Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 97 (2016) 447 – 453

Energy Procedia

European Geosciences Union General Assembly 2016, EGU
Division Energy, Resources & Environment, ERE

Data-driven surrogate model approach for improving the performance of reactive transport simulations

Janis Jatnieks[a,c,*], Marco De Lucia[b], Doris Dransch[a,c], Mike Sips[a,d]

[a] *GFZ German Research Centre for Geosciences, Section 1.5 Geoinformatics, Potsdam, Germany*
[b] *GFZ German Research Centre for Geosciences, Section 3.4 Fluid Systems Modeling, Potsdam, Germany*
[c] *Humboldt University of Berlin, Department of Geography, Berlin, Germany*
[d] *Humboldt University of Berlin, Department of Computer Science, Berlin, Germany*

Abstract

Geochemical simulation models are the computational bottleneck for coupled reactive transport simulations. We investigate the use of a data-driven surrogate model in place of a geochemical simulation model to speed up the run-times of reactive transport simulations. This is a challenge because the surrogate model needs to use results of its predictions as inputs at each subsequent simulation step. We test the suitability of surrogate model approach on a popular reactive transport benchmark problem, often used for evaluating simulation models. We show that the concept is feasible and can make the simulations many times faster, however several open areas for future work remain.

Peer-review under responsibility of the organizing committee of the General Assembly of the European Geosciences Union (EGU)

Keywords: Data-driven surrogate; reactive transport; coupled geochemical simulation; surrogate model

1. Introduction

Reactive transport models deal with simulation of geochemical reactions together with transport of fluids in the geological subsurface. These models have a wide range of important applications, such as the assessment of long-term outcome of underground gas storage [1], geothermal energy, and many others.

Reactive transport simulations are often implemented as a coupling of two distinct simulation models. One is responsible for the hydrodynamical processes - fluid flow and solute transport. Another is a geochemical simulation model responsible for geochemical reactions. Typically, geochemical simulations are the main computational bottleneck in coupled reactive transport simulations. This limitation allows only reactive transport simulations of coarse spatial resolution to be performed. In comparison, hydrodynamic simulations can have detailed geometries ranging into millions of elements that can be solved on single workstations [3]. Coupled reactive transport simulations of such detail are not currently feasible.

* Corresponding author. Tel.: +49-157-3245-1188.
E-mail address: jatnieks@janis.es

Peer-review under responsibility of the organizing committee of the General Assembly of the European Geosciences Union (EGU)
doi:10.1016/j.egypro.2016.10.047

The main reasons for the computational cost associated with geochemical simulations are the large number of parameter involved in each run and the large number of runs required at each simulation time step. Depending on the scenario and user interest, the number of input and output variables can range into hundreds [4]. In the framework of sequential coupling between hydrodynamics and chemistry, one geochemical simulation model is executed for every grid element at each time step of the coupled simulation. Executing the geochemical simulation that has such a large number of variables many times results in long simulation run times. For these reasons, the reduction of run times for geochemical simulations is of great interest for the scientific community.

2. Approach and challenge

One way to reduce the high computational cost of geochemical simulation is to replace the geochemical simulation model with a surrogate model. A surrogate model is a fast-running approximation that can be used as a replacement for the geochemical simulation model [5].

In this study, we focus on a data-driven surrogate approach. It is important to note that a data-driven approach is different from model reduction in mathematical modelling. Model reduction aims at preserving the underlying physical principles behind the simulation model while creating a simplified version of it. Instead, the data-driven approach considers input-output data produced from a coarse sampling of the parameter space of the simulation model, but without any physical considerations. As long as the surrogate model can predict the simulator output based on the input values, it may be possible to use the data-driven surrogate in place of the geochemical simulation model.

There are several reasons leading us to believe that only a necessary subset of geochemical models capabilities must be contained in a data-driven surrogate. First, geochemical simulations are often performed repeatedly for studying a specific site. This is computationally costly when using a geochemical simulation model. However, this site is associated with particular ranges of input and output data. When a reactive transport simulation for the site is performed with reduced spatial resolution, this provides input and output data samples from the geochemical simulation model. Second, the capabilities required from the surrogate model can be limited to a specific foreseen set of application scenarios. Finally, there is a margin for errors that may be acceptable if a perfect surrogate model is not obtainable. Geochemical models are affected by uncertainties concerning input parameters and the definition of a representative initial state for a particular site and scenario. Such uncertainties are inherent to the current state-of-the-art in geochemical simulations [6]. Because of this, the surrogate model may not always need to perfectly reproduce the simulator output. However, a surrogate model does have to be significantly faster than the "full physics" geochemistry simulator it is replacing.

To evaluate our approach in a reactive transport simulation scenario, we used a popular benchmark problem which has been defined to evaluate and compare different geochemical simulation models. Therefore, we believe it is also an appropriate test for a surrogate model.

There is a particular challenge associated with the use of a surrogate model instead of a "full physics" geochemical simulator for reactive transport. The surrogate model has to take its own prediction results, that are modified by the transport model, and use them as inputs at the next time step as shown in Figure 1. Most statistical models make some prediction errors. This can mean that with each subsequent time-step the changes in input parametrization can take the surrogate model outside the parameter space region for which it was trained. The large number of studies relying on surrogate models for various tasks in water resources research [5,7] do not use surrogate models in this recursive way. Our aim was to answer the question if it is feasible to use a surrogate model as a replacement for a coupled geochemical simulation model.

3. Related work

Surrogate modeling is a well established engineering approach for dealing with long-running simulation problems [5,8,9]. This can also be seen in the number of terms that are often used as synonymous for this approach. Among these are proxy models [10], emulators [11], meta-models [12], reduced order models [13], lower fidelity models [14] and response surface models [15]. Significant work has been invested in using surrogate models for speeding up computationally expensive simulations for water resources applications in general. The most comprehensive recent