

Towards a comprehensive assessment of model structural adequacy

Hoshin V. Gupta,¹ Martyn P. Clark,² Jasper A. Vrugt,^{3,4} Gab Abramowitz,⁵ and Ming Ye⁶

Received 12 June 2011; revised 19 June 2012; accepted 24 June 2012; published 25 August 2012.

[1] The past decade has seen significant progress in characterizing uncertainty in environmental systems models, through statistical treatment of incomplete knowledge regarding parameters, model structure, and observational data. Attention has now turned to the issue of model structural adequacy (MSA, a term we prefer over model structure “error”). In reviewing philosophical perspectives from the groundwater, unsaturated zone, terrestrial hydrometeorology, and surface water communities about how to model the terrestrial hydrosphere, we identify several areas where different subcommunities can learn from each other. In this paper, we (a) propose a consistent and systematic “unifying conceptual framework” consisting of five formal steps for comprehensive assessment of MSA; (b) discuss the need for a pluralistic definition of adequacy; (c) investigate how MSA has been addressed in the literature; and (d) identify four important issues that require detailed attention—structured model evaluation, diagnosis of epistemic cause, attention to appropriate model complexity, and a multihypothesis approach to inference. We believe that there exists tremendous scope to collectively improve the scientific fidelity of our models and that the proposed framework can help to overcome barriers to communication. By doing so, we can make better progress toward addressing the question “How can we use data to detect, characterize, and resolve model structural inadequacies?”

Citation: Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resour. Res.*, 48, W08301, doi:10.1029/2011WR011044.

1. Introduction

[2] The literature of different modeling communities—groundwater (GW), unsaturated zone (UZ), terrestrial hydro-meteorology (THM), and surface water (SW)—suggests that each applies seemingly different philosophical approaches to dynamical modeling of the terrestrial hydrosphere. These differences in philosophy are evident in the emphases placed on different aspects of the modeling problem. For example, a key step in the application of GW and UZ models is to conceptualize the (one-, two-, or three-dimensional) hydrostratigraphy of the system, whereas SW and THM typically use a one-dimensional model (often disaggregated into sub-basins or tiles) with system structure specified in terms of different stores of water, energy, and carbon.

[3] Such differences in philosophy are also evident in the mathematical forms of the equations used. GW and UZ models use deterministic/stochastic differential equations (such as Richards’ equation and/or Darcy’s law) based ultimately on the Navier-Stokes equations and/or the Buckingham continuity requirement for describing flow through porous media. Meanwhile SW and THM models use sets of loosely coupled differential equations to describe the interdependent hydrometeorological processes at the land surface and within the soil.

[4] Even within each modeling community, there can be disputes about the “appropriate” approach to model development. Such disputes are acutely evident in SW, where debates have raged on the relative merits, applicability, and desirability of “physics-based” (sometimes called physically based) models. More than 20 years ago, *Beven* [1989] and *Grayson et al.* [1992] questioned whether physically based SW modeling is even realistic. We follow Grayson et al. and proceed with their arguments as follows (italicized comments added):

[5] 1. Physically based SW models assume (*incorrectly*) that the conditions under which their equations are derived (a column of homogenous soil) are the same as those in the field, and that the spatial variability of the catchment can be (*acceptably well*) represented by distributed values of the model parameters.

[6] 2. These assumptions are limiting because the mathematics describing hydrological processes is poorly defined *at the model scale* (hillslope, subcatchment, grid) and because there is inadequate data to describe the spatial variability of soil characteristics, even in highly instrumented catchments.

¹Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA.

²National Center for Atmospheric Research, Boulder, Colorado, USA.

³Department of Civil and Environmental Engineering, University of California, Irvine, California, USA.

⁴Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands.

⁵Climate Change Research Center, University of New South Wales, Sydney, New South Wales, Australia.

⁶Department of Scientific Computing, Florida State University, Tallahassee, Florida, USA.

Corresponding author: H. V. Gupta, Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ 85721, USA. (hoshin.gupta@hwr.arizona.edu)

[7] 3. As a result, physically based SW models use “effective” parameters that implicitly represent the impact of spatial variability at spatial scales that are smaller than the model scale.

[8] Others have advanced analogous critiques of physics-based modeling approaches [Beven, 2002; Kirchner, 2006] and some in the SW community have argued the need for a unified theory of hydrology at the catchment scale [Reggiani et al., 1998, 1999; Sivapalan, 2005; Troch et al., 2009].

[9] Such differences in philosophy inevitably become manifest as different interpretations of model structural error (we prefer the term model structural adequacy—see Appendix A), an issue now receiving considerable attention in the context of system identification, data assimilation, and quantification of prediction uncertainty. In SW, model structural adequacy stems from both (a) selection of which hydrologic state variables to represent and how to represent them and (b) choice of equations to compute the hydrologic fluxes. However, it is possible to simplify the representation of storage and flux heterogeneity to the point that the flux equations used become largely empirical guesses [Clark et al., 2008a; Bulygina and Gupta, 2009]. In contrast, model structural adequacy in GW is intimately linked to the amount of detail used to represent the three-dimensional hydrostratigraphy (hydrogeologic structure) of the system [Koltermann and Gorelick, 1996; Hill, 2006; Hunt et al., 2007; Ye et al., 2007; Ye and Khaleel, 2008], and the underlying physics and driving boundary conditions are usually assumed to be correct [National Research Council, 2001; Neuman, 2003; Neuman and Wierenga, 2003; Bredehoeft, 2003, 2005, 2010]. Given this perspective, Doherty and Welter [2010] argue that model structural adequacy can be enhanced simply by inferring additional parameters via calibration. Yet this approach is of limited use in situations where forcing data errors play a significant role. Lack of clarity about such differences in perspective can make it difficult to have a meaningful discussion about the sources and nature of model structural inadequacy.

[10] We suggest the need for a broader perspective. Regardless of apparent differences in philosophy, the GW, UZ, THM, and SW communities all share a common challenge of improving process knowledge and reducing discrepancies between models and the natural system. We present here a systematic conceptual representation of model development that shows how approaches to modeling the terrestrial hydrosphere can be viewed within a common framework. By recognizing and adopting a unified conceptual framework, the communities can communicate more effectively in a common language and jointly contribute to addressing the fundamental question “How can we use data to detect, characterize, and resolve model structural inadequacies?”

[11] To be clear about scope, this paper is not a comprehensive review of the literature on modeling of the terrestrial hydrosphere. References cited are selected as examples relevant to issues we wish to discuss. For pedagogical reasons we refer only to the broad divisions of GW, UZ, THM, and SW; this is not to ignore other elements of the terrestrial hydrosphere but to help focus the discussion. Similarly, we omit discussion of important issues such as data informativeness and of distinguishing between systematic data errors and model structural deficiencies; we leave these for later discussion.

[12] Section 2 introduces a conceptualization of the model building process as a framework for discussing common challenges faced by the community. In section 3, we point to the need for a pluralistic definition of model structural adequacy. Section 4 examines how model structural adequacy has been addressed to date, with many of the steps discussed in section 2 remaining poorly addressed. Section 5 discusses the implications of our framework and makes several recommendations toward stimulating discussion and progress. Finally, section 6 raises complementary issues that need to be examined and discussed.

2. Systematic Conceptualization of the Model Building Process

[13] Broadly speaking, there are generally considered to be three formal stages to model building (although we will later suggest that there are really five formal steps). Prior to these is a preliminary, informal stage—development of a “perceptual model” of the system—that involves a tight loop of purely sensory perceptions coupled with interpretations of the data. This perceptual/interpretive process can be strongly influenced by prior concepts (ideas about reality), and modified by new ones that form during the investigative process (see Roman et al. [1998], Beven [2001], Neuman and Wierenga [2003], and others). In our characterization, this “perceptual model” exists only in the mind of the investigator (being colored by mental concepts, it might actually be called a “perceptual-conceptual” model). As such, it cannot readily be subjected to formal analysis (since formal analysis requires symbolic representation). At the formal level, the three formal stages involve development of the following:

[14] 1. Conceptual model: Summarizes our abstract state of knowledge about the structure and workings of the system.

[15] 2. Mathematical model: Defines the computational states, fluxes, and parameters of the system and the choices regarding how system processes will be mathematically handled.

[16] 3. Computational model: Provides numerical solutions for specific initial states, material properties (parameters), and boundary conditions.

[17] All of these stages have been discussed previously [see Iliev, 1984; Beven, 2001; Anderson and Woessner, 2002; Neuman and Wierenga, 2003; Refsgaard et al., 2006; Gupta et al., 2008; Clark and Kavetski, 2010; Kumar, 2011], albeit with varied terminology. (For example, Beven [2001] combines the perceptual and conceptual models into one stage called the “perceptual model” and refers to the mathematical model as a “conceptual model” and the computational model as a “procedural model.” We believe our terminology to be more precise, with less scope for introducing confusion, and encourage its use by the community.) However, previous formulations do not investigate each stage in depth, nor do they explicitly address how these stages are manifest in different communities (including commonalities and differences between them). While we do not debate the critical importance of the development of “perceptual” models, this paper will not dwell on them because of their lack of explicit formal expression (theory requires explicit expression using abstract symbols that can be manipulated). To the extent that our *expressions* of the perceptual models can be improved through the investigative

process, this improvement takes form as modifications to the consequent “conceptual models” which *can* be subject to formal analysis. Here, we focus on the three formal stages, whose investigation we seek to promote. Below we present a more penetrating description of each formal stage, illustrated with examples from GW, UZ, SW, and THM.

2.1. The “Conceptual” Model

[18] A model is a simplified representation of a system, with twofold purpose to enable reasoning within an idealized framework and to enable testable predictions of what might happen under new circumstances (this paragraph closely follows the ideas in *Gupta et al.* [2008]). Building upon our (perceptual) understanding of the system, one or more “conceptual models” emerge, represented usually as verbal and pictorial descriptions. A complete conceptual model of a system will include a clear specification of relevant system boundaries, inputs and outputs, state variables (sometimes called prognostic variables), physical and behavioral laws to be obeyed (conservation of mass, momentum, etc.), facts to be properly incorporated (e.g., spatiotemporal distribution of “static” material properties such as soils), uncertainties to be considered, and simplifying assumptions to be made. In other words, the conceptual model characterizes the “architecture” of the system [*Bulygina and Gupta*, 2011]. Note that relationships among elements need not be rigorously specified and investigated but are conceptually explained through drawings, maps, tables, papers, reports, oral presentations, etc.

[19] The conceptual model summarizes, and depends upon, our abstract state of knowledge (degree of belief) about the structure and workings of the system. Our “degree of belief”—i.e., our process understanding—depends fundamentally on training and experience, including the effectiveness of the dialog between field hydrologists and modelers [*Seibert and McDonnell*, 2002]. Alternative conceptual models can represent competing hypotheses about the structure and functioning of an observed system, conditioned on qualitative and quantitative observations, and on prior facts, knowledge, and ideas. Of course, the conceptual models may not be faithful expressions of the aforementioned perceptual model(s) and are often simplified expressions of such. Such simplification can be due to lack of knowledge, ideas, or imagination about how to express the perceptual model in a proper way, but may also be due to explicit decisions to build simplified expressions based on assumptions regarding what is more important and less important. Together, the conceptual model of the system and the conditioning prior knowledge form the rudimentary levels of “theory” about the system.

[20] To be more specific, development of the conceptual model actually involves two steps, the order of which may vary (or even be comingled) among disciplines:

[21] 1. A conceptual model of the *physical structure of the system* in absence of water, energy, and/or other less solid quantities whose storage, movement, and behaviors the system mediates. In hydrology, this refers primarily to the geology and topography of the porous (and not so porous) media that constrain and direct the storage and movement of water. This is mainly the soil matrix but can include other aspects such as vegetation, engineered structures, etc. On the other hand, for a model of atmospheric flows, the physical structure refers mainly to the structural nature of the upper

and lower boundary conditions (land surface topography including vegetation and engineered structures, ocean surface, and so on), as well as other impediments, obstacles, or structures the air must navigate.

[22] 2. A conceptual model of the *process structure of the system*. This specifies the dominant processes that are mediated by the physical structure of the system and, importantly, how different factors affect the storage and movement of water and energy through the system. For example, our conceptual model of the process structure of the system may include a description of the role of bedrock topography in controlling the storage of water on hillslopes and subsurface flow during storm events, and it may include description of the heterogeneity in flow paths within a catchment. More generally, the process structure of the system arises as a consequence of interactions between water, energy, and other mobile constituents, with the physical structure. Underlying physical principles (physics) govern interactions of the fluids with the physical structure, giving rise to the characteristic behaviors of the system.

[23] In the longer term, the system processes can possibly act to modify the physical structure of the system, which can have profound implications for predictability [*Peters et al.*, 2007; *Kumar*, 2011]. However, under the stationarity paradigm (now severely in question; see *Milly et al.* [2008]), it is common to assume that the physical structure of the system is not noticeably affected on time scales relevant to the modeling study. Under this assumption, the “physical structure” of the system is sometimes thought of as those aspects that are (assumed to be) time-invariant over the period of simulation. However, many other aspects of model structure can also be assumed time-invariant (process structure, spatial variability structure, equation structure, etc., see below) and the term “time-invariant” does not (by itself) cover this distinction well.

[24] The practical outcome of these two steps is a conceptual model and a working definition of the “conceptual states and fluxes” of the system. To be clear, this definition does not require that (a) spatial and temporal “resolution” be specified (including size/shape/extent of model elements and temporal discretization of model fluxes, etc.) or (b) mathematical equations be formally selected, although some specification of physical laws may take place (conservation of mass and momentum; laws of thermodynamics; fluid flows governed by Navier-Stokes). Major decisions involve the architecture (extent and structure) of the modeling domain and which processes to include/exclude.

[25] This conceptual model, along with the numerical, graphical, and other results generated using the computational model developed from it (discussed below), defines the level at which communication actually occurs between scientists, stakeholders, decision makers, policy analysts, and others. Of course, the conceptual model can be updated as new data and information become available.

2.2. The “Mathematical” Model

[26] The modeler next proceeds toward realization of a mathematical model. In the past, when digital computing was not readily available and/or the theory insufficiently advanced, it was common to construct an actual physical model of the system to perform the desired “computations” (e.g., scale models of Dam systems designed to obey scaling

laws). Today, it is common to construct a mathematical model for use in analytical or digital computation.

[27] Development of the mathematical model also involves two steps:

[28] 1. A mathematical model of the *spatial variability structure of the system*. This specifies how horizontal and vertical variability is represented in the model and (consequently) determines the properties of the associated porous and nonporous geohydrologic media.

[29] In GW the system is commonly treated as layered 2-D, or fully 3-D, with some spatial resolution selected to represent the variability structure. Many UZ applications are 1-D [Šimůnek and van Genuchten, 2008] since 2-D/3-D modeling is only productive if a significant horizontal flow component is observed. However, there is a trend for 2-D and 3-D applications [Vesselinov et al., 2001a, 2001b; Vrugt et al., 2001; Yeh et al., 2002; Vrugt et al., 2004; Ye and Khaleel, 2008; Klaus and Zehe, 2010], although more difficult than in GW due to the increased computational cost of solving variably saturated flow equations, and the challenge of accurately representing the spatial variability of soil hydraulic properties. Some recent efforts even have 3-D representations from the bedrock to the top of the atmosphere [Maxwell et al., 2011], effectively combining modeling approaches from GW, SW, and THM. In THM, 1-D vertically layered representations are typically assumed, with grid cells disaggregated into land-cover categories and/or elevation bands and horizontal resolution dictated by the atmospheric model [Oleson et al., 2010; Krinner et al., 2005; Wang et al., 2011]. In SW, where this is often called catchment architecture, a wide range of approaches is used, from spatially lumped/vertically layered (1-D bucket-style) to spatially distributed/vertically layered (quasi 3-D) and fully 3-D (so-called physically based).

[30] To be clear, while this step defines the kind of spatial variability to be represented, the spatiotemporal resolution for computing is not yet specified.

[31] 2. A mathematical model of the *equation structure of the system*. This selects the equations used to represent the dynamics of each process and their interactions. Sometimes, this may involve decisions regarding whether temporal variability is to be considered (as in steady state versus transient state modeling). Since each equation will have parameters representing the properties of the system, together with the previous step this also defines the spatial variability structure of system parameters.

[32] While in some cases (such as GW and UZ) the basis for these equations will have been determined during conceptual modeling (e.g., Darcy's Law and Richards' equation are macroscopic solutions of Navier-Stokes; Neuman and Wierenga [2003] call this conceptual-mathematical modeling), this step may still involve far-reaching simplifications or approximations (such as neglecting pore-scale processes and the effects of contact angle, dispersion, and/or diffusion on saturated/unsaturated flow and perhaps most importantly preferential flow). In other cases, the basis for selecting equations may be less constrained by computational theory. For example, SW, UZ, and THM commonly use macroscopic solutions to the Navier-Stokes equations (e.g., Richards' and Boussinesq equations) to represent flow through soils and plants and empirical parameterizations for interception.

[33] Nonetheless, the selected equation structure may be some combination of ordinary differential equations (ODEs),

partial differential equations (PDEs), and/or stochastic ODEs/PDEs, depending in part on decisions regarding spatial variability structure.

[34] The practical outcome of these two steps is a working definition of the "computational states, fluxes, and parameters" of the system and a choice regarding how the "system processes" will be mathematically handled. To be clear, (a) the spatial variability structure of the system/model parameters need not correspond to the spatial variability structure of the states and fluxes (parameters can be treated as homogeneous/lumped even if states and fluxes are treated as spatially distributed), and (b) the "computational resolution" is not introduced explicitly at this stage. Major decisions involve the method for representing spatial variability (spatial structure) and the equations used to model process dynamics ("physics" of the system).

[35] This mathematical model forms a basis for discussion among scientists; it is rare for stakeholders, decision makers, and policy analysts to be intimately involved.

2.3. The "Computational" Model

[36] Implementation of the mathematical model results in a computational model that, when realized as a computer program, can provide numerical solutions for specific initial states, material properties (parameters), and boundary conditions. Here, the details include (a) selecting a numerical formulation for representing the spatial relationships (finite difference or finite element), (b) defining the spatial resolution for computations within the model domain, and (c) selecting a procedure for time integration of the governing model equations.

[37] It is important to recognize that, in making these choices, strong interactions can exist between choice of model equations and choice of spatial and temporal scales [e.g., Clark et al., 2011a] because the decision to explicitly resolve a given process will dictate a minimum spatial and temporal resolution of computation. For example, Mott and Lehning [2010] point out that explicit simulation of snowdrift formation requires a horizontal resolution of <5 m.

[38] The practical outcome of this stage is a "numerical basis for solving the model equations." The approach varies somewhat with modeling community. In GW and UZ, solutions are typically achieved using finite difference [Harbaugh, 2005; Wang and Anderson, 1982] or finite element methods [Pinder and Huyakorn, 1977; Lin et al., 1996; Yeh, 1999], where the differential equations are implemented using global solvers. In THM, solutions are typically achieved via loose coupling of the differential equations; many implementations decouple processes and use a different local solver for each subprocess [see Oleson et al., 2010]. While this allows the numerical strategy to be tailored to the dynamics of each subprocess, it requires arbitrary selection of the solution sequence and does not properly account for sub-time-step interactions, and feedbacks among subprocesses, that can give rise to numerical errors or artifacts. In SW, many model codes use unreliable numerical schemes that can unnecessarily complicate model calibration, and cause parameter estimates to compensate for numerical errors [Kavetski et al., 2003; Kavetski and Clark, 2010; Schoups et al., 2010]. Overall, the computational strategy is a critical step in model development and deserves as much scrutiny and care as the development of conceptual and mathematical models.

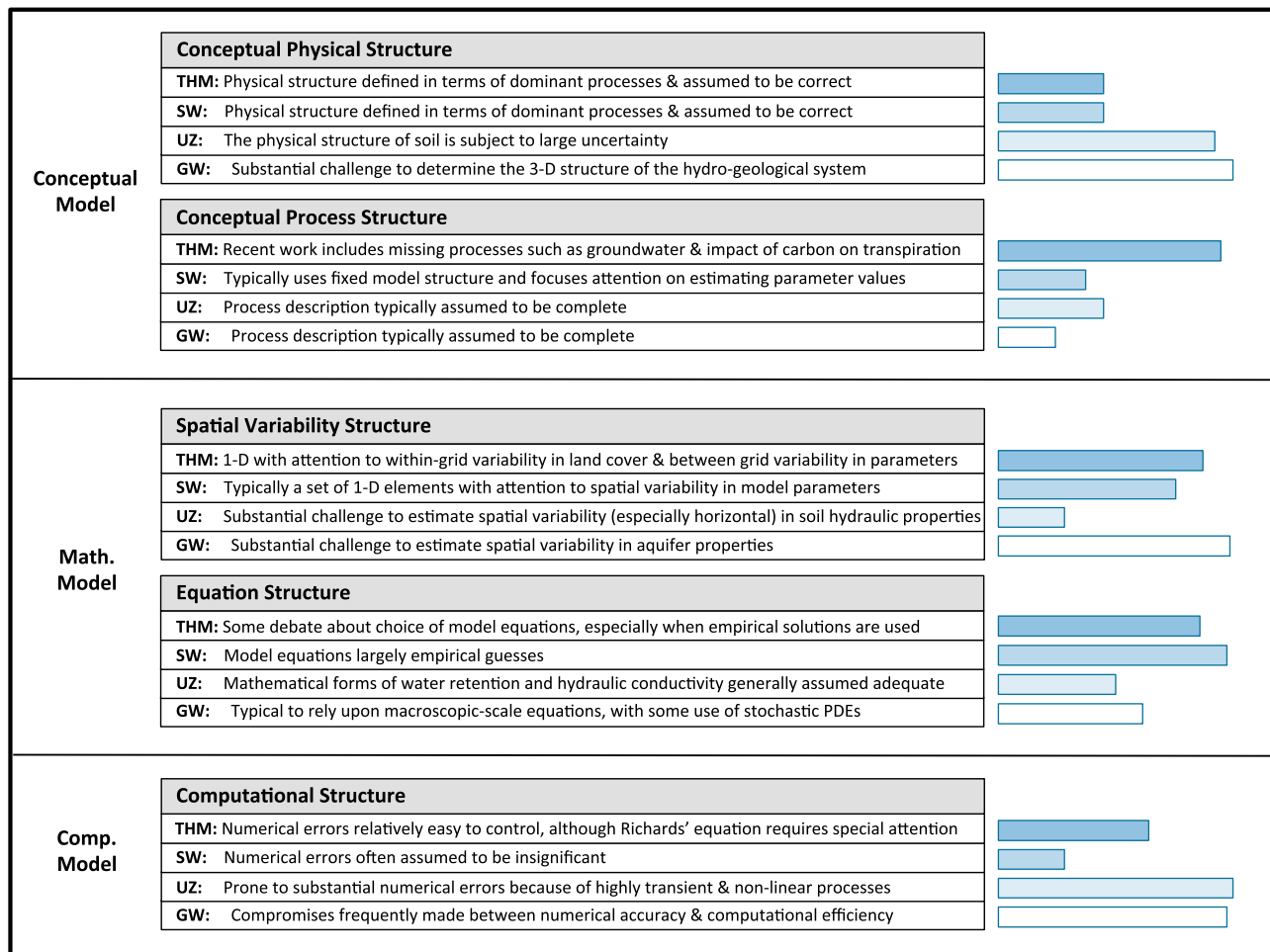


Figure 1. Subjective assessment of the attention/emphasis (indicated by length of bars) given by different modeling communities to various sources of model inadequacy; colors are used to distinguish between THM, SW, UZ, and GW.

2.4. Common Challenges Faced by the Community

[39] Considering the steps in sections 2.1 and 2.2 to deserve focused attention, we can now think of the model building process as really having five formal steps (conceptual physical structure, conceptual process structure, spatial variability structure, equation structure, and computational structure). Figure 1 illustrates the typical amount of attention given by different communities to each of these aspects of model adequacy. The first two steps (conceptual model) specify the architectural extent and topology of the modeling domain and what processes to include/exclude, resulting in a conceptual definition of the states and fluxes, without concern for spatial or temporal resolution or equations to be used. Whereas all models of the terrestrial hydrosphere are based on these first two developmental steps, different communities place different emphases on specific decisions and so develop these conceptual components in different ways. For example, GW may focus on the underlying (hydrostratigraphic) structure of the aquifer (confined or unconfined), UZ may emphasize the properties of the porous medium (layering and pore size distribution), while THM and SW may focus on specifying the dominant surface and subsurface processes (process structure). Nonetheless, the common challenge is that

of “epistemic uncertainty” arising from incomplete (due to lack of knowledge) or inadequate characterization of the system [Beven, 2008; Clark et al., 2008a, 2008b].

[40] The next two steps (mathematical model) specify the spatial representation and form of the process equations, resulting in explicit mathematical definitions of the states, fluxes, and parameters and their spatial resolution. Although different applications may call for different choices regarding spatial configuration and parameterization, the development of a mathematical model of variability and equation structure follows essentially the same approach in all communities. In regards to model adequacy, the challenge is in proper specification of “spatiotemporal detail” and “parameterization” (degree of model complexity).

[41] The final step (computational model) expresses the precise sequence of operations required to generate solutions. It is rare for any stakeholder, decision maker, or policy analyst to care about or be involved at this level, and computational details are usually only the subject of discussion among “nerds” who understand and care about the details of numerical implementation.

[42] Of course, in any given modeling process, the decisions required to implement these five steps can render them

(explicitly or implicitly) interdependent. Assumptions about dominant processes/structures can imply assumptions about the dimensionality of flow process and symmetry in the system, which can in turn determine the choice of differential equations. Nonetheless, the point of breaking them out into separate steps is to emphasize the differences (and relationships) and further that the model development process tends to be strongly directional through the steps as indicated.

3. The Need for a Pluralistic Definition of Model Structural Adequacy

[43] In constructing a model, how does one decide when the model is structurally adequate? Arguably, this “evaluation-of-structure” problem is commonly misunderstood, a key reason being that it can be viewed from at least three somewhat independent perspectives; we will call these the “engineering” viewpoint, the “physical science” viewpoint, and the “systems science” viewpoint. We discuss each in turn.

3.1. The “Engineering” Viewpoint

[44] The engineering view of model structural adequacy is firmly focused on functional adequacy and usually takes a decision-making perspective. It is often defined in the context of a particular application, so that adequacy benchmarks are necessarily usually defined by individual practitioners, rather than agreed upon by the community. In this sense, an adequate model structure is equivalent to a “valid” model as defined by *Oreskes et al.* [1994] and *Medlyn et al.* [2005]. Deficiencies in structural adequacy are often expressed as uncertainties relating to a risk analysis in the decision making process. As long as the uncertainties are well characterized, one can effectively and efficiently emulate the system input-output behavior using nonphysically realistic models (e.g., databased regression models). Clearly, however, this approach can cause problem in nonstationary systems.

3.2. The “Physical Science” Viewpoint

[45] This viewpoint defines model adequacy purely in terms of consistency with the physical system. Model variables, states, and parameters should correspond to physical quantities. Difficulties can arise when aspects of a model are perceived to be physically representative but when no data to support or falsify them is available.

3.3. The “Systems Science” Viewpoint

[46] This viewpoint is essentially a databased hybrid of the first two. It stresses physical consistency to the extent that it is observable in data and values the simplicity of the engineering approach by stressing the principle of parsimony. Typically, a simpler model structure that is consistent with the physical system and performs as well as a more complex model structure is preferred. This approach may focus on techniques that quantify the information content of data as a means to characterize which aspects of system behavior are evident in observational data.

3.4. The Need for Community Attention to This Problem

[47] Despite considerable progress in particular aspects of model structure evaluation, it is clear that the hydrological community as a whole is still far from any consensus

regarding a consistent methodology for model evaluation. The development of consensually agreed benchmarks, including the establishment of a “list of key characteristics” to be reproduced, should, over time, help to characterize structural inadequacy and direct model improvement. While the THM community has begun to recognize the benefits of this approach (e.g., the International Land-Atmosphere Benchmarking activity, www.ilamb.org), the process is still in its infancy. Through discussion and debate it should be possible to move toward a consensual, more pluralistic, and better structured definition of model structural adequacy that acknowledges multiple disciplines and viewpoints (engineering, physical science, systems science) while recognizing the common basis underlying model development (section 2).

4. How Model Structure Adequacy Has Been Addressed to Date

[48] To examine how model structure adequacy challenges have been addressed in the literature, we examine the steps in reverse order (arguably that of increasing difficulty). We find that many of these steps remain poorly understood and that only ad hoc approaches for dealing with them have been proposed.

4.1. Numerical Error

[49] The problem of “numerical error” is easily avoided by carefully applying robust numerical methods when constructing the computer code. Since it can sometimes be necessary to compromise between numerical robustness and computational efficiency, it is important that numerical accuracy be assessed, characterized, and quantified, and the implications for model predictions documented. Each community has literature dealing with this issue. We provide only a few examples.

[50] In GW, *Zyvoloski and Vesselinov* [2006] illustrate the effects of numerical resolution on Darcy Law computations and point out “To keep numerical models computationally efficient, compromises are frequently made in the model development, particularly, about resolution of the computational grid and numerical representation of the governing flow equation. The compromise is required so that the model can be used in calibration, parameter estimation, performance assessment, and analysis of sensitivity and uncertainty in model predictions.” They show that use of different grid resolutions can significantly impact parameter estimates during inverse estimation, and that a control volume finite element approach may be superior to the finite difference approach in terms of accuracy of the parameter estimates for a given grid resolution. For other discussions of numerical issues, see *Lenhard et al.* [1995], *Zheng and Bennett* [2002], *Wu et al.* [2004], and *Konikow* [2011].

[51] In UZ and THM, a key consideration is the method for solving Richards’ equation. While a number of iterative and noniterative schemes are in use, the best way to improve the numerical solution remains an active area of research. The amount of numerical error depends on the system being simulated and the number of nodes used to represent the governing equations (e.g., Navier-Stokes and advection-dispersion). UZ models are prone to significant numerical errors (even with the mass-conservative scheme of *Celia et al.* [1990]), particularly when simulating contaminant and/or tracer transport, due to transient behaviors, high

nonlinearity of the retention and hydraulic conductivity functions, and the significant heterogeneity of soil. To limit numerical error (particularly numerical diffusion), dense discretization and hence a large number of nodes is required, which poses significant computational challenges [Kollet *et al.*, 2010]. This makes it difficult to simulate complex 2-D or 3-D subsurface (hydrogeological) systems. While approaches such as particle tracking offer promise, each comes with its own set of limitations [Roubinet *et al.*, 2010]. In THM, Boone and Wetzel [1996] and Lee and Abriola [1999] discuss the impacts of low-dimensional implementations of Richards' equation, and the sensitivity to infiltration parameterizations and the representation of the lower boundary condition.

[52] In SW, poor numerical schemes have been shown to result in incorrect evolution of spatiotemporal state/flux trajectories (wrong numerical solutions, as opposed to incorrect simulations of reality), and to result in high-frequency oscillations of the Likelihood function, causing severe problems during model identification [Kavetski *et al.*, 2003, 2006a, 2006b; Kavetski and Clark, 2010; Kavetski and Kuczera, 2007; Schoups *et al.*, 2010].

4.2. Adequacy of Equation Structure

[53] The problem of equation structure adequacy is commonly addressed using stochastic differential equations to hedge against uncertainties associated with a precise deterministic formulation. This approach is common in GW [Unny, 1989; Serrano and Unny, 1987a, 1987b, 1988, 1990; Gelhar, 1989; Dagan, 1989; Dagan and Neuman, 1997; Zhang, 2002; Rubin, 2003] where an adequate description of groundwater hydraulic head and solute concentration can be given only in a stochastic sense, and the assessment of uncertainty is integral to analysis and modeling (being as important as the predictions themselves). In UZ, attention to equation structure uncertainty is usually focused on the mathematical forms of the water retention and hydraulic conductivity functions, while soil moisture flow, root water uptake, and infiltration are generally assumed to be adequate. While various stochastic descriptions of the vadose zone have been proposed [Dagan and Bresler, 1979; Jury, 1982; Yeh *et al.*, 1985a, 1985b, 1985c; Destouni and Cvetkovic, 1989, 1991; Harter and Yeh, 1996a, 1996b; Harter and Zhang, 1999; Russo and Fiori, 2009], errors in data, parameters, and model structure have only recently been investigated [Vrugt *et al.*, 2005b; Pan *et al.*, 2009; Wöhling and Vrugt, 2011].

[54] In SW, uncertainty in the equation structure is commonly represented via stochastic perturbations to the model state variables and/or parameters [Vrugt *et al.*, 2005a, 2006; Moradkhani *et al.*, 2005a, 2005b; Beven and Freer, 2001] or by using stochastic time varying parameters [Kuczera *et al.*, 2006; Reichert and Mieleitner, 2009]. It is often argued that the mean (or median) of a stochastic ensemble can provide a better estimate than any individual ensemble member because the ensemble better represents the nonlinear behavior and subgrid variability characteristics of natural systems. Further, Hollenbeck and Jensen [1998b] provide experimental evidence that repeated experiments on the same soil sample can result in quite different estimates of the hydraulic parameters and functions, bringing into question assumptions of both stationarity and determinism.

[55] Complementing the stochastic-equation approach is the ensemble-equation approach that samples multiple alternative equations for each process from a prespecified set [Vrugt and Robinson, 2007; Clark *et al.*, 2008a; Fencia *et al.*, 2008; Niu *et al.*, 2011]; we discuss ensemble conceptual models later in section 3.4. The ensemble-equation approach provides an opportunity to reject equation hypotheses that are inconsistent with data and process understanding, while retaining ones that provide plausible representations of the natural system [Clark *et al.*, 2011b]. Due to uncertainties in data and deficiencies in process understanding, it is generally impossible to identify a single set of model equations, and the ones not rejected by hypothesis testing implicitly represent uncertainty in the equation structure. The main problem is that due to epistemic uncertainty this set of equations will generally be incomplete (not fully span the space of viable equations), potentially resulting in bias and inadequate characterization of uncertainty.

[56] In this context, an interesting approach is the attempt to infer model equation structures directly from data (inverse estimation of equation structure). This is common in SW where streamflow recessions are analyzed to infer storage-discharge relationships [Brutsaert and Nieber, 1977; Ambroise *et al.*, 1996; Atkinson *et al.*, 2003; Rupp and Selker, 2006; Clark *et al.*, 2009; Harman *et al.*, 2009; Kirchner, 2009]. Alternatively, the Bayesian Estimation of Structure approach [Bulygina and Gupta, 2009, 2010, 2011] seeks to use data to refine all of the equations in a model, not just those associated with streamflow recessions. Beginning with a set of assumed process equations, the approach represents them as a prior joint probability density function (pdf), applies Bayesian Data Assimilation to obtain a data-augmented posterior pdf, and then extracts a "corrected" set of process equations. In a different approach, Young [1998, 2003], Young and Ratto [2008], Young *et al.* [1996], and McIntyre *et al.* [2011] use databased mechanistic modeling to progressively identify the nonlinear forms of process equations from data. The approach detects time and process dependent patterns of parameter variability and modifies the relevant model equations to remove such patterns, but essentially ignores a prior model concept. For more complex models, equation structure inadequacies can be detected and corrected by analysis of the spatiotemporal properties of the state updates [Vrugt *et al.*, 2005a; Vrugt and Clark, 2006].

[57] In THM it is not common to infer model equation structures from data. However, in GW the use of dimensionless drawdown plots of pumping test data to infer structural characteristics such as wellbore storage, leaky confining layers, penetrating hydraulic fractures, faults, stratified layers, etc., has been common [Hantush, 1967; Neuman *et al.*, 2004; Mishra and Neuman, 2010], and in reactive transport modeling, a recent trend is to identify structures of chemical reactions from data [Matott and Rabideau, 2008].

[58] An interesting aspect of inferring model equations directly from data, is that it attempts to deal with problems of scale mismatch that arise when process equations derived at one scale (e.g., the laboratory soil column) are applied at a very different model scale (e.g., the hillslope, watershed, or grid). While innovative solutions have been proposed to derive model equations directly at the spatial scale of interest [Reggiani *et al.*, 1998, 1999], their general applicability has

yet to be determined, and much uncertainty regarding this issue remains [Beven, 2006].

4.3. Adequacy in Representation of Spatial Variability

[59] The representation of spatial variability is subtly different among different modeling communities. GW and UZ typically use three-dimensional finite difference/element representations of the spatial relationships between states and fluxes, while variability of soil and vegetation properties (parameters) is inferred from data by direct or inverse methods. To account for the enormous heterogeneity in aquifer properties, spatial parameter variability is considered necessary and a significant portion of the literature deals with methods for its estimation; *Doherty and Welter* [2010] recently suggested that all the important GW modeling uncertainties can be effectively resolved simply by adding parameters to be estimated via inference. However, considerable structural adequacy dangers can be associated with adding unmeasurable parameters [Ginzburg and Jensen, 2004] and, as mentioned earlier, simply adding parameters to capture spatial detail does not address errors associated with boundary conditions and/or structural inadequacies of the governing equations.

[60] While some SW and THM models use a fully 3-D representation, a quasi 3-D approach that represents the vertical movement of water and energy via multiple one-dimensional columns (grids or subbasins) is more common in these communities, with runoff fluxes routed separately through an overland flow and channel network. In this approach, the horizontal flow between different soil columns is effectively ignored. A key decision is which processes to represent explicitly (e.g., spatial variability in precipitation, temperature, and solar radiation; spatial variability in the interception capacity of vegetation), and which processes to represent implicitly (e.g., via use of pdfs to simulate aggregate impacts of unresolved processes). These decisions regarding process complexity then determine the spatial scale necessary for the model simulations [Clark et al., 2011a]. While some investigations have suggested the possibility of an ideal model scale or representative elementary volume for representing within-grid and between-grid variability [Band and Wood, 1988; Blöschl et al., 1995], the notion of an ideal model scale has proved difficult to implement in practice. This may be because (a) various catchment processes occur at different scales, and (b) it can be difficult to implement models at the spatial scale for which the equations were actually derived (e.g., the point scale as in Richards' equation), due to both computational constraints and limited knowledge regarding the spatial variability of parameters. In practice, therefore, the spatial size of model units is commonly selected based on computational considerations and data resolution, while attention is focused on the spatial characterization of parameter variation.

[61] In THM, spatial parameter variability is typically represented at regional scales based on soils and vegetation data [Sen et al., 2001], whereas in SW it is usually inferred from some combination of prior estimation using soils and vegetation data [Koren et al., 2000; Samaniego et al., 2010] and spatially regularized inverse estimation from rainfall-runoff data [Pokhrel et al., 2008, 2009; Pokhrel and Gupta, 2010]. Methods for prior estimation of parameters are limited by problems in conceptualizing the relationship of parameters to data, and it has therefore been common to implement

parsimonious representations of parameter heterogeneity so that inverse methods can be applied [Boyle et al., 2001; Feyen et al., 2008; McMillan and Clark, 2009].

[62] An important, and common, problem arises when the degree of spatial variability resolvable by inverse methods is limited by the amount of information available in the available data. In GW and UZ studies, inadequate availability of head and concentration measurement locations can make it impossible to resolve the underlying spatial variability of the aquifer and vadose zone [Konikow and Bredehoeft, 1992] let alone distinguish between alternative conceptual models [Refsgaard et al., 2006] (see sections 3.4 and 3.5). While there has been much research on calibration of GW models, the problem of evaluating competing representations of spatial variability is not well understood [Wood, 2005]. Open questions remain regarding type and amount of data needed to distinguish between alternative models and of how to provide reliable estimates of model plausibility [Tsai and Yeh, 2004; Tsai and Li, 2008; Ye et al., 2007; Ye and Khaleel, 2008]. In SW, the damping/dispersive effects of routing can cause catchment outlet hydrographs to become insensitive to spatial variability in the parameter and rainfall fields [Kirkby, 1975; Beven and Wood, 1983; Pokhrel and Gupta, 2011]; consequently, basin-mean parameters can be sufficient for simulating flows at the outlet (though not at interior points). Meanwhile, such "lumping" can result in severe parameter errors that diminish correlations between the parameter estimates and catchment properties even when theoretically strong relationships exist [Kling and Gupta, 2009]. Together, these issues seriously complicate attempts at parameter regionalization necessary for predictions in ungaged basins.

4.4. Adequacy in Conceptual Representation of Process Structure

[63] Perhaps the most interesting source of model structure inadequacy arises from missing state variables and processes. Put simply, how can we know if all of the important states and processes are included (and included correctly) in the model? In other words, how do we know that our model resolves the correct dimensionality of the state-space? A compelling recent example of this problem comes from THM, where numerous groups have demonstrated that simulations of land-atmosphere fluxes improve when explicit representations of the water table are incorporated [Liang et al., 2003; Niu et al., 2007; Pan et al., 2008], affecting the storage and transmission of water in the unsaturated zone and providing a more complete representation of system behavior.

[64] In general, the adequacy of conceptual representations of state and process structure is usually assessed by comparing models of differing complexity. This can involve comparing very different conceptualizations of the state-space, in some cases with significantly different definitions and/or dimensionality. THM has examined increasingly complex representations of the land-atmosphere energy balance [Desborough, 1999; Leplastrier et al., 2002; Pitman et al., 2003; Xia et al., 2002; Hogue et al., 2006]. In SW, the "top-down" approach to model development explores increasingly complex representations applied at increasingly finer time scales [Jothityangkoon et al., 2001; Atkinson et al., 2003; Eder et al., 2003; Farmer et al., 2003; Bai et al., 2009]. Assessments of state/process complexity have

also been undertaken in experimental basins, where tracer and groundwater levels data are used to identify the range of processes to include in hydrologic models [Son and Sivapalan, 2007; Fenicia et al., 2008].

[65] An increasingly popular approach to acknowledging uncertainty in state/process representation, now widely used in many fields, is the ensemble-model approach (distinct from the ensemble-equation approach discussed in section 3.2) in which multiple model formulations (each having a specific set of selected state/process representations) are run to assess prediction uncertainty. It can be shown theoretically that (when the models of the ensemble are not calibrated) the mean of the model ensemble can provide better average performance than any individual ensemble member but only if the members are selected to provide characteristically different simulations [Winter and Nychka, 2009]. In SW, multimodel ensembles have been shown to provide improved streamflow forecasts when the set of models has sufficient diversity [Georgakakos et al., 2004]. In THM, output from multiple land-surface models has been used to provide estimates of global terrestrial water and energy balance, with the spread among models used as a metric to describe the uncertainty in representing energy and water balance [Dirmeyer et al., 2006]. In GW, the Maximum Likelihood Bayesian Model Averaging (MLBMA) approach assigns relative weights to each model based on their ability to explain the observed data [Neuman, 2003; Ye et al., 2004, 2008a]; the relative weights can also be obtained using other methods [Poeter and Anderson, 2005; Poeter and Hill, 2007; Tsai and Li, 2008]. Applications to UZ and SW have demonstrated the superiority of Granger–Ramanathan averaging [Granger and Ramanathan, 1984] for ensemble modeling [Diks and Vrugt, 2010].

[66] However, while model averaging provides a framework for explicitly considering (conceptual) model uncertainty, it currently lumps all errors into a single misfit term and does not provide insights into model structural adequacy. As such, it is an engineering approach to resolving structural model uncertainty; individual models are typically selected in an ad hoc manner, without considering the extent to which the models span the model hypothesis space [Clark et al., 2011b].

4.5. Adequacy in Conceptual Representation of Physical Structure

[67] This final source of model structure inadequacy arises from incorrect specification of the elements of physical system structure upon which all subsequent steps are based. Put simply, how can we know if all of the important physical structures that constrain and mediate the movements of water, energy, and chemicals through the system are included and their spatiotemporal geometry properly represented? In general, these elements include all impermeable (typically rigid) and permeable boundaries, porous and potentially deformable media, elements and composition of vegetation (roots, stems, leaves, etc.), and even (in cases involving reactive processes) biochemical properties of all of the above. Adequate specification of these can be particularly difficult when (as in modeling the subsurface) the structure of the system is difficult to observe directly. For example, Bond et al. [2007] and Ye et al. [2008b, 2010] demonstrate that significant conceptual uncertainty is inherent in interpretation of data related to subsurface geology. Clearly, much of the discussion related to conceptual representation of process structure also applies here; again detection of the existence

of epistemic uncertainty will be a major issue of concern, along with characterization of its sources. Where appropriate, system nonstationarity in the form of “processes that change the physical structure” will also need to be incorporated [Kumar, 2011].

5. Implications to the Investigation of Model Structure Adequacy

[68] The review presented above illustrates the problems that exist in all stages of the development of models of terrestrial systems. While some of these may yield quickly to improved mathematical approaches and/or improved availability of data, others are not so easily resolved. Although one can simply judge model adequacy from the engineering perspective of minimizing the risk in meeting some desired decision goal, in looking to the future we must give sufficient importance to the scientific perspective of advancing knowledge so that model-based predictions are “right for the right reasons” [Kirchner, 2006]. Only the latter is likely to provide insight into pressing concerns such as system nonstationarity [Milly et al., 2008] and the various types of predictability problems (novel phenomena versus future events [Kumar, 2011]).

[69] Given this perspective, we must be able to establish whether a model structure is adequate to the task of simulating system behaviors under past, current, and potential future conditions for both similar and relatively different locations and/or modeling conditions [Klemes, 1986], a question that largely motivates the “Prediction in Ungaged Basins” initiative [Sivapalan et al., 2003]. In particular, the epistemic question regarding completeness and adequacy of our characterization of the system “How can we know when important physical structures, states and processes, spatial variability, and process interconnections are not included (or not included correctly) in the model?” is of critical importance. To address this, we must ask the following questions regarding evaluation and diagnosis:

[70] 1. How can we know when the behaviors of states and fluxes of interest are simulated with sufficient accuracy and precision?

[71] 2. When deemed inadequate, what are the underlying cause(s) of inadequate model performance?

[72] 3. When deemed adequate, how can we know when models are not overly complex?

[73] In this context, we believe that the following issues require detailed attention.

5.1. Model Evaluation Needs to Be Better Structured

[74] To address question 1 above requires a sufficiently well posed basis for model evaluation. It is clear that any single metric will not generally provide a sufficient basis for characterizing all relevant aspects of model performance and instead some combination of hydrological and statistical metrics will be necessary. In this regard, what seems necessary is a general hierarchical system of metrics that covers the dimensions of space, time, state/process, and application. Here, we must be concerned with which, and how many, metrics are necessary and sufficient to characterize adequacy in model performance. More specifically, we must be concerned with designing such metrics to clarify the causes of model inadequacy [Gupta et al., 2008; Clark et al., 2011b]. While acceptance of a general system may be

difficult to achieve across modeling communities (given their diverse needs), we believe this to be an important goal worth striving toward. As it stands, in using measures such as mean squared error (MSE), cross-correlation coefficient, and other similar statistics across virtually all modeling communities we are already doing something like this, albeit crudely (see discussion of MSE decomposition in *Gupta et al.* [2009]), while to a large extent ignoring the very real problems associated with such poorly informative metrics.

[75] In this context, then, diversity seems important, as it will make it more difficult to “win the game” via calibration, by helping to ensure that models have a complete representation of system behavior. For example, bucket-style SW models with limited representation of spatial variability (both in the vertical and horizontal dimensions) are typically unable to reproduce observations of internal states and fluxes. Clearly, the ability to represent observable states and fluxes should be required by a model used for simulating the natural system.

[76] Finally, in any real-world application of models, we must move beyond blind reliance on statistical metrics to also consider contextual metrics specific to the problem at hand; our vision of a hierarchical system of metrics does not preclude (and even encourages) developments relevant and specific to each area or problem.

5.2. Methods for Diagnosing Epistemic Cause Must Be Developed

[77] Question 2 above points to the basis for improving understanding by revealing where “gaps” in knowledge exist. While numerical issues, deterministic bias in model equations, insufficient detail in spatial variability structure, and underlying random (aleatory) causes are receiving attention, it is to the problem of underlying epistemic cause (things we could in principle, but do not in practice, know) that our collective brainpower, curiosity, and investigation must be directed. While we are certainly not alone in saying this, we point out that we specifically need to understand how an epistemic cause can be detected and characterized by “bringing the model up against the data.” Fundamentally, this must be possible for any kind of structural problem that has an impact on model behavior; if it has an impact then it must be detectable; the challenges are in how to detect it, characterize its impact, attribute its cause, and correct it.

[78] In this regard, methods from the fields of nonlinear time series analysis [*Kantz and Schreiber*, 2004], Bayesian belief networks [*Pearl*, 1986], graph theory [*Gerasoulis and Yang*, 1993], information theory [*Cover and Thomas*, 2006], and manifold learning [*Costa and Hero*, 2004] are likely to prove useful. In particular, we conjecture that insights may specifically be gained by comparing model-inferred and data-inferred characteristic behaviors (signature patterns) to detect when aspects of one are missing from the other [*Gupta et al.*, 2008].

5.3. Methods for Establishing Appropriate Model Complexity Must Be Refined

[79] Question 3 above points to the basis for establishing just how much model structural detail is relevant to the level of performance adequacy desired. *Clement* [2011] poses this as “How should we decide how much is enough?” Here one seeks to balance increasing model complexity against diminishing understandability and increasing uncertainty of

the conceptual physical and process structure during inverse estimation. In this regard, it is common to start from simple conceptual models and tools (mathematical and computational) and build detail. Clearly, perspective (engineering, physical-science, systems-science) has an impact on the decision.

[80] However, more difficult is the issue of whether increasing model complexity has significant impacts on simulations of important but unobserved system behaviors. Information gained from modeling does not always increase with the level of model complexity (or adequacy), especially once a certain level of complexity is attained [see *Clement*, 2011, Figure 4]. This is relevant when aspects of a model established based on prior information (theory and qualitative observations of system properties other than inputs, outputs, and state variables) cannot be quantitatively tested against data; for example, “actual evapotranspiration” is commonly unobserved. Similarly, this is always true of model extrapolations such as projections into the future. In this regard, sophisticated methods of sensitivity analysis [*Saltelli et al.*, 1999; *Rosolem et al.*, 2012] have shown promise in detecting when increasing model complexity provides no tangible benefit. Further development and creativity is needed here.

[81] Additionally, it may be useful to consider model adequacy (or complexity) from the Bayesian perspective in terms of the probability that a model is either inadequate or “more-than-adequate” (overly complex) given available data (quantity and quality), information, knowledge, and project objectives. To achieve this, one must understand what kinds of data/information/knowledge can help to limit the development of overly complex models and how model adequacy can be improved. *Clark et al.* [2011b] note that assessment of model adequacy “must be done thoughtfully, without imposing simplistic solutions on complex problems,” as practical issues of data availability and data quality inevitably constrain our ability to meaningfully discriminate among competing hypothesis of system behavior. The Bayesian perspective naturally balances prior knowledge with data uncertainty, as strong new evidence is required to invalidate strong prior expectations. An advantage of model selection based on the marginal likelihood value is that it helps to prevent overfitting to the data used for model identification and can objectively point to the most appropriate model (of a set) given the information content of the data (*Schoups, G., and J. A. Vrugt*, Bayesian selection of hydrological models using Sequential Monte Carlo sampling, *Water Resour. Res.*, in review, 2012).

[82] Of course, in the context of decision-making, a relevant question is “Who should decide how much is enough [*Clement*, 2011]?” Here, we refrain from taking sides between the engineering, physical-science, and systems-science perspectives but suggest that if all of these perspectives are suitably encapsulated in a hierarchical system of metrics (as expressed above) then a more meaningful (better informed) answer to the model structural adequacy question can be arrived at in any specific situation.

5.4. Methodological Approaches Must Be Multihypothesis in Nature

[83] In learning from data, it seems clear that much can be gained by combining “bottom-up” and “top-down” approaches to progressive model development; see

recommendations of the recent workshop on modeling of complex systems for subsurface fate and transport [U. S. Department of Energy, 2010]. Evolutionary science suggests that any single diagnostic path along a model evolution tree can lead prematurely to a dead end (locally optimal solution), while some previously rejected branch of the tree may ultimately lead to a more satisfactory solution. Top-down investigations are therefore likely to achieve more power by maintaining several parallel lines of investigation beginning with multiple prior conceptualizations [Clark *et al.*, 2011b].

[84] In this regard, we risk stating the obvious fact that when exploiting prior information we must be careful to not become trapped by personal interpretations of such priors. As models become progressively more complex relative to the data (see need for hyperresolution models proposed by Wood *et al.* [2011]), they become progressively harder to falsify, both by becoming more correct (adequate) and by requiring exponentially larger amounts and types of data to be tested against. Unavoidably, therefore, the process of testing will be unable to provide a “complete” test of the model [Clark *et al.*, 2011b]. Here again, a structured hierarchy will help to establish the limits of testing, so that diagnostic interpretations drawn from data are limited to only those aspects of the model that the data can penetrate (inform) and no further.

6. Concluding Remarks

[85] To conclude, we reiterate that a unifying framework for representing model structural development, and a common perspective on the problems of model structure adequacy, can help provide a basis for collective progress toward answering the question “How can we use data to detect, characterize and resolve model structural adequacy?” We present a framework, consisting of five formal steps, as a basis for further discussion, and pose four research issues that require immediate and in-depth investigation. Our hope is that this will promote collaboration and catalyze rapid progress.

[86] From a practical perspective, as we move forward, it will be important for each community to consider whether or not it has become myopic in its perspective. Specifically, are there some steps in the model building process that are being given too much attention, while other areas are being systematically ignored? In SW it is obvious and necessary to question the continued use of “bucket-style” models, given the limited ability of bucket-style models to represent the myriad of hydrologic processes that occur in natural catchments. In THM we may question whether improving grid-scale representations of vertical hydro-meteorological processes in land-surface models is the most critical current research task as long as representations of spatial variability remain so poor. In UZ we may question whether soil hydraulic functions are the most important component of structural uncertainty when issues regarding preferential flow, the importance of hysteresis, and the validity of Richards’ equation over the entire range of modeled soil moisture values, need to be better addressed. In GW we may question whether the spatial representation of hydro-geological structure constitutes the most important source of structural inadequacy, given that Darcy’s Law ignores pore-scale processes and may not be applicable under all the flow conditions to be simulated, and that preferential flow paths

can give rise to systematic model–data mismatch in aspects such as tracer arrival and contaminant transport.

[87] Beyond all of this lies the critical issue of data adequacy, a topic complementary (arguably dual) to that explored in this paper. This issue will involve consideration of both measurement noise and data informativeness (type, amount, spatiotemporal resolution, support, and extent) and of the related problem of designing adequate observational strategies (the observational model) appropriate to the system and problem at hand. As with model adequacy, the issue of data adequacy must inevitably be pursued in an iterative manner.

[88] In conclusion, advancing the science of terrestrial hydrology requires that we adopt a more comprehensive and better-structured assessment of model adequacy. Since each modeling community has tended to place different emphases on different kinds of model adequacy, there exists tremendous scope to learn from each other and to collectively improve the scientific fidelity of our models. Ultimately, a framework that adequately treats and encapsulates both structural uncertainty and forcing data uncertainty must be pursued. In doing so, a systematic characterization of different aspects of model structural adequacy will help by explicitly recognizing the role of each aspect in shaping the overall adequacy of the model.

Appendix A: Model Structural Error Versus Adequacy

[89] The term “model structure error” has become common currency [e.g., Beven, 2005]. However, the term “error” implies the existence of some “true” value from which the difference can (in principle) be measured. Since any model is necessarily a simplified and “analogous” representation of reality, the notion of a “true model” (and its counterpart, “model error”) is in some sense a contradiction in terms.

[90] A more coherent concept is “model structure adequacy” where adequacy has two major attributes, structural and functional. Structural adequacy can be discussed in terms of how well the model emulates key “physical structures” of the system, and functional adequacy in terms of how well it simulates the spatiotemporal “dynamical” behaviors of interest.

[91] The interesting scientific investigation, then, is to what extent enhanced model structural adequacy corresponds to enhanced functional adequacy. Since the goal of modeling is often to maximize functional adequacy with minimal structural representation (the principle of parsimony), we need to be clear that the goal of consistency is also important [Martinez and Gupta, 2011]. This means that enhanced structural adequacy must imply not just the representation of additional system features but also increased consistency in the representation of existing system features. This distinction is especially important for out-of-sample prediction problems (such as in ungauged basins or climate change scenarios), where the value of functional adequacy (determined on existing data sets) can be considerably less than the value of structural adequacy.

[92] Finally, in promoting the term “adequacy” we do not mean the statistical notion of model adequacy used often in evaluating regression models, defined below:

$$p_{\text{adeq}} = 1 - Q(OF_{\min}(\mathbf{p}), N_o - N_p),$$

where $Q(\cdot)$ is the χ^2 cumulative density function of the weighted sum of optimized squared errors objective function $OF_{\min}(\mathbf{p})$ for parameter vector \mathbf{p} with $N_o - N_p$ degrees of freedom, N_o being the number of observations, and N_p being the number of calibration parameters [Hollenbeck and Jensen, 1998a]. This statistical metric is sensitive to outliers, does not differentiate between structural and functional adequacy, and assumes the mismatch between model and data to originate from model deficiencies only.

[93] **Acknowledgments.** We thank the editor and associate editor, and reviewers Erwin Zehe and Keith Beven, and Peter Reichert for highly constructive comments, which contributed significantly to improvement of this manuscript. We acknowledge and thank Shlomo Neuman for his efforts in helping to clarify the (seeming) differences in modeling perspective adopted by the surface water and groundwater modeling communities. The first author received 2011–2012 sabbatical support from (a) Jesus Carrera of the Institute of Environmental Assessment and Water Research (IDAEA), Spanish Council of Scientific Research (CSIC) and the Spanish Ministry of Science and Innovation (MEC), and (b) the Department of Hydrology and Water Resources, University of Arizona. Support for the first author was also provided by Andy Pitman and the Australian Research Council through the Centre of Excellence for Climate System Science (CE110001028). The fifth author received partial support from NSF-EAR grant 0911074 and the DOE Early Career Award DE-SC0008272.

References

- Ambrose, B., K. Beven, and J. Freer (1996), Toward a generalization of the TOPMODEL concepts: Topographic indices of hydrological similarity, *Water Resour. Res.*, *32*(7), 2135–2145, doi:10.1029/95WR03716.
- Anderson, M. P., and W. W. Woessner (2002), *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*, 381 pp., Academic, San Diego, Calif.
- Atkinson, S. E., M. Sivapalan, R. A. Woods, and N. R. Viney (2003), Dominant physical controls on hourly flow predictions and the role of spatial variability: Mahurangi catchment, New Zealand, *Adv. Water Resour.*, *26*, 219–235, doi:10.1016/S0309-1708(02)00183-5.
- Bai, Y., T. Wagener, and P. Reed (2009), A top-down framework for watershed model evaluation and selection under uncertainty, *Environ. Model. Softw.*, *24*, 901–916, doi:10.1016/j.envsoft.2008.12.012.
- Band, L. E., and E. F. Wood (1988), Strategies for large-scale distributed hydrologic simulation, *Appl. Math. Comput.*, *27*(1), 23–37, doi:10.1016/0096-3003(88)90096-3.
- Beven, K. (1989), Changing ideas in hydrology - The case of physically based models, *J. Hydrol.*, *105*, 157–172, doi:10.1016/0022-1694(89)90101-7.
- Beven, K. J. (2001), *Rainfall-Runoff Modeling—The Primer*, 360 pp., John Wiley, Chichester, U.K.
- Beven, K. J. (2002), Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system, *Hydrol. Processes*, *16*, 189–206, doi:10.1002/hyp.343.
- Beven, K. J. (2005), On the concept of model structural error, *Water Sci. Technol.*, *52*, 167–175.
- Beven, K. J. (2006), Searching for the Holy Grail of scientific hydrology: as closure, *Hydrol. Earth Syst. Sci.*, *10*, 609–618, doi:10.5194/hess-10-609-2006.
- Beven, K. J. (2008), On doing better hydrological science, *Hydrol. Processes*, *22*, 3549–3553, doi:10.1002/hyp.7108.
- Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, *249*, 11–29, doi:10.1016/S0022-1694(01)00421-8.
- Beven, K., and E. F. Wood (1983), Catchment geomorphology and the dynamics of runoff contributing areas, *J. Hydrol.*, *65*(1–3), 139–158, doi:10.1016/0022-1694(83)90214-7.
- Blöschl, G., R. B. Grayson, and M. Sivapalan (1995), On the representative elementary area (REA) concept and its utility for distributed rainfall-runoff modelling, *Hydrol. Processes*, *9*, 313–330, doi:10.1002/hyp.3360090307.
- Bond, C. E., A. D. Gibbs, Z. K. Shipton, and S. Jones (2007), What do you think this is? “Conceptual uncertainty” in geoscience interpretation, *GSA Today*, *17*(11), 4, doi:10.1130/GSAT01711A.1.
- Boone, A., and P. J. Wetzel (1996), Issues related to low resolution modeling of soil moisture: Experience with the PLACE model, *Global Planet. Change*, *13*, 161–181, doi:10.1016/0921-8181(95)00044-5.
- Boyle, D. P., H. V. Gupta, and S. Sorooshian (2001), Toward improved streamflow forecasts: Value of semidistributed modeling, *Water Resour. Res.*, *37*(11), 2749–2759, doi:10.1029/2000WR000207.
- Bredehoeft, J. D. (2003), From models to performance assessment—the conceptualization problem, *Ground Water*, *41*, 571–577, doi:10.1111/j.1745-6584.2003.tb02395.x.
- Bredehoeft, J. D. (2005), The conceptualization model problem—surprise, *Hydrogeol. J.*, *13*, 37–46, doi:10.1007/s10040-004-0430-5.
- Bredehoeft, J. D. (2010), Models and model analysis, *Ground Water*, *48*(3), 328, doi:10.1111/j.1745-6584.2009.00631.x.
- Brutsaert, W., and J. L. Nieber (1977), Regionalized drought flow hydrographs from a mature glaciated plateau, *Water Resour. Res.*, *13*, 637–643, doi:10.1029/WR013i003p00637.
- Bulygina, N., and H. V. Gupta (2009), Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, *Water Resour. Res.*, *45*, W00B13, doi:10.1029/2007WR006749.
- Bulygina, N., and H. V. Gupta (2010), How Bayesian data assimilation can be used to estimate the mathematical structure of a model, *Stochastic Environ. Res. Risk Assess.*, *24*, 925–937, doi:10.1007/s00477-010-0387-y.
- Bulygina, N., and H. V. Gupta (2011), Correcting the mathematical structure of a hydrological model via Bayesian data assimilation, *Water Resour. Res.*, *47*, W02503, doi:10.1029/2010WR009240.
- Celia, M. A., E. T. Bouloutas, and R. L. Zarba (1990), A general mass-conservative numerical solution for the unsaturated flow equation, *Water Resour. Res.*, *26*(7), 1483–1496, doi:10.1029/WR026i007p01483.
- Clark, M., and D. Kavetski (2010), Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes, *Water Resour. Res.*, *46*, W10510, doi:10.1029/2009WR008894.
- Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008a), FUSE: A modular framework to diagnose differences between hydrological models, *Water Resour. Res.*, *44*, W00B02, doi:10.1029/2007WR006735.
- Clark, M. P., D. E. Rupp, R. A. Woods, X. Zheng, R. P. Ibbitt, A. G. Slater, J. Schmidt, and M. Uddstrom (2008b), Hydrological data assimilation with the Ensemble Kalman Filter: Use of streamflow data to update the states in a distributed hydrological model, *Adv. Water Resour.*, *31*, 1309–1324, doi:10.1016/j.advwatres.2008.06.005.
- Clark, M., D. Rupp, R. Woods, H. Tromp-van Meerveld, N. Peters, and J. Freer (2009), Consistency between hydrological models and field observations: Linking processes at the hillslope scale to hydrological responses at the watershed scale, *Hydrol. Processes*, *23*, 311–319, doi:10.1002/hyp.7154.
- Clark, M. P., J. Hendrikx, A. G. Slater, B. Anderson, T. R. Kerr, E. Örn Hreinsson, and R. A. Woods (2011a), Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review, *Water Resour. Res.*, *47*, W07539, doi:10.1029/2011WR010745.
- Clark, M. P., D. Kavetski, and F. Fenicia (2011b), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, *47*, W09301, doi:10.1029/2010WR009827.
- Clement, T. P. (2011), Complexities in hindcasting models—When should we say enough is enough?, *Ground Water*, *49*, 620–629, doi:10.1111/j.1745-6584.2010.00765.x.
- Costa, J. A., and A. O. Hero (2004), Geodesic entropic graphs for dimension and entropy estimation in manifold learning, *IEEE Trans. Signal Process.*, *52*(8), 2210–2221, doi:10.1109/TSP.2004.831130.
- Cover, T. M., and J. A. Thomas (2006), *Elements of Information Theory*, 2nd ed., 776 pp., John Wiley, Hoboken, N. J.
- Dagan, G. (1989), *Flow and Transport in Porous Formations*, Springer, Berlin, doi:10.1007/978-3-642-75015-1.
- Dagan, G., and E. Bresler (1979), Solute transport in unsaturated heterogeneous soil at field scale, I. Theory, *Soil Sci. Soc. Am. J.*, *43*, 461–467, doi:10.2136/sssaj1979.03615995004300030008x.
- Dagan, G., and S. P. Neuman (Eds.) (1997), *Subsurface Flow and Transport: A Stochastic Approach*. *Int. Hydrol. Ser.*, Cambridge Univ. Press, Cambridge, U.K., doi:10.1017/CBO9780511600081.
- Desborough, C. E. (1999), Surface energy balance complexity in GCM land surface models, *Clim. Dyn.*, *15*, 389–403, doi:10.1007/s003820050289.
- Destouni, G., and V. Cvetkovic (1989), The effect of heterogeneity on large scale solute transport in the unsaturated zone, *Nord. Hydrol.*, *20*, 43–52.

- Destouni, G., and V. Cvetkovic (1991), Field scale mass arrival of sorptive solute into the groundwater, *Water Resour. Res.*, 27, 1315–1325, doi:10.1029/91WR00182.
- Diks, C. G. H., and J. A. Vrugt (2010), Comparison of point forecast accuracy of model averaging methods in hydrologic applications, *Stochastic Environ. Res. Risk Assess.*, 24(6), 809–820, doi:10.1007/s00477-010-0378-z.
- Dirmeyer, P. A., R. D. Koster, and Z. Guo (2006), Do global models properly represent the feedback between land and atmosphere?, *J. Hydrometeorol.*, 7, 1177–1198, doi:10.1175/JHM532.1.
- Doherty, J., and D. Welter (2010), A short exploration of structural noise, *Water Resour. Res.*, 46, W05525, doi:10.1029/2009WR008377.
- Eder, G., M. Sivapalan, and H. P. Nachtanbel (2003), Modelling of water balances in an Alpine catchment through exploitation of emerging properties over changing time scales, *Hydrol. Processes*, 17, 2125–2149, doi:10.1002/hyp.1325.
- Farmer, D., M. Sivapalan, and C. Jothityangkoon (2003), Climate, soil and vegetation controls upon the variability of water balance in temperature and semiarid landscapes: Downward approach to water balance analysis, *Water Resour. Res.*, 39(2), 1035, doi:10.1029/2001WR000328.
- Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2008), Understanding catchment behavior through stepwise model concept improvement, *Water Resour. Res.*, 44, W01402, doi:10.1029/2006WR005563.
- Feyen, L., M. Kalas, and J. A. Vrugt (2008), Semi-distributed parameter optimization and uncertainty assessment for large-scale streamflow simulation using global optimization, *Hydrol. Sci.*, 53(2), 293–308, doi:10.1623/hysj.53.2.293.
- Gelhar, L. W. (1989), *Stochastic Subsurface Hydrology*, Prentice Hall, Englewood Cliffs, N. J.
- Georgakakos, K. P., D. Seo, H. V. Gupta, J. Schaake, and M. B. Butts (2004), Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, *J. Hydrol.*, 298(1–4), 222–241.
- Gerasoulis, A., and T. Yang (1993), On the granularity and clustering of directed acyclic task graphs, *IEEE Trans. Parallel Distrib. Syst.*, 4(6), 686–701, doi:10.1109/71.242154.
- Ginzburg, L. R., and C. X. J. Jensen (2004), Rules of thumb for judging ecological theories, *Trends Ecol. Evol.*, 19, 121–126, doi:10.1016/j.tree.2003.11.004.
- Granger, C. W. J., and R. Ramanathan (1984), Improved methods of combining forecasts, *J. Forecast.*, 3, 197–204, doi:10.1002/for.3980030207.
- Grayson, R. B., I. D. Moore, and T. A. McMahon (1992), Physically based hydrologic modeling. 2. Is the concept realistic?, *Water Resour. Res.*, 28, 2659–2666, doi:10.1029/92WR01259.
- Gupta, H. V., T. Wagener, and Y. Q. Liu (2008), Reconciling theory with observations: Towards a diagnostic approach to model evaluation, *Hydrol. Processes*, 22(18), 3802–3813, doi:10.1002/hyp.6989.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez-Baquero (2009), Decomposition of the mean squared error & NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377, 80–91, doi:10.1016/j.jhydrol.2009.08.003.
- Hantush, M. S. (1967), Growth and decay of groundwater-mounds in response to uniform percolation, *Water Resour. Res.*, 3(1), 227–234, doi:10.1029/WR003i001p00227.
- Harbaugh, A. W. (2005), MODFLOW-2005, the U.S. Geological Survey modular ground-water model—The Ground-Water Flow Process, *Tech. and Methods 6–A16*, U. S. Geol. Surv., Reston, Va.
- Harman, C. J., M. Sivapalan, and P. Kumar (2009), Power law catchment-scale recessions arising from heterogeneous linear 638 small-scale dynamics, *Water Resour. Res.*, 45, W09404, doi:10.1029/2008WR007392.
- Harter, T., and T.-C. Yeh (1996a), Stochastic analysis of solute transport in heterogeneous, variably saturated soils, *Water Resour. Res.*, 32, 1585–1595, doi:10.1029/96WR00502.
- Harter, T., and T.-C. Yeh (1996b), Conditional stochastic analysis of solute transport in heterogeneous, variably saturated soils, *Water Resour. Res.*, 32, 1597–1609, doi:10.1029/96WR00503.
- Harter, T., and D. Zhang (1999), Water flow and solute spreading in heterogeneous soils with spatially variable water content, *Water Resour. Res.*, 35, 415–426, doi:10.1029/1998WR900027.
- Hill, M. C. (2006), The practical use of simplicity in developing ground water models, *Ground Water*, 44(6), 775–781, doi:10.1111/j.1745-6584.2006.00227.x.
- Hogue, T. S., L. A. Bastidas, H. V. Gupta, and S. Sorooshian (2006), Evaluating model performance and parameter behavior for varying levels of land surface model complexity, *Water Resour. Res.*, 42, W08430, doi:10.1029/2005WR004440.
- Hollenbeck, K. J., and K. H. Jensen (1998a), Maximum-likelihood estimation of unsaturated hydraulic parameters, *J. Hydrol.*, 210, 192–205, doi:10.1016/S0022-1694(98)00185-1.
- Hollenbeck, K. J., and K. H. Jensen (1998b), Experimental evidence of randomness and nonuniqueness in unsaturated outflow experiments designed for hydraulic parameter estimation, *Water Resour. Res.*, 34, 595–602, doi:10.1029/97WR03609.
- Hunt, R. J., J. Doherty, and M. J. Tonkin (2007), Are models too simple? Argument for increased parameterization, *Ground Water*, 45(3), 254–262, doi:10.1111/j.1745-6584.2007.00316.x.
- Iliev, L. (1984), *Theory of Modelling*, 75 pp., Bulgarian Acad. of Sci., Sofia.
- Jothityangkoon, C., M. Sivapalan, and D. Farmer (2001), Process controls on water balance variability in a large semi-arid catchment: Downward approach to hydrological model development, *J. Hydrol.*, 254, 174–198, doi:10.1016/S0022-1694(01)00496-6.
- Jury, W. A. (1982), Simulation of solute transport using a transfer function model, *Water Resour. Res.*, 18, 363–368, doi:10.1029/WR018i002p00363.
- Kantz, H., and T. Schreiber (2004), *Nonlinear Time Series Analysis*, Cambridge Univ. Press, New York.
- Kavetski, D., and M. Clark (2010), Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, *Water Resour. Res.*, 46, W10511, doi:10.1029/2009WR008896.
- Kavetski, D., and G. Kuczera (2007), Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration, *Water Resour. Res.*, 43, W03411, doi:10.1029/2006WR005195.
- Kavetski, D., G. Kuczera, and D. W. Franks (2003), Semidistributed hydrological modeling: A “saturation path” perspective on TOPMODEL and VIC, *Water Resour. Res.*, 39(9), 1246, doi:10.1029/2003WR002122.
- Kavetski, D., G. Kuczera, and D. W. Franks (2006a), Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts, *J. Hydrol.*, 320, 173–186, doi:10.1016/j.jhydrol.2005.07.012.
- Kavetski, D., G. Kuczera, and D. W. Franks (2006b), Calibration of conceptual hydrological models revisited: 2. Improving optimisation and analysis, *J. Hydrol.*, 320, 187–201, doi:10.1016/j.jhydrol.2005.07.013.
- Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42, W03S04, doi:10.1029/2005WR004362.
- Kirchner, J. W. (2009), Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resour. Res.*, 45, W02429, doi:10.1029/2008WR006912.
- Kirkby, M. J. (1975), Hydrograph modelling strategies, in *Processes in Physical and Human Geography*, edited by R. Peel, M. Chisholm, and P. Haggett, pp. 69–90, Heinemann, Portsmouth, N. H.
- Klaus, J., and E. Zehe (2010), Modeling rapid flow response of a tile-drained field site using a 2D physically based model: Assessment of ‘equifinal’ model setups, *Hydrol. Processes*, 24(12), 1595–1609, doi:10.1002/hyp.7687.
- Klemes, V. (1986), Operational testing of hydrological simulation models, *Hydrol. Sci. J.*, 31(1), 13–24, doi:10.1080/02626668609491024.
- Kling, H., and H. V. Gupta (2009), On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability, *J. Hydrol.*, 373, 337–351, doi:10.1016/j.jhydrol.2009.04.031.
- Kollet, S. J., R. M. Maxwell, C. S. Woodward, S. Smith, J. Vanderborght, H. Vereecken, and C. Simmer (2010), Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources, *Water Resour. Res.*, 46, W04201, doi:10.1029/2009WR008730.
- Koltermann, C. E., and S. M. Gorelick (1996), Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches, *Water Resour. Res.*, 32(9), 2617–2658, doi:10.1029/96WR00025.
- Konikow, L. F. (2011), The secret to successful solute-transport modeling, *Ground Water*, 49(2), 144–159, doi:10.1111/j.1745-6584.2010.00764.x.
- Konikow, L. F., and J. D. Bredehoeft (1992), Groundwater models cannot be validated, *Adv. Water Resour.*, 15(1), 75–83, doi:10.1016/0309-1708(92)90033-X.
- Koren, V. I., M. Smith, D. Wang, and Z. Zhang (2000), Use of soil property data in the derivation of conceptual rainfall runoff model parameters, paper presented at 15th Conference on Hydrology, Am. Meteorol. Soc., Long Beach, Calif.
- Krinner, G., N. Viogy, N. de Noblet-Ducoudré, J. Ogée, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice (2005), A dynamic

- global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochem. Cycles*, *19*, GB1015, doi:10.1029/2003GB002199.
- Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, *331*(1–2), 161–177, doi:10.1016/j.jhydrol.2006.05.010.
- Kumar, P. (2011), Typology of hydrologic predictability, *Water Resour. Res.*, *47*, W00H05, doi:10.1029/2010WR009769.
- Lee, D. H., and L. M. Abriola (1999), Use of the Richards equation in land surface parameterizations, *J. Geophys. Res.*, *104*(D22), 27,519–27,526, doi:10.1029/1999JD900951.
- Lenhard, R. J., M. Oostrom, and M. D. White (1995), Modeling fluid flow and transport in variably saturated porous media with the STOMP simulation. 2. Verification and validation exercises, *Adv. Water Resour.*, *18*(6), 365–373, doi:10.1016/0309-1708(95)00019-F.
- Leplastrier, M., A. J. Pitman, H. V. Gupta, and Y. Xia (2002), Exploring the relationship between complexity and performance in a land surface model using the multicriteria method, *J. Geophys. Res.*, *107*(D20), 4443, doi:10.1029/2001JD000931.
- Liang, X., Z. H. Xie, and M. Y. Huang (2003), A new parameterization for surface and groundwater interactions and its impact on water budgets with the variable infiltration capacity (VIC) land surface model, *J. Geophys. Res.*, *108*(D16), 8613, doi:10.1029/2002JD003090.
- Lin, H. C., D. R. Richards, G. T. Yeh, J. R. Cheng, H. P. Chang, and N. L. Jones (1996), FEMWATER: A three-dimensional finite element computer model for simulating density dependent flow and transport, technical report, 129 pp., U.S. Army Eng. Waterways Exp. Station, Vicksburg, Miss.
- Martinez, G. F., and H. V. Gupta (2011), Hydrologic consistency as a basis for assessing complexity of water balance models for the continental United States, *Water Resour. Res.*, *47*, W12540, doi:10.1029/2011WR011229.
- Matott, L. S., and A. J. Rabideau (2008), Calibration of complex subsurface reaction models using a surrogate-model approach, *Adv. Water Resour.*, *31*(12), 1697–1707, doi:10.1016/j.advwatres.2008.08.006.
- Maxwell, R. M., J. K. Lundquist, J. D. Mirocha, S. G. Smith, C. S. Woodward, and A. F. B. Tompson (2011), Development of a coupled groundwater-atmosphere model, *Mon. Weather Rev.*, *139*(1), 96–116, doi:10.1175/2010MWR3392.1.
- McIntyre, N., P. Young, B. Orellana, M. Marshall, B. Reynolds, and H. Wheeler (2011), Identification of nonlinearity in rainfall-flow response using data-based mechanistic modeling, *Water Resour. Res.*, *47*, W03515, doi:10.1029/2010WR009851.
- McMillan, H., and M. Clark (2009), Rainfall-runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme, *Water Resour. Res.*, *45*, W04418, doi:10.1029/2008WR007288.
- Medlyn, B. E., A. P. Robinson, R. Clement, and R. E. McMurtrie (2005), On the validation of models of forest CO₂ exchange using eddy covariance data: Some perils and pitfalls, *Tree Physiol.*, *25*(7), 839–857, doi:10.1093/treephys/25.7.839.
- Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008), Stationarity is dead: Whither water management?, *Science*, *319*, 573, doi:10.1126/science.1151915.
- Mishra, P. K., and S. P. Neuman (2010), Improved forward and inverse analyses of saturated-unsaturated flow toward a well in a compressible unconfined aquifer, *Water Resour. Res.*, *46*, W07508, doi:10.1029/2009WR008899.
- Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. A. Houser (2005a), Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, *28*, 135–147, doi:10.1016/j.advwatres.2004.09.002.
- Moradkhani, H., K. L. Hsu, H. V. Gupta, and S. Sorooshian (2005b), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using particle filter, *Water Resour. Res.*, *41*, W05012, doi:10.1029/2004WR003604.
- Mott, R., and M. Lehning (2010), Meteorological modeling of very high-resolution wind fields and snow deposition for mountains, *J. Hydrometeorol.*, *11*, 934–949, doi:10.1175/2010JHM1216.1.
- National Research Council (2001), *Conceptual Models of Flow and Transport in the Vadose Zone*, Natl. Acad. Press, Washington, D. C.
- Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environ. Res. Risk Assess.*, *17*, 291–305, doi:10.1007/s00477-003-0151-7.
- Neuman, S. P., and P. J. Wierenga (2003), A comprehensive strategy of hydrogeologic modeling and uncertainty analysis for nuclear facilities and sites, *Rep. NUREG/CR-6805*, U.S. Nucl. Regul. Comm, Washington, D. C.
- Neuman, S. P., A. Guadagnini, and M. Riva (2004), Type-curve estimation of statistical heterogeneity, *Water Resour. Res.*, *40*, W04201, doi:10.1029/2003WR002405.
- Niu, G. Y., Z. L. Yang, R. E. Dickinson, L. E. Gulden, and H. Su (2007), Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, *J. Geophys. Res.*, *112*(D7), D07103, doi:10.1029/2006JD007522.
- Niu, G. Y., et al. (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, *J. Geophys. Res.*, *116*, D12109, doi:10.1029/2010JD015139.
- Oleson, K. W., X. Zeng, and X. Zeng (2010), Technical description of version 4.0 of the Community Land Model (CLM), *NCAR Tech. Note NCAR/TN-478+STR*, 257 pp., Natl. Cent. for Atmos. Res., Boulder, Colo.
- Oreskes, N., K. Shraderfrechette, and K. Belitz (1994), Verification, validation, and confirmation of numerical models in the Earth sciences, *Science*, *263*(5147), 641–646, doi:10.1126/science.263.5147.641.
- Pan, F., M. Ye, J. Zhu, Y. S. Wu, B. Hu, and Z. Yu (2009), Effect of water retention parameters on predictive uncertainty of unsaturated flow and contaminant transport, *Vadose Zone J.*, *8*, 158–166, doi:10.2136/vzj2008.0092.
- Pan, L., J. Jin, N. Miller, Y. Wu, and G. Bodvarsson (2008), Modeling hydraulic responses to meteorological forcing: From canopy to aquifer, *Vadose Zone J.*, *7*(1), 325–331, doi:10.2136/vzj2006.0106.
- Pearl, J. (1986), Fusion, propagation, and structuring in belief networks, *Artif. Intell.*, *29*(3), 241–288, doi:10.1016/0004-3702(86)90072-X.
- Peters, D. P. C., B. T. Bestelmeyer, and M. G. Turner (2007), Cross-scale interactions and changing pattern-process relationships: Consequences for system dynamics, *Ecosystems (N. Y.)*, *10*(5), 790–796, doi:10.1007/s10021-007-9055-6.
- Pinder, G. F., and P. S. Huyakorn (1977), *Computational Methods in Subsurface Flow*, Academic, San Diego, Calif.
- Pitman, A. J., Y. Xia, M. Leplastrier, and A. Henderson-Sellers (2003), The Chameleon Surface Model: Description and use with the PILPS phase 2(e) forcing data, *Global Planet. Change*, *38*, 121–135, doi:10.1016/S0921-8181(03)00015-8.
- Poeter, E. P., and D. A. Anderson (2005), Multimodel ranking and inference in ground water modeling, *Ground Water*, *43*(4), 597–605, doi:10.1111/j.1745-6584.2005.0061.x.
- Poeter, E. P., and M. C. Hill (2007), *MMA: A Computer Code for Multi-Model Analysis, Techniques and Methods TM6-E3*, U.S. Geol. Surv., Boulder, Colo.
- Pokhrel, P., and H. V. Gupta (2010), On the use of spatial regularization strategies to improve calibration of distributed watershed models, *Water Resour. Res.*, *46*, W01505, doi:10.1029/2009WR008066.
- Pokhrel, P., and H. V. Gupta (2011), On the ability to infer spatial catchment variability using streamflow hydrographs, *Water Resour. Res.*, *47*, W08534, doi:10.1029/2010WR009873.
- Pokhrel, P., H. V. Gupta, and T. Wagener (2008), A spatial regularization approach to parameter estimation for a distributed watershed model, *Water Resour. Res.*, *44*, W12419, doi:10.1029/2007WR006615.
- Pokhrel, P., K. Yilmaz, and H. V. Gupta (2009), Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures, *J. Hydrol.*, *418–419*, 49–60, doi:10.1016/j.jhydrol.2008.12.004.
- Refsgaard, J. C., J. P. van der Sluijs, J. Brown, and P. van der Keur (2006), A framework for dealing with uncertainty due to model structure error, *Adv. Water Resour.*, *29*, 1586–1597, doi:10.1016/j.advwatres.2005.11.013.
- Reggiani, P., M. Sivapalan, and S. M. Hassanizadeh (1998), A unifying framework for watershed thermodynamics: Balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics, *Adv. Water Resour.*, *22*(4), 367–398, doi:10.1016/S0309-1708(98)00012-8.
- Reggiani, P., S. M. Hassanizadeh, M. Sivapalan, and W. G. Gray (1999), A unifying framework for watershed thermodynamics: Constitutive relationships, *Adv. Water Resour.*, *23*(1), 15–39, doi:10.1016/S0309-1708(99)00005-6.
- Reichert, P., and J. Mieleitner (2009), Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, *Water Resour. Res.*, *45*, W10402, doi:10.1029/2009WR007814.

- Roman, P., A. A. Brooks, R. G. Newton, and L. De La Torre (1998), The constructs of physics and the role of math—Revisited, *Phys. Today*, 51(1), 91, doi:10.1063/1.882124.
- Rosolem, R., H. V. Gupta, W. J. Shuttleworth, L. G. G. de Goncalves, X. Zeng, and N. Restrepo-Coupe (2012), A fully multiple-criteria implementation of the Sobol method for parameter sensitivity analysis, *J. Geophys. Res.*, 117, D07103, doi:10.1029/2011JD016355.
- Roubinet, D., H. H. Liu, and J. R. de Dreuzy (2010), A new particle tracking approach to simulating transport in heterogeneous fractured porous media, *Water Resour. Res.*, 46, W11507, doi:10.1029/2010WR009371.
- Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, Oxford Univ. Press, New York.
- Rupp, D. E., and J. S. Selker (2006), On the use of the Boussinesq equation for interpreting recession hydrographs from sloping aquifers, *Water Resour. Res.*, 42, W12421, doi:10.1029/2006WR005080.
- Russo, D., and A. Fiori (2009), Stochastic analysis of transport in a combined heterogeneous vadose zone–groundwater flow system, *Water Resour. Res.*, 45, W03426, doi:10.1029/2008WR007157.
- Saltelli, A., S. Tarantola, and K. P. S. Chan (1999), A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics*, 41(1), 39–56.
- Samaniego, L., R. Kumar, and S. Attinger (2010), Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resour. Res.*, 46, W05523, doi:10.1029/2008WR007327.
- Schoups, G., J. A. Vrugt, F. Fenicia, and N. C. van de Giesen (2010), Corruption of accuracy and efficiency of Markov chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models, *Water Resour. Res.*, 46, W10530, doi:10.1029/2009WR008648.
- Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, 38(11), 1241, doi:10.1029/2001WR000978.
- Sen, O. L., L. A. Bastidas, W. J. Shuttleworth, Z. L. Yang, H. V. Gupta, and S. Sorooshian (2001), Impact of field-calibrated vegetation parameters on GCM climate simulations, *Q. J. R. Meteorol. Soc.*, 127(574), 1199–1224.
- Serrano, S. E., and T. E. Unny (1987a), Semigroup solutions of the unsteady groundwater flow equation with stochastic parameters, *J. Stochastic Hydrol. Hydraul.*, 1, 281–296, doi:10.1007/BF01543100.
- Serrano, S. E., and T. E. Unny (1987b), Predicting groundwater flow in a phreatic aquifer, *J. Hydrol.*, 95, 241–268, doi:10.1016/0022-1694(87)90004-7.
- Serrano, S. E., and T. E. Unny (1988), General solution of random advective-dispersive transport equation in porous media. Part II: Stochasticity in the sources and the boundaries, *J. Stochastic Hydrol. Hydraul.*, 2(3), 20–34.
- Serrano, S. E., and T. E. Unny (1990), Random evolution equations in hydrology, *Appl. Math. Comput.*, 38(3), 201–226, doi:10.1016/0096-3003(90)90023-V.
- Šimůnek, J., and M. T. van Genuchten (2008), Modeling nonequilibrium flow and transport with HYDRUS, *Vadose Zone J.*, 7(2), 782–797, doi:10.2136/VZJ2007.0074.
- Sivapalan, M. (2005), Pattern, process and function: Elements of a new unified hydrologic theory at the catchment scale, in *Encyclopaedia of Hydrologic Sciences*, vol. 13, edited by M. G. Anderson, pp. 193–219, John Wiley, Hoboken, N. J.
- Sivapalan, M., et al. (2003), IAHS decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, 48(6), 857–880, doi:10.1623/hysj.48.6.857.51421.
- Son, K., and M. Sivapalan (2007), Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, *Water Resour. Res.*, 43, W01415, doi:10.1029/2006WR005032.
- Troch, P. A., G. F. Martinez, V. R. N. Pauwels, M. Durcik, M. Sivapalan, C. Harman, P. D. Brooks, H. V. Gupta, and T. Huxman (2009), Climate and vegetation water-use efficiency at catchment scales, *Hydrol. Processes*, 23(16), 2409–2414, doi:10.1002/hyp.7358.
- Tsai, F. T. C., and X. Li (2008), Multiple parameterization for hydraulic conductivity identification, *Ground Water*, 46(6), 851–864.
- Tsai, F. T. C., and W. W. G. Yeh (2004), Characterization and identification of aquifer heterogeneity with generalized parameterization and Bayesian estimation, *Water Resour. Res.*, 40, W10102, doi:10.1029/2003WR002893.
- Unny, T. E. (1989), Stochastic partial differential equations in groundwater hydrology Part I: Theory, *J. Stochastic Hydrol. Hydraul.*, 3, 135–153, doi:10.1007/BF01544077.
- U.S. Department of Energy (2010), Complex systems science for subsurface fate and transport: Report from the August 2009 workshop, *DOE/SC-0123*, Washington, D. C.
- Vesselinov, V. V., S. P. Neuman, and W. A. Illman (2001a), Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff: 1. Methodology and borehole effects, *Water Resour. Res.*, 37(12), 3001–3017, doi:10.1029/2000WR000133.
- Vesselinov, V. V., S. P. Neuman, and W. A. Illman (2001b), Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff: 2. Equivalent parameters, high-resolution stochastic imaging and scale effects, *Water Resour. Res.*, 37(12), 3019–3041, doi:10.1029/2000WR000135.
- Vrugt, J. A., and B. A. Robinson (2007), Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging, *Water Resour. Res.*, 43, W01411, doi:10.1029/2005WR004838.
- Vrugt, J. A., M. T. van Wijk, J. W. Hopmans, and J. Simunek (2001), One, two, and three-dimensional root water uptake functions for transient modeling, *Water Resour. Res.*, 37(10), 2457–2470, doi:10.1029/2000WR000027.
- Vrugt, J. A., G. Schoups, J. W. Hopmans, C. Young, W. W. Wallender, T. Harter, and W. Bouten (2004), Inverse modeling of large-scale spatially distributed vadose zone properties using global optimization, *Water Resour. Res.*, 40, W06503, doi:10.1029/2003WR002706.
- Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005a), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, 41, W01017, doi:10.1029/2004WR003059.
- Vrugt, J. A., B. A. Robinson, and V. V. Vesselinov (2005b), Improved inverse modeling of flow and transport in subsurface media: Combined parameter and state estimation, *Geophys. Res. Lett.*, 32, L18408, doi:10.1029/2005GL023940.
- Vrugt, J. A., H. V. Gupta, B. O. Nuallain, and W. Bouten (2006), Real time data assimilation for operational ensemble streamflow forecasting, *J. Hydrometeorol.*, 7(3), 548–565, doi:10.1175/JHM504.1.
- Wang, H. F., and M. P. Anderson (1982), *Introduction to Groundwater Modeling: Finite Difference and Finite Element Methods*, 237 pp., W.H. Freeman, San Francisco, Calif.
- Wang, Y., E. Kowalczyk, R. Leuning, G. Abramowitz, M. R. Raupach, B. Pak, E. van Gorsel, and A. Luhar (2011), Diagnosing errors in a land surface model (CABLE) in the time and frequency domains, *J. Geophys. Res.*, 116, G01034, doi:10.1029/2010JG001385.
- Winter, C. L., and D. Nychka (2009), Forecasting skill of model averages, *Stochastic Environ. Res. Risk Assess.*, 24, 633–638, doi:10.1007/s00477-009-0350-y.
- Wöhling, T., and J. A. Vrugt (2011), Multi-response multi-layer vadose zone model calibration using Markov chain Monte Carlo simulation and field water retention data, *Water Resour. Res.*, 47, W04510, doi:10.1029/2010WR009265.
- Wood, E. F., et al. (2011), Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resour. Res.*, 47, W05301, doi:10.1029/2010WR010090.
- Wood, W. W. (2005), It's the heterogeneity!, *Ground Water*, 38(1), 1, doi:10.1111/j.1745-6584.2000.tb00187.x.
- Wu, Y. S., L. Pan, and K. Pruess (2004), A physically based approach for modeling multiphase fracture-matrix interaction in fractured porous media, *Adv. Water Resour.*, 27, 875–887, doi:10.1016/j.advwatres.2004.07.002.
- Xia, Y., A. J. Pitman, H. V. Gupta, M. Leplastrier, A. Henderson-Sellers, and L. A. Bastidas (2002), Calibrating a land surface model of varying complexity using multi-criteria methods and the Cabauw data set, *J. Hydrometeorol.*, 3, 181–194, doi:10.1175/1525-7541(2002)003<0181:CALSMO>2.0.CO;2.
- Ye, M., and R. Khaleel (2008), A Markov chain model for characterizing medium heterogeneity and sediment layering structure, *Water Resour. Res.*, 44, W09427, doi:10.1029/2008WR006924.
- Ye, M., S. P. Neuman, and P. D. Meyer (2004), Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resour. Res.*, 40, W05113, doi:10.1029/2003WR002557.
- Ye, M., R. Khaleel, M. G. Schaap, and J. Zhu (2007), Simulation of field injection experiments in heterogeneous unsaturated media using cokriging and artificial neural network, *Water Resour. Res.*, 43, W07413, doi:10.1029/2006WR005030.
- Ye, M., P. D. Meyer, and S. P. Neuman (2008a), On model selection criteria in multimodel analysis, *Water Resour. Res.*, 44, W03428, doi:10.1029/2008WR006803.

- Ye, M., K. F. Pohlmann, and J. B. Chapman (2008b), Expert elicitation of recharge model probabilities for the Death Valley regional flow system, *J. Hydrol.*, *354*, 102–115, doi:10.1016/j.jhydrol.2008.03.001.
- Ye, M., K. F. Pohlmann, J. B. Chapman, G. M. Pohl, and D. M. Reeves (2010), A model-averaging method for assessing groundwater conceptual model uncertainty, *Ground Water*, *48*, 716–728, doi:10.1111/j.1745-6584.2009.00633.x.
- Yeh, G. T. (1999), *Computational Subsurface Hydrology Fluid Flows*, Kluwer Acad., Norwell, Mass.
- Yeh, T. C. J., L. W. Gelhar, and A. L. Gutjahr (1985a), Stochastic analysis of unsaturated flow in heterogeneous soils: 1. Statistically isotropic media, *Water Resour. Res.*, *21*, 447–456, doi:10.1029/WR021i004p00447.
- Yeh, T. C. J., L. W. Gelhar, and A. L. Gutjahr (1985b), Stochastic analysis of unsaturated flow in heterogeneous soils: 2. Statistically anisotropic media, *Water Resour. Res.*, *21*, 457–464, doi:10.1029/WR021i004p00457.
- Yeh, T. C. J., L. W. Gelhar, and A. L. Gutjahr (1985c), Stochastic analysis of unsaturated flow in heterogeneous soils: 3. Observations and applications, *Water Resour. Res.*, *21*, 465–471, doi:10.1029/WR021i004p00465.
- Yeh, T. C. J., S. Liu, R. J. Glass, K. Baker, J. R. Brainard, D. L. Alumbaugh, and D. LaBrecque (2002), A geostatistically based inverse model for electrical resistivity surveys and its applications to vadose zone hydrology, *Water Resour. Res.*, *38*(12), 1278, doi:10.1029/2001WR001204.
- Young, P. (1998), Data-based mechanistic modelling of environmental, ecological, economic, and engineering systems, *Environ. Model. Softw.*, *13*, 105–122, doi:10.1016/S1364-8152(98)00011-5.
- Young, P. (2003), Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale, *Hydrol. Processes*, *17*, 2195–2217, doi:10.1002/hyp.1328.
- Young, P. C., and M. Ratto (2008), A unified approach to environmental systems modelling, *Stochastic Environ. Res. Risk Assess.*, *24*, 665–674, doi:10.1007/s00477-008-0271-1.
- Young, P., S. Parkinson, and M. Lees (1996), Simplicity out of complexity in environmental modelling: Occam's razor revisited, *J. Appl. Stat.*, *23*, 165–210, doi:10.1080/02664769624206.
- Zhang, D. (2002), *Stochastic Methods for Flow in Porous Media: Coping With Uncertainties*, Academic, San Diego, Calif.
- Zheng, C., and G. D. Bennett (2002), *Applied Contaminant Transport Modeling*, 2nd ed., 621 pp., John Wiley, New York.
- Zyvoloski, G. A., and V. V. Vesselinov (2006), An investigation of numerical grid effects in automated calibration, *Ground Water*, *44*(6), 814–825, doi:10.1111/j.1745-6584.2006.00203.x.