Geophysical Research Abstracts
Vol. 19, EGU2017-10924, 2017
EGU General Assembly 2017

Semi-supervised Machine Learning for Analysis of Hydrogeochemical Data and Models

Velimir Vesselinov, Daniel O'Malley, Boian Alexandrov, and Bryan Moore
Los Alamos National Laboratory, United States (vvv@lanl.gov)

Data- and model-based analyses such as uncertainty quantification, sensitivity analysis, and decision support using complex physics models with numerous model parameters and typically require a huge number of model evaluations (on order of 10^6). Furthermore, model simulations of complex physics may require substantial computational time. For example, accounting for simultaneously occurring physical processes such as fluid flow and biogeochemical reactions in heterogeneous porous medium may require several hours of wall-clock computational time. To address these issues, we have developed a novel methodology for semi-supervised machine learning based on Non-negative Matrix Factorization (NMF) coupled with customized k-means clustering. The algorithm allows for automated, robust Blind Source Separation (BSS) of groundwater types (contamination sources) based on model-free analyses of observed hydrogeochemical data. We have also developed reduced order modeling tools, which coupling support vector regression (SVR), genetic algorithms (GA) and artificial and convolutional neural network (ANN/CNN). SVR is applied to predict the model behavior within prior uncertainty ranges associated with the model parameters. ANN and CNN procedures are applied to upscale heterogeneity of the porous medium. In the upscaling process, fine-scale high-resolution models of heterogeneity are applied to inform coarse-resolution models which have improved computational efficiency while capturing the impact of fine-scale effects at the course scale of interest. These techniques are tested independently on a series of synthetic problems. We also present a decision analysis related to contaminant remediation where the developed reduced order models are applied to reproduce groundwater flow and contaminant transport in a synthetic heterogeneous aquifer. The tools are coded in Julia and are a part of the MADS high-performance computational framework (https://github.com/madsjulia/Mads.jl; http://madsjulia.github.io/Mads.jl).