Goldschmidt 2021 Abstract
https://doi.org/10.7185/gold2021.5922

Acceleration of reactive transport simulations with data-driven and hybrid physics-artificial intelligence surrogates

MARCO DE LUCIA[1], MORGAN TRANTER[1,2] AND MICHAEL KÜHN[1,2]

[1]GFZ German Research Centre for Geosciences

[2]University of Potsdam

Presenting Author: delucia@gfz-potsdam.de

The heavy computational loads of coupled reactive transport simulations, mostly due to the chemical subsystem, represents a major barrier for their wide application. It is however possible to achieve decisive speedups of the coupled simulations by replacing the equation-based "full physics" geochemical sub-process simulation with a pre-trained statistical surrogate at the price of small accuracy loss, enabled by efficient machine learning and artificial intelligence (ML/AI) algorithms. In this contribution we compare two distinct surrogate-development approaches: the first is purely data-driven, meaning that it is implemented in a completely process-agnostic way, and it treats the geochemical sub-process as a "black box" in which only input and output data are considered. This generically applicable strategy relies on large, refined training datasets and is error-prone when encountering previously unseen data. Among the many regression algorithms available, we are most successful with gradient-boosting trees such as xgboost. The second, novel approach builds the surrogate based on geochemical knowledge, which is leveraged by feature engineering shaped according to mass action laws and stoichiometric reactions, and reduces the learning process to a search for bijective monovariate relationships on recursive partitions of the training data. This "hybrid physics-AI" approach, which ultimately results in a decision-tree, is more robust towards previously unseen data, and is explicable by design, since it contributes to a better understanding of the underlying geochemical process. However, its construction is currently not thoroughly automated and relies to some extent on manual calibration. The two methods are demonstrated on fully kinetic benchmarks derived from literature and from recent experimental and modelling works relevant for subsurface utilization, notably for energy storage, geothermal energy and spent nuclear fuel repository. With both approaches, which should be considered complementary and not alternative, we are able to achieve speedup factors of around 5 for the 1D benchmarks used for training and, by extrapolation, of 50 to 100 on grids with 10^5- 10^6 elements.