2015

Metamodels to Bridge the Gap Between Modeling and Decision Support

Michael N. Fienen
U.S. Geological Survey, Wisconsin Water Science Center, Middleton, WI, mnfienen@usgs.gov

Bernard T. Nolan
USGS National Water Quality Assessment Program, Reston, VA, btnolan@usgs.gov

Daniel T. Feinstein
USGS WIWSC, Milwaukee, WI, dtfeinst@usgs.gov

J. Jeffrey Starn
USGS NAWQA Program, East Hartford, CT, jjstarn@usgs.gov

Groundwater

Metamodels to Bridge the Gap Between Modeling and Decision Support

Michael N. Fienen[1], Bernard T. Nolan[2], Daniel T. Feinstein[3], and J. Jeffrey Starn[4]

Insights from process-based models are a mainstay of many groundwater investigations; however, long runtimes often preclude their use in the decision-making process. Screening-level predictions are often needed in areas lacking time or funding for rigorous process-based modeling. The U.S. Geological Survey (USGS) Groundwater Resources and National Water Quality Assessment Programs are addressing these issues by evaluating the "metamodel" to bridge these gaps. A metamodel is a statistical model founded on a computationally expensive model. Although faster, the question remains: Can a statistical model provide similar insights to a numerical model with faster results?

Metamodeling was developed to overcome long runtimes for sensitivity analysis (Blanning 1975); our focus is decision support applications. Two representative groundwater applications are: (1) the contribution of surface water to wells in shallow groundwater systems (e.g., Fienen and Plant 2014), and (2) unsaturated zone nitrate flux to groundwater (e.g., Nolan et al. 2012).

The first step is to generate a representative sample of input/output combinations from the numerical model over a range of conditions. This variability is especially important when propagating uncertainty to predictions. Variability can be represented by many model runs using different input values or by few model runs with samples scattered in space/time experiencing the range of natural system variability.

In the second step, a statistical learning technique is selected with which a predictive model can be "learned" from the data derived from the model. Techniques include Bayesian networks, artificial neural networks, gradient-boosted regression trees, and support vector machines. These methods learn relationships among inputs and outputs and accommodate expert knowledge to inform whether relationships are also causal. That the dataset is obtained from a process-based model implies a causal connection, where connections among input and output stem from underlying processes simulated by the model. However, not all input variables are explicitly connected to all outputs. In some techniques, such as Bayesian networks, connections among the dataset are defined a priori through expert knowledge. In others, connections are learned and reinforced as the dataset is learned by the algorithm.

The "learning" concept is important because the dataset derived from the process-based model cannot account for every possible configuration of input values encountered in nature. The statistical model is made up of functions relating behavior of output values to inputs, creating predictions based on new input values. As expected, precision is lost in this generalization, but predictions are made with the statistical model nigh instantly. For Bayesian networks, inputs/outputs are probability density functions so the uncertainty of both is explicit and propagated through calculations. Other methods are deterministic, but uncertainty can be considered through Monte Carlo or other techniques.

The final step is to incorporate the metamodel into a decision-making framework. The speed of imperfect but reasonable predictions (often 60% to >90% of the insight from the process-based model [e.g., Nolan et al. 2012; Fienen et al. 2013; Fienen and Plant 2014]), made in nearly real time, is more valuable for screening sometimes than more precise predictions requiring long runtimes. Such screening models can run quickly and easily in a web browser using digital data sources. Alternatively, response maps, graphically depicting predictions over large regions, can be made where input variable values are obtained for a region of similar conditions where the metamodel is considered valid.

[1]Corresponding author: U.S. Geological Survey (USGS), Wisconsin Water Science Center (WIWSC), Middleton, WI 53562; (608) 821-3894; mnfienen@usgs.gov

[2]USGS National Water Quality Assessment (NAWQA) Program, Reston, VA; btnolan@usgs.gov

[3]USGS WIWSC, Milwaukee, WI; dtfeinst@usgs.gov

[4]USGS NAWQA Program, East Hartford, CT; jjstarn@usgs.gov

doi: 10.1111/gwat.12339

Despite advances in groundwater model sophistication, a societal need for quick low-cost answers remains strong. Metamodeling is one approach to leverage insight contained in a complex groundwater model, often with a measure of uncertainty. Tools widely used in many other fields provide an attractive approach for today's rapid decision making.

References

Blanning, R.W. 1975. The construction and implementation of metamodels. *Simulation* 24: 177–184. DOI:10.1177/003754977502400606.

Fienen, M., and N. Plant. 2014. A cross-validation package driving netica with python. *Environmental Modelling and Software* 63: 14–23. DOI:10.1016/j.envsoft.2014.09.007.

Fienen, M.N., J.P. Masterson, N.G. Plant, B.T. Gutierrez, and E.R. Thieler. 2013. Bridging groundwater models and decision support with a Bayesian network. *Water Resources Research* 49: 6459–6473. DOI:10.1002/wrcr.20496.

Nolan, B.T., R.W. Malone, J.A. Gronberg, K.R. Thorp, and L. Ma. 2012. Verifiable metamodels for nitrate losses to drains and groundwater in the Corn Belt, USA. *Environmental Science and Technology* 46: 901–908. DOI:10.1021/es202875e.