

Impact of Calibration Data on Evaluating Plausibility of Alternative Groundwater Models

Hydrologic environments are open and complex, rendering them prone to multiple interpretations and mathematical descriptions regardless of the quantity and quality of available data. This recognition has led to a growing tendency among hydrologists to postulate several alternative hydrologic models for a site. Models here are not limited to governing equations and associated boundary/initial conditions, but refer to conceptual-mathematical representations of hydrologic systems (e.g., their processes and interactions). Facing the alternative models, the scientific question to be answered is how to evaluate plausibility of the models so that the models can be properly used to yield optimum predictions. Evaluating model plausibility considers the entire modeling process (including model formulation, calibration, and validation), and calibration/validation data play a key role in the evaluation process. Although calibrating a single model has been studied for decades, impact of calibration data on evaluating plausibility of multiple models has not been well understood. Open questions are as follows: What kinds of calibration data can be used to most effectively discriminate between models? How many data are needed to reliably evaluate model plausibility? How does data correlation (spatial and temporal) affect the evaluation? How does biased evaluation of model plausibility influence predictive performance? These fundamental questions will be addressed using an interdisciplinary approach combining Bayesian statistical and computational methods. Model plausibility will be quantified using model probability, which is estimated, in a Bayesian framework, based on conformity of model simulations to calibration data, complexity of models, and expert judgment. Based on the model probability, one can choose a single model (i.e., Bayesian model selection) or use multiple models (i.e., Bayesian model averaging) to make predictions. Predictive performance of the Bayesian model selection or averaging will be investigated. Hypothesis will be tested using a two-pronged strategy based on both synthetic and real-world modeling. For the synthetic case, alternative groundwater models will be developed based on different representations of site heterogeneity and boundary conditions. The real-world modeling will be conducted at the Naturita site, Colorado, where a risk exists that uranium may reach the Colorado River. Alternative models will be developed based on different ways of formulations of uranium reactive transport models, such as surface complexation models with different numbers of functional groups. For the real-world modeling, model predictive performance will be evaluated using cross-validation methods such as leave-one-out and K-fold. The synthetic and real-world modeling will be conducted together with USGS scientists at Boulder and Menlo Park. Scientific insights gained in this project will be valuable to any environmental modeling through cost-effective data collection for refining existing models and developing new models for environmental restoration and protection.