Improving Prediction of Subsurface Flow and Transport through Exploratory Data Analysis and Complementary Modeling

Principal Investigators
Albert J. Valocchi, Department of Civil and Environmental Engineering
Eyal Amir, Department of Computer Science
University of Illinois at Urbana-Champaign

Computer models of groundwater flow and solute transport are important tools for scientists and decision makers to manage scarce groundwater resources and assess the risk of groundwater contamination. These models are used to support decisions that have enormous social, political, and economic impacts upon stakeholders. For example, in semi-arid regions such as the Republican River Basin, a farmer's irrigation pumping can be restricted based upon the numerically simulated impacts of groundwater withdrawal upon surface water. Unfortunately, these models are fraught with many different types of uncertainties. Current state-of-the-art practice is to assume that model parameters are uncertain and to use statistical regression to estimate these parameters from historical data; this calibrated model is then used for prediction and extrapolation. However, the unavoidable simplification and uncertainty associated with physically-based hydrological models often results in both random and systematic model calibration errors, thus violating key foundational assumptions for regression-based parameter estimation and uncertainty analysis. Our research project develops a new complementary (or hybrid) framework to handle the systematic error by integrating physically-based groundwater flow models with error-mapping data-driven models, thereby improving the predictive ability of the calibrated models. We hypothesize that model error, i.e., the mismatch between observed and modeled values, should reflect the main sources of modeling uncertainty, and that methods from the computer science field of machine learning can be used to discover underlying patterns and relationships in this error. Specific tasks include: (1) evaluating different types of machine learning and model integration techniques, (2) evaluating the ability of machine learning to identify the underlying sources of the systematic error thereby leading to an improved physically-based model, and (3) developing comprehensive uncertainty analysis techniques. Our complementary modeling framework will be tested and validated using real-world data and applications. We will also partner with campus programs and work with Illinois high school science and mathematics teachers to develop curriculum modules that incorporate concepts of mathematical simulation of natural systems, statistical fitting of models to observed data, and model uncertainty.

The research will lead to development of a new complementary modeling framework that integrates widely-used groundwater simulation models with field observations. Due to advances in environmental sensors and information technology, groundwater and surface water data are becoming increasingly available, but the conventional approaches for groundwater model calibration were developed in a data-poor era. Our framework will lead to more accurate local forecasts of groundwater response to natural and human disturbances, thereby improving the decision support for water resources management. Although we work in the area of groundwater resources, our research findings will also have broader implications for many other fields where calibrated physically-based computer models are used to forecast the behavior of complex natural systems.