiration, Appl Environ Microbiol, 69, 2340–2348, https://doi.org/10.1128/AEM.69.4.2340-2348.2003, 2003.

Jin, Q. and Bethke, C. M.: Predicting the rate of microbial respiration in geochemical environments, Geochimica et Cosmochimica Acta, 69, 1133–1143, 2005.

Jin, Q. and Bethke, C. M.: The thermodynamics and kinetics of microbial metabolism, American Journal of Science, 307, 643–677, https://doi.org/10.2475/04.2007.01, 2007.

Jin, Q. and Kirk, M. F.: Thermodynamic and Kinetic Response of Microbial Reactions to High CO2, Frontiers in Microbiology, 7, 1696, https://doi.org/10.3389/fmicb.2016.01696, 2016.

Jin, Q. and Kirk, M. F.: pH as a Primary Control in Environmental Microbiology: 1. Thermodynamic Perspective, Frontiers in Environmental Science, 6, 21, https://doi.org/10.3389/fenvs.2018.00021, 2018.

Johnson, J. W., Nitao, J. J., and Knauss, K. G.: Reactive transport modelling of CO2 storage in saline aquifers to elucidate fundamental processes, trapping mechanisms and sequestration partitioning, Geological Society, London, Special Publications, 233, 107–128, https://doi.org/10.1144/GSL.SP.2004.233.01.08, 2004.

Jung, H. and Navarre-Sitchler, A.: Scale effect on the time dependence of mineral dissolution rates in physically heterogeneous porous media, Geochimica et Cosmochimica Acta, 234, 70–83, https://doi.org/10.1016/j.gca.2018.05.009, 2018.

Kashinath, K., Mustafa, M., Albert, A., Wu, J.-L., Jiang, C., Esmaeilzadeh, S., Azizzadenesheli, K., Wang, R., Chattopadhyay, A., Singh, A., Manepalli, A., Chirila, D., Yu, R., Walters, R., White, B., Xiao, H., Tchelepi, H. A., Marcus, P., Anandkumar, A., Hassanzadeh, P., and Prabhat: Physics-informed machine learning: case studies for weather and climate modelling, Phil. Trans. R. Soc. A., 379, 20200093, https://doi.org/10.1098/rsta.2020.0093, 2021.

764 Komlos, J., Peacock, A., Kukkadapu, R. K., and Jaffé, P. R.: Long-term dynamics of uranium
765 reduction/reoxidation under low sulfate conditions, Geochimica et Cosmochimica Acta, 72, 3603–
766 3615, https://doi.org/10.1016/j.gca.2008.05.040, 2008.

767 Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Chalikov, D. V.: New Approach to Calculation of
768 Atmospheric Model Physics: Accurate and Fast Neural Network Emulation of Longwave Radiation
769 in a Climate Model, Monthly Weather Review, 133, 1370–1383,
770 https://doi.org/10.1175/MWR2923.1, 2005.

771 Kyas, S., Volpatto, D., Saar, M. O., and Leal, A. M. M.: Accelerated reactive transport simulations in
772 heterogeneous porous media using Reaktoro and Firedrake, Comput Geosci, 26, 295–327,
773 https://doi.org/10.1007/s10596-021-10126-2, 2022.

774 Laloy, E. and Jacques, D.: Speeding up reactive transport simulations in cement systems by surrogate
775 geochemical modeling: deep neural networks and k-nearest neighbors,
776 http://arxiv.org/abs/2107.07598, 30 November 2021.

777 Li, L., Maher, K., Navarre-Sitchler, A., Druhan, J., Meile, C., Lawrence, C., Moore, J., Perdrial, J.,
778 Sullivan, P., Thompson, A., Jin, L., Bolton, E. W., Brantley, S. L., Dietrich, W. E., Mayer, K. U.,
779 Steefel, C. I., Valocchi, A., Zachara, J., Kocar, B., Mcintosh, J., Tutolo, B. M., Kumar, M.,
780 Sonnenthal, E., Bao, C., and Beisman, J.: Expanding the role of reactive transport models in critical
781 zone processes, Earth-Science Reviews, 165, 280–301,
782 https://doi.org/10.1016/j.earscirev.2016.09.001, 2017.

783 Li, Y., Lu, P., and Zhang, G.: An artificial-neural-network-based surrogate modeling workflow for
784 reactive transport modeling, Petroleum Research, 7, 13–20,
785 https://doi.org/10.1016/j.ptlrs.2021.06.002, 2022.

786 Long, P. E., Williams, K. H., Davis, J. A., Fox, P. M., Wilkins, M. J., Yabusaki, S. B., Fang, Y.,
787 Waichler, S. R., Berman, E. S. F., Gupta, M., Chandler, D. P., Murray, C., Peacock, A. D., Giloteaux,
788 L., Handley, K. M., Lovley, D. R., and Banfield, J. F.: Bicarbonate impact on U(VI) bioreduction in a
789 shallow alluvial aquifer, Geochimica et Cosmochimica Acta, 150, 106–124,
790 https://doi.org/10.1016/j.gca.2014.11.013, 2015.

791 Lu, H., Ermakova, D., Wainwright, H. M., Zheng, L., and Tartakovsky, D. M.: DATA-INFORMED
792 EMULATORS FOR MULTI-PHYSICS SIMULATIONS, J Mach Learn Model Comput, 2, 33–54,
793 https://doi.org/10.1615/JMachLearnModelComput.2021038577, 2021.

794 Maavara, T., Siirila-Woodburn, E. R., Maina, F., Maxwell, R. M., Sample, J. E., Chadwick, K. D.,
795 Carroll, R., Newcomer, M. E., Dong, W., Williams, K. H., Steefel, C. I., and Bouskill, N. J.:
796 Modeling geogenic and atmospheric nitrogen through the East River Watershed, Colorado Rocky
797 Mountains, PLOS ONE, 16, e0247907, https://doi.org/10.1371/journal.pone.0247907, 2021a.

798 Maavara, T., Siirila-Woodburn, E. R., Maina, F., Maxwell, R. M., Sample, J. E., Chadwick, K. D.,
799 Carroll, R., Newcomer, M. E., Dong, W., Williams, K. H., Steefel, C. I., and Bouskill, N. J.: Nitrate,
800 ammonium, and DON mass time series output for East River stream, vadose zone and groundwater
801 subwatersheds from HAN-SoMo model, Environmental System Science Data Infrastructure for a
802 Virtual Ecosystem (ESS-DIVE) (United States); Watershed Function SFA,
803 https://doi.org/10.15485/1766811, 2021b.

804 Malaguerra, F., Albrechtsen, H.-J., and Binning, P. J.: Assessment of the contamination of drinking
805 water supply wells by pesticides from surface water resources using a finite element reactive
806 transport model and global sensitivity analysis techniques, Journal of Hydrology, 476, 321–331,
807 https://doi.org/10.1016/j.jhydrol.2012.11.010, 2013.

808 Martinez, B. C., DeJong, J. T., and Ginn, T. R.: Bio-geochemical reactive transport modeling of
809 microbial induced calcite precipitation to predict the treatment of sand in one-dimensional flow,
810 Computers and Geotechnics, 58, 1–13, https://doi.org/10.1016/j.compgeo.2014.01.013, 2014.

811 Molins, S. and Knabner, P.: 2. Multiscale Approaches in Reactive Transport Modeling, in: 2.
812 Multiscale Approaches in Reactive Transport Modeling, De Gruyter, 27–48,
813 https://doi.org/10.1515/9781501512001-003, 2020.

814 Moon, H. S., McGuinness, L., Kukkadapu, R. K., Peacock, A. D., Komlos, J., Kerkhof, L. J., Long,
815 P. E., and Jaffé, P. R.: Microbial reduction of uranium under iron- and sulfate-reducing conditions:
816 Effect of amended goethite on microbial community composition and dynamics, Water Research, 44,
817 4015–4028, https://doi.org/10.1016/j.watres.2010.05.003, 2010.

818 Paper, J. M., Flynn, T. M., Boyanov, M. I., Kemner, K. M., Haller, B. R., Crank, K., Lower, A., Jin,
819 Q., and Kirk, M. F.: Influences of pH and substrate supply on the ratio of iron to sulfate reduction,
820 Geobiology, 19, 405–420, https://doi.org/10.1111/gbi.12444, 2021.

821 Richter, F. M. and DePaolo, D. J.: Numerical models for diagenesis and the Neogene Sr isotopic
822 evolution of seawater from DSDP Site 590B, Earth and Planetary Science Letters, 83, 27–38,
823 https://doi.org/10.1016/0012-821X(87)90048-3, 1987.

824 Rolle, M. and Borgne, T. L.: 5. Mixing and Reactive Fronts in the Subsurface, in: 5. Mixing and
825 Reactive Fronts in the Subsurface, De Gruyter, 111–142, https://doi.org/10.1515/9781501512001-
826 006, 2020.

827 Seigneur, N., Vriens, B., Beckie, R. D., and Mayer, K. U.: Reactive transport modelling to
828 investigate multi-scale waste rock weathering processes, Journal of Contaminant Hydrology, 236,
829 103752, https://doi.org/10.1016/j.jconhyd.2020.103752, 2021.

830 Steefel, C. I., DePaolo, D. J., and Lichtner, P. C.: Reactive transport modeling: An essential tool and
831 a new research approach for the Earth sciences, Earth and Planetary Science Letters, 240, 539–558,
832 https://doi.org/10.1016/j.epsl.2005.09.017, 2005a.

833 Steefel, C. I., DePaolo, D. J., and Lichtner, P. C.: Reactive transport modeling: An essential tool and
834 a new research approach for the Earth sciences, Earth and Planetary Science Letters, 240, 539–558,
835 https://doi.org/10.1016/j.epsl.2005.09.017, 2005b.

836 Steefel, C. I., Appelo, C. A. J., Arora, B., Jacques, D., Kalbacher, T., Kolditz, O., Lagneau, V.,
837 Lichtner, P. C., Mayer, K. U., Meeussen, J. C. L., Molins, S., Moulton, D., Shao, H., Šimůnek, J.,
838 Spycher, N., Yabusaki, S. B., and Yeh, G. T.: Reactive transport codes for subsurface environmental
839 simulation, Comput Geosci, 19, 445–478, https://doi.org/10.1007/s10596-014-9443-x, 2015.

840 Torres, E., Couture, R. M., Shafei, B., Nardi, A., Ayora, C., and Van Cappellen, P.: Reactive
841 transport modeling of early diagenesis in a reservoir lake affected by acid mine drainage: Trace

metals, lake overturn, benthic fluxes and remediation, Chemical Geology, 419, 75–91, https://doi.org/10.1016/j.chemgeo.2015.10.023, 2015.

Williams, K. H., Long, P. E., Davis, J. A., Wilkins, M. J., N'Guessan, A. L., Steefel, C. I., Yang, L., Newcomer, D., Spane, F. A., Kerkhof, L. J., McGuinness, L., Dayvault, R., and Lovley, D. R.: Acetate Availability and its Influence on Sustainable Bioremediation of Uranium-Contaminated Groundwater, Geomicrobiology Journal, 28, 519–539, https://doi.org/10.1080/01490451.2010.520074, 2011.

Wu, W.-M., Carley, J., Gentry, T., Ginder-Vogel, M. A., Fienen, M., Mehlhorn, T., Yan, H., Caroll, S., Pace, M. N., Nyman, J., Luo, J., Gentile, M. E., Fields, M. W., Hickey, R. F., Gu, B., Watson, D., Cirpka, O. A., Zhou, J., Fendorf, S., Kitanidis, P. K., Jardine, P. M., and Criddle, C. S.: Pilot-Scale in Situ Bioremedation of Uranium in a Highly Contaminated Aquifer. 2. Reduction of U(VI) and Geochemical Control of U(VI) Bioavailability, Environ. Sci. Technol., 40, 3986–3995, https://doi.org/10.1021/es051960u, 2006.

Yoo, A. B., Jette, M. A., and Grondona, M.: SLURM: Simple Linux Utility for Resource Management, in: Job Scheduling Strategies for Parallel Processing, vol. 2862, edited by: Feitelson, D., Rudolph, L., and Schwiegelshohn, U., Springer Berlin Heidelberg, Berlin, Heidelberg, 44–60, https://doi.org/10.1007/10968987_3, 2003.