

Uncertainty quantification for groundwater management in the Danish buried valley systems by means of regression tree-based surrogate models

*Jihoon Park¹, Céline Scheidt¹, Jef Caers²

1. Department of Energy Resources Engineering, Stanford University, 2. Department of Geological Sciences, Stanford University

Uncertainty quantification is a key component for decision making in groundwater management. Such applications involve the building of large complex spatial models, the application of computationally intensive forward modeling codes and the integration of heterogeneous sources of uncertainty. An integral step for uncertainty quantification is to condition models to a variety of data. In the Danish groundwater management this consists of head, streamflow, recharge, well and geophysical (SkyTEM) data. Uncertainty quantification requires model calibration. This is a challenging problem when dealing with complex systems (such as the Danish buried valley system) and a wealth of data. Another difficulty is computational cost, since a proper model calibration should account for all data, all model variables and geological heterogeneity requires running many forward flow models.

In this research, a workflow is proposed to find posterior multivariate distribution of model parameters and predictions. First, dimensionality reduction with mixed principle component analysis (PCA) is performed to incorporate different types of available data. A regression model is built for uncertain model parameters and misfit between simulated and observed data. As a regression model, we use a boosted regression tree because it offers high quality predictive model in nonlinear problems. Another advantage of tree-based approach is that we can obtain predictor importance, which can be directly used in sensitivity analysis.

Models that match the data are found by Approximate Bayesian Computations (ABC), where the likelihood is simply an indicator function of data mismatch. ABC requires exhaustive Monte Carlo sampling and running forward models. By using the regression model as surrogate forward model, we can obtain models conditioned to the data without intensive full forward runs. Regression models can also be constructed for predictions, such as the effect of establishing new wells for extraction.

We illustrate our method using a real field problem of decision making in the Danish groundwater system. Decisions include where to relocate drinking wells while minimizing the change of water produced and effects on farms and industrial areas. Well head and stream data are observed from monitoring wells. The proposed workflow is used to understand the effect of each parameter and to obtain the posterior distribution of 20 forecasts with newly acquired data.

Keywords: uncertainty quantification, regression analysis, model calibration, groundwater management