



Placing stochastic simulation in a system-based context that promotes transparency and refutability

M.C. Hill (1), D. Kavetski (2), M.P. Clark (3), B.T. Nolan (4), M. Arabi (5), L. Foglia (6), S. Mehl (7), and M. Ye (8)

(1) U.S. Geological Survey, Boulder, Colorado, USA (mchill@usgs.gov), (2) School of Civil, Environmental and Mining Engineering, University of Adelaide, SA, Australia, (3) National Center for Atmospheric Research, Boulder, Colorado, USA, (4) U.S. Geological Survey, Reston, Virginia, USA, (5) Civil Engineering, Colorado State University, Fort Collins, Colorado, USA, (6) Department of Air, Land, and Water, University of California, Davis, California, USA, (7) Civil Engineering, California State University, Chico, California, USA, (8) Florida State University, Tallahassee, Florida, USA

Stochastic simulation often is used to evaluate the consequences of small-scale variation of selected system properties. In models of flow and transport models in the subsurface, often stochastic simulations focus on heterogeneity of the hydraulic conductivity field. This spatial variation exists within the context of larger-scale hydraulic conductivity variations and of other properties and boundary conditions that often, though perhaps erroneously, are represented at a larger scale. Understanding the small-scale stochastic variation in the context of the larger-scale properties becomes difficult when calibration, sensitivity analysis, and(or) uncertainty evaluation of the larger-scale properties require 1,000s to 100,000s of model runs. For example, multiobjective optimization, FAST, Markov-Chain Monte Carlo, and cross-validation all require many model runs. While all of these methods can be very useful, the high computational cost limits their applicability. An alternative is to consider computationally frugal local methods that often use 10s of 100s of highly parallelizable model runs, but these methods are often criticized because of their underlying assumptions related to weighting and linearity. The ability to obtain insight with so few model runs and the resulting opportunity to better understand the context within which detail is explored using stochastic methods is tempting, but only if the computationally frugal methods provide enough valuable insights. In this talk the problematic underlying assumptions are considered in the context of ideas about accounting for data error (including epistemic error) using error-based weighting and ideas about addressing model nonlinearity using robust models. Transparency is increased because measures of what is important to various objectives are available even for process models with lengthy execution times. Indeed, the ability to consider such models allows exploration of processes that would otherwise be impractical. Refutability is increased through better understanding of how data and their errors are propagated through parameter estimation and into predictions. Both transparency and refutability are thought by many to be required for defensible models of any environmental system.

Here we propose a set of computationally frugal methods to evaluate model fit to observations, sensitivity, data needs, predictions, and uncertainty. Examples highlight methods that focus on model fit and sensitivity analysis. This includes the use of error-based weighting and the maximum likelihood variance to detect model overfitting and underfitting. It also includes how composite scaled sensitivities and parameter correlation coefficients can be combined with parameter identifiability statistics to evaluate models for which parameters are estimated using SVD parameter transformation. Results suggest that in many groundwater problems, including saturated and unsaturated systems, other processes and properties such as recharge, bulk density, water content at field capacity, and so on are as important as subsurface hydraulic conductivity to calibration and prediction. This supports recent efforts to consider variability of multiple properties in stochastic evaluations. The ability to provide insight about large-scale processes quickly has exciting consequences for stochastic simulation. For example, it allows greater opportunity to screen more alternative large-scale and stochastic models in a meaningful way.