Abstract

Effective water resources management typically relies on numerical models to analyze groundwater flow and solute transport processes. Since the important hydrogeological parameters of these models (e.g., hydraulic conductivity) cannot be measured, they are normally estimated by model calibration. In addition, groundwater models are often subject to input data errors, as some of the input forcings (such as recharge and well pumping rates) are unknown or estimated. In addition, model structural error is ubiquitous, due to simplification and/or misrepresentation of the real system. The presence of errors in input data and model structure raises questions regarding the suitability of conventional least squares regression-based (LSR) calibration.

We present a Bayesian framework that explicitly recognizes errors in input forcings and model structure and is tailored for groundwater models. The framework implements a marginalizing step to account for input data variability when evaluating the likelihood, and explicitly describes the model structural error statistically in an inductive, data-driven way. We adopt a fully Bayesian approach that integrates Gaussian process error models into the calibration, prediction and uncertainty analysis.

We test the usefulness of the fully Bayesian approach with synthetic case studies of the impact of pumping on surface-ground water interaction and a real-world case study. In the real-world case study, Bayesian inference is facilitated using high performance computing and fast surrogate models (based on machine learning technique) as a substitute for the computationally expensive groundwater model. We demonstrate in the case studies that when input error or model structural error is present but not explicitly taken into account, the parameters can be overly adjusted to compensate for input data and model structural error, thus leading to biased and overconfident parameter estimates. Parameter compensation is found to have deleterious impact when the model is used to make prediction under new scenarios. In contrast, the presented Bayesian framework effectively alleviates parameter compensation and gives predictions that are more consistent with validation data in all case studies. The results highlight the importance of explicit treatment of errors in input forcings and model structure especially in circumstances where subsequent decision-making and risk analysis require accurate prediction and uncertainty quantification. Follow-up studies will further investigate the feasibility of joint inference of input and model structural errors, particularly for real-world modeling practice.