

Surrogates and caching of results in lookup tables: tools to speedup reactive transport simulations

M. DE LUCIA¹, M. KÜHN^{1,2}, M. LÜBKE³, B. SCHNOR³

¹GFZ German Research Centre for Geosciences, Fluid Systems Modelling, Potsdam, Germany, delucia@gfz-potsdam.de

²University of Potsdam, Institute of Geosciences, Hydrogeology, Potsdam, Germany

³University of Potsdam, Institute of Computer Science, Operating Systems and Distributed Systems, Potsdam, Germany

The computation of geochemistry normally represents the bottleneck in coupled reactive transport simulations. Since geochemistry is affected by large uncertainties, a natural idea to achieve speedup at price of acceptable accuracy loss is to approximate the geochemical subsystem with data-driven statistical surrogates instead of classical equation-based models. A surrogate is a multivariate regressor, often treated as “black box”, trained on a set of pre-calculated geochemical simulations. Many regression algorithms are available such as gradient-boosting tree-based regressors (xgboost, carboost), Neural Networks and Gaussian Processes. Their performance depend on the problem at hand. In our implementation, surrogate predictions offending a given tolerance on mass and charge balance are rejected and the classical chemical simulations run instead. Thus the practical speedup of this strategy is a tradeoff between careful training of the surrogate and run-time efficiency. We demonstrate how using surrogates leads to dramatic decrease of computing time, with speedup factors up to 100 in favorable cases.

Furthermore, we showcase our implementation of Distributed Hash Tables (DHT) for caching geochemical simulation results and further reuse in subsequent time steps in the coupled simulations. Query and retrieval from DHT is much faster than both classical geochemical simulations and surrogate predictions, and introduces negligible to no loss of accuracy in the simulations. Using DHT speeds up large scale reactive transport simulations up to a factor of four even when computing on several hundred CPUs.

These developments are demonstrated on already published reactive transport benchmarks and on a real-life scenario inspired to subsurface CO₂ storage at the Ketzin pilot site in Germany.

This work was partly funded by HGF - Initiative and Networking Fund (ZT-I-0010) in project *Reduced Complexity Models*.