Comments on Model Uncertainty

by

Charles M. Brendecke, PhD, PE

For consideration by the Eastern Snake Hydrologic Modeling Committee
January 13-14, 2009

At the ESHMC meeting in August of last year we began discussion on the issue of ESPA model uncertainty. Allen Wylie presented two potential approaches to quantifying model uncertainty, a "multiple models" approach and a "bend-but-don't-break" approach, and asked for comments so as to further the discussion of this matter. This memo conveys my thoughts on the matter of model uncertainty and on Allen's suggested approaches.

Sources of Model Uncertainty

While there are semantic differences among experts and practitioners, I find it helpful to consider four fundamental types of model uncertainty.

- Conceptual uncertainty, aka "structural" or "geological" uncertainty...this arises because we can't know and fully represent the actual hydrogeologic structure, stratigraphy and boundaries of the aquifer.

- Parameter uncertainty...this arises because, for whatever model structure we adopt, we can't precisely quantify all the aquifer water budget terms we need as inputs.

- Calibration uncertainty (internal)...this arises because, for any given model structure and water budget, there are many combinations of calibration parameter (e.g., transmissivity, conductance) values that can lead to a similarly well-calibrated model.

- Calibration uncertainty (external)...this arises because we are calibrating the model to a set of targets (water levels and reach gains) that are themselves uncertain.

Overall model uncertainty is a function of all of these and model uncertainty may be different for different scenarios, depending on how different the water budget terms (e.g., aquifer stresses) are between scenarios. Conceptual uncertainty is the most difficult to quantify in a systematic way, partly because it does not lend itself to repetitive, incrementally changing calculations that can be programmed and executed in batch processes.

I think the Committee should discuss and be clear about which of these sources of uncertainty we are attempting to address in our overall quantification of model uncertainty.

Quantifying Model Uncertainty

Our goal should be to express model predictions in probabilistic terms, ideally as probability distributions. The probability distribution of a model prediction is a joint distribution, transformed through the model algorithm, of the probability distributions of all model inputs. This joint distribution is, in turn, one input distribution to the joint distribution derived from all possible models.

The most rigorous approach to generating these distributions in practice is Monte Carlo ("MC") simulation. This approach involves making large numbers of model runs (or "realizations") based on input values drawn randomly from their respective probability distributions (or their joint distribution). The results of this large number of runs are then displayed as the probability distribution of model predictions. Care must be taken in the set up of the random sampling process so as to account for statistical independence (or lack thereof) among the inputs. Unconditional MC simulation allows parameters to vary regardless of their effect on calibration, while conditional MC simulation filters input parameter vectors to keep simulation results within a pre-defined uncertainty range. Needless to say, this is a computationally intensive and time-consuming process.

A less intensive approach is a parametric sensitivity analysis, in which incremental changes are made in model parameters (one at a time) to cover their acknowledged uncertainty range. Effects on model results and calibration "fit" are assessed with each adjustment. While it is more tractable computationally, this approach does not lend itself as well to a probabilistic definition of uncertainty.

In both unconditional MC simulation and sensitivity analysis, there is the risk that the variation in input parameters will lead the model out of calibration. It is my understanding that PEST2000 offers a predictive analysis capability that supports calculation of predictive uncertainty associated with input parameter uncertainties while ensuring that the model remains in calibration. This would appear to avoid the risk in sensitivity analysis and in unconditioned MC simulation that the model is operating outside the range of calibration.

Allen's proposed approaches appear to come closest to the parameter sensitivity method. The "multiple models" approach could accommodate different conceptual models, though it is not clear we are considered these, while the "bend-but-don't-break" approach cannot. The latter would seem to require that the Committee agree *a priori* on a single conceptual model that would be subjected to "bending." Both of Allen's approaches could address parameter and calibration uncertainty. In all cases we will need a definition of the term "calibrated."

Some Suggestions for Ground Work and Discussion

To advance the ball on model uncertainty, I think the following foundational efforts should be considered. Many of these bear on the issue of conceptual uncertainty.

Definition of Calibration. In the past we simply looked at various calibration results and picked the model we thought "looked best." This time around we need to find a more objective approach, since we may be trying to compare calibrations from conceptually different models. We might, for example, prepare a map showing the kriged standard error of observed water levels across the ESPA. Any calibration run that produced heads within this standard error would be considered "calibrated." It is a bit harder to see how this would be applied to reach gains; perhaps reach gain error could be derived from the errors in the various components of the relatively simple water balance equation used for gains calculation.

Layering. In the past we adopted the concept of a single layer model. Over the course of the intervening years, arguments have been advanced that there may be multiple operative layers in some portions of the ESPA. Recent MODFLOW versions (HFB package) now accommodate representation of multiple layers in portions of the model domain. I would suggest a review of well logs and development of fence diagrams in certain areas of the aquifer (where such data is reasonably available), as well as water level maps stratified by well completion depth, to review whether there is reason to adopt multiple layers in portions of the aquifer.

Local Distribution of T and S. In the model, transmissivity and storativity are assumed uniform throughout each model cell. In some areas of the aquifer there is probably enough pump test data to examine how much variability exists in observed T and S values within a square mile cell. Understanding this variability would help inform the certainty that should be placed on model simulations that ask ever more detailed (spatially) questions, and might suggest where a finer grid should be considered.

Confined v. Unconfined. Most model simulations have been run in superposition mode. It is my understanding that this approach was adopted at least partly for convenience. There are some areas of the aquifer (e.g., the edges) where this assumption may lead to erroneous conclusions. Since computational power is now much more readily available, we should reconsider whether we ought to be making unconfined model runs and generating differences rather than doing everything with single runs.

Anisotropy. The model in its present form assumes isotropic conditions throughout the aquifer. Although it could readily be done, it does not appear that the MODFLOW isotropy parameter was subjected to PEST manipulation in the original calibration. Since then, arguments have been put forward that the aquifer may be anisotropic in some areas, and MODFLOW now allows isotropy to be set on a cell by cell basis (LPF package). I would suggest a more thorough evaluation of aquifer anisotropy, starting with a statistical analysis of observed T and S values in various parts of the aquifer to see if they contain any directional distribution.