Updating the Debate on Model Complexity

Craig T. Simmons, *National Centre for Groundwater Research and Training, Flinders University, Adelaide SA, Australia; and* ***Randall J. Hunt,*** *U.S. Geological Survey, 8505 Research Way, Middleton, Wisconsin 53562, USA, rjhunt@usgs.gov.*

As scientists who are trying to understand a complex natural world that cannot be fully characterized in the field, how can we best inform the society in which we live? This founding context was addressed in a special session, "Complexity in Modeling: How Much is Too Much?" convened at the 2011 Geological Society of America Annual Meeting. The session had a variety of thought-provoking presentations—ranging from philosophy to cost-benefit analyses—and provided some areas of broad agreement that were not evident in discussions of the topic in 1998 (Hunt and Zheng, 1999). The session began with a short introduction during which model complexity was framed borrowing from an economic concept, the Law of Diminishing Returns, and an example of enjoyment derived by eating ice cream. Initially, there is increasing satisfaction gained from eating more ice cream, to a point where the gain in satisfaction starts to decrease, ending at a point when the eater sees *no* value in eating more ice cream. A traditional view of model complexity is similar—understanding gained from modeling can actually decrease if models become unnecessarily complex. However, oversimplified models—those that omit important aspects of the problem needed to make a good prediction—can also limit and confound our understanding. Thus, the goal of all modeling is to find the "sweet spot" of model sophistication—regardless of whether complexity was added sequentially to an overly simple model or collapsed from an initial highly parameterized framework that uses mathematics and statistics to attain an optimum (e.g., Hunt et al., 2007). Thus, holistic parsimony is attained, incorporating "as simple as possible," as well as the equally important corollary "but no simpler."

Complexity will not go away simply by fiat; too many problems require complexity to adequately address societal needs and expectations. In recognition of the need to at times live in a complex world, Anne-Sophie Høyer discussed a new capability to tune 3-D geological modeling for water resource problems (Høyer et al., 2011). Lars Nebel followed with a demonstration of manual and semi-automated ways to manipulate voxel (short for *volume element*, analogue to a 2-D pixel) modeling of complex geology (Nebel et al., 2011). The visualization and investigation of possible realizations can appreciably influence end products such as hydrological models. Hunt's presentation showed models that seem like "big hammers" but fall short in predictive capability because data number and type available were not sufficient to constrain processes important for the prediction of interest (Hunt and Walker, 2011). Although Hunt's example focused on coupled groundwater–surface water modeling, imperfect characterizations of uncertainty can be expected in any modeling endeavor that relies on limited observations to constrain complex and highly parameterized processes.

Complexity needs can be expected to change over time as well, owing to changes in system properties and societal objectives. Denis Peach addressed the complexity resulting from the societal need for a holistic basin-scale integrated model of the River Thames in the UK (Hughes et al., 2011). The basin is characterized by a wide variety of bedrock and sediments that may not be in hydrologic connection even if proximal in location. The river is also actively managed. Rather than putting all eggs in one basket, flexibility is built in from the beginning, as underscored by a reliance on an open standard for linking current and future models.

Henk Haitjema also emphasized the need for flexibility as one looks to find not "true" or "optimal" but "adequate" complexity—where adequate is derived from the societally relevant topic of cost-benefit analysis (Haitjema, 2011). That is, if 80% of the answer can be obtained with 10% of the work, might that be enough to sufficiently answer the question? Haitjema suggested an approach relying on very simple conceptualizations and calculations that are progressively extended until an adequate depiction of the system is reached. However, no one is born knowing how to add all necessary complexity for all problems. Therefore, Haitjema also underscored an associated inherent need of efficient stepwise modeling: heightened development of intuition and hydrosense. Making this insight a primary objective of professional development will assist all modeling endeavors regardless of relative simplicity or complexity.

Fred Molz expanded the philosophical underpinnings of model complexity with an example of "computer-aided thinking." This term suggests a utility for models even if mathematical chaos violates the concept of a single unique reality or where a premise of classic model prediction fails (Molz et al., 2011). Molz described a number of analyses performed to explain plutonium transport in field lysimeters, starting with steady-state, then transient soil-water movement. Simulations included geochemical reactions that account for changes in mobility due to oxidized and reduced conditions. Further extensions to the conceptual model were needed, culminating with plant models and lab experiments of plant transport pathways unknown at the beginning. This work accentuated the place of models in the scientific method and how various hypotheses are formalized and tested. Moreover, Molz demonstrated the importance of interdisciplinary thinking for

GSA Today, v. 22, no. 8, doi: 10.1130/GSATG150GW.1.

today's problems: No matter how sophisticated the representation of soil water movement and abiotic geochemistry, field results could not be duplicated. The field results were only simulated after including movement via plant transport. This underscores potential ecohydrological drivers of many of today's seemingly abiotic problems.

Daniel Abrams compared end-extremes of the complexity-simplicity scale, where insights gained from intensive particle tracking from a 3-D groundwater flow model were also obtained using a simple exponential solution for predictions of watershed-scale transit time distributions (Abrams and Haitjema, 2011). The exponential formulation was also extended to watershed-scale nitrate transit time distributions. These simple conceptualizations are vital for quickly assessing the effects of actions over very large watershed scales—for example, the relation of Upper Mississippi nitrate transport to societally important hypoxia in the Gulf of Mexico. Jeff Starn presented a case in which overly simplistic conceptualizations limited the usefulness of tritium tracer data to investigate drivers of changing water quality (Starn and Green, 2011). However, rather than simply moving to an overly complex model, his work demonstrated a middle ground for complexity and simplicity, one that recognizes the potential artifacts of large model grids but addresses the issue with additional simple methods. Philip Brunner also demonstrated the sliding scale nature of the complexity versus simplicity issue—in some cases, good predictions can be had, not because all parameters were accurately estimated, but because only certain combinations of these parameters were sufficiently accurate (Brunner et al., 2011). Thus, it is possible that a model's predictive power may lose little if it were simplified appropriately.

John Doherty highlighted a need to move beyond the "either/or" framing of the complexity/simplicity question (Doherty, 2011). He cautioned against expectations of widespread utility from any single model conceptualization given an unknowably complex world and today's multifaceted decision making. Often the best model use is to represent the uncertainty in a model prediction, reduce that uncertainty to the extent possible given the available field data, and provide these critical outputs at the speed of real-world decision making.

Chunmiao Zheng offered a succinct summary on his experiences using very simple models for regulatory decision making and extremely complex models and very large field data collection to elucidate salient processes to better understand and define salient simplicity (Zheng, 2011). He took issue with a focus on models being too complex or too simple. The focus instead should be, is the model "good" or "bad"? If we recognize all models are a simplification of reality but have different objectives, some are necessarily more complex than others. The optimal level of complexity for any model should be dictated by its purpose. Such a pragmatic handling of the overarching topic recognizes both the underlying scientific issues of non-uniqueness as well as the societal realm in which most models are consumed. Given that all models should be constructed for a reason, the model objective becomes the primary prism for any and all discussions of model complexity.

REFERENCES CITED

Abrams, D.B., and Haitjema, H., 2011, Groundwater transit time distributions in watersheds with recharge controlled or topography controlled water tables: Geological Society of America Abstracts with Programs, v. 43, no. 5, p. 353, http://gsa.confex.com/gsa/2011AM/finalprogram/abstract_193561.htm (last accessed 25 May 2012).

Brunner, P., Doherty, J., Christensen, S., Simmons, C., 2011, To what extent does the degree of meaningful model complexity depend on observation data?: Geological Society of America Abstracts with Programs, v. 43, no. 5, p. 352, http://gsa.confex.com/gsa/2011AM/finalprogram/abstract_191148.htm (last accessed 25 May 2012).

Doherty, J., 2011, Model complexity and simplicity: Combining the strengths of both: Geological Society of America Abstracts with Programs, v. 43, no. 5, p. 353, http://gsa.confex.com/gsa/2011AM/finalprogram/abstract_192309.htm (last accessed 25 May 2012).

Haitjema, H., 2011, Model complexity: A cost-benefit issue: Geological Society of America Abstracts with Programs, v. 43, no. 5, p. 354, http://gsa.confex.com/gsa/2011AM/finalprogram/abstract_197453.htm (last accessed 25 May 2012).

Høyer, A.S., Jørgensen, F., Møller, R.R., and Christiansen, A.V., 2011, Geostatistically tuned hydrogeological 3D modeling of complex geology—Based on AEM, seismics and boreholes: Geological Society of America Abstracts with Programs, v. 43, no. 5, p. 353, http://gsa.confex.com/gsa/2011AM/finalprogram/abstract_197360.htm (last accessed 25 May 2012).

Hughes, A., Peach, D., Jackson, C., Bricker, S., Mansour, M., Ford, J., and Kessler, H., 2011, Towards holistic basin scale integrated modelling: A case study using the Thames Basin, UK: Geological Society of America Abstracts with Programs, v. 43, no. 5, p. 354, http://gsa.confex.com/gsa/2011AM/finalprogram/abstract_196216.htm (last accessed 25 May 2012).

Hunt, R.J., and Walker, J.F., 2011, The big hammer: Complexity, coupling, and calibration of groundwater and surface water systems: Geological Society of America Abstracts with Programs, v. 43, no. 5, p. 354, http://gsa.confex.com/gsa/2011AM/finalprogram/abstract_197264.htm (last accessed 25 May 2012).

Hunt, R.J., and Zheng, C., 1999, Debating complexity in modeling: Eos (Transactions, American Geophysical Union), v. 80, no. 3, p. 29.

Hunt, R.J., Doherty, J., and Tonkin, M.J., 2007, Are models too simple? Arguments for increased parameterization: Ground Water, v. 45, no. 3, p. 254–262, doi: 10.1111/j.1745-6584.2007.00316.x.

Molz, F., Demirkanli, D., Thompson, S., and Kaplan, D.I., 2011, Mathematical models, computer-aided thinking and the scientific method: Geological Society of America Abstracts with Programs, v. 43, no. 5, p. 353, http://gsa.confex.com/gsa/2011AM/finalprogram/abstract_192897.htm (last accessed 25 May 2012).

Nebel, L., Jensen, N.P., Jørgensen, F., and Møller, R.R., 2011, Manual and semi-automated 3D voxel modeling of complex geology using GeoScene3D: Geological Society of America Abstracts with Programs, v. 43, no. 5, p. 353, http://gsa.confex.com/gsa/2011AM/finalprogram/abstract_191996.htm (last accessed 25 May 2012).

Starn, J.J., and Green, C.T., 2011, Simplifying complexity—Comparison of particle and grid methods with coarse grids: Geological Society of America Abstracts with Programs, v. 43, no. 5, p. 354, http://gsa.confex.com/gsa/2011AM/finalprogram/abstract_196255.htm (last accessed 25 May 2012).

Zheng, C., 2011, Model complexity: A personal odyssey: Geological Society of America Abstracts with Programs, v. 43, no. 5, p. 352, http://gsa.confex.com/gsa/2011AM/finalprogram/abstract_195792.htm (last accessed 25 May 2012).

Manuscript received 18 Feb. 2012; accepted 21 May 2012.