Speeding up reactive transport simulations: statistical surrogates and caching of simulation results in lookup tables

Marco De Lucia[1], Robert Engelmann[2], Michael Kühn[1,3], Alexander Lindemann[2], Max Lübke[2], and Bettina Schnor[2]

[1]GFZ German Research Centre for Geosciences, Fluid Systems Modelling, Potsdam, Germany (delucia@gfz-potsdam.de)
[2]University of Potsdam, Institute of Computer Science, Operating Systems and Distributed Systems, Potsdam, Germany
[3]University of Potsdam, Institute of Geosciences, Hydrogeology, Potsdam, Germany

A successful strategy for speeding up coupled reactive transport simulations at price of acceptable accuracy loss is to compute geochemistry, which represents the bottleneck of these simulations, through data-driven surrogates instead of 'full physics' equation-based models [1]. A surrogate is a multivariate regressor trained on a set of pre-calculated geochemical simulations or potentially even at runtime during the coupled simulations. Many available algorithms and implementations are available from the thriving Machine Learning community: tree-based regressors such as Random Forests or xgboost, Artificial Neural Networks, Gaussian Processes and Support Vector Machines just to name a few. Given the 'black-box' nature of the surrogates, however, they generally disregard physical constraints such as mass and charge balance, which are of course of paramount importance for coupled transport simulations. A runtime check of error of balances in the surrogate outcomes is therefore necessary: predictions offending a given tolerance must be rejected and the full physics chemical simulations run instead. Thus the practical speedup of this strategy is a tradeoff between careful training of the surrogate and run-time efficiency.

In this contribution we demonstrate that the use of surrogates can lead to a dramatic decrease of required computing time, with speedup factors in the order of 10 or even 100 in the most favorable cases. Thus, large scale simulations with some 10^6 grid elements are feasible on common workstations without requiring computation on HPC clusters [2].

Furthermore, we showcase our implementation of Distributed Hash Tables caching geochemical simulation results for further reuse in subsequent time steps. The computational advantage here stems from the fact that query and retrieval from lookup tables is much faster than both full physics geochemical simulations and surrogate predictions. Another advantage of this algorithm is that virtually no loss of accuracy is introduced in the simulations. Enabling the caching of geochemical simulations through DHT speeds up large scale reactive transport simulations up to a factor of four even when computing on several hundred cores.

These algorithmical developments are demonstrated in comparison with published reactive transport benchmarks and on a real-life scenario of CO_2 storage.

[1] Jatnieks, J., De Lucia, M., Dransch, D., Sips, M. (2016): Data-driven surrogate model approach for improving the performance of reactive transport simulations. Energy Procedia 97, pp. 447-453. DOI: 10.1016/j.egypro.2016.10.047

[2] De Lucia, M., Kempka, T., Jatnieks, J., Kühn, M. (2017): Integrating surrogate models into subsurface simulation framework allows computation of complex reactive transport scenarios. Energy Procedia 125, pp. 580-587. DOI: 10.1016/j.egypro.2017.08.200