Improving the theoretical underpinnings of process-based hydrologic models

Martyn P. Clark, Bettina Schaefli, Stanislaus J. Schymanski, Luis Samaniego, Charlie H. Luce, Bethanna M. Jackson, Jim E. Freer, Jeffrey R. Arnold, R. Dan Moore, Erkan Istanbulluoglu, Serena Ceola, and Ciaran Harman

Paper submitted as a commentary to *Water Resources Research*, 28 July 2015; revised 9 November 2015

Abstract

In this Commentary we argue that it is possible to improve the physical realism of hydrologic models by making better use of existing hydrologic theory. We address the following questions: (1)what are some key elements of current hydrologic theory; (2)how can those elements best be incorporated where they may be missing in current models; and (3)how can we evaluate competing hydrologic theories across scales and locations? We propose that hydrologic science would benefit from a model-based community synthesis effort to reframe, integrate and evaluate different explanations of hydrologic behavior, and provide a controlled avenue to find where understanding falls short.

<u>Key points</u>:

- We seek to increase the realism of hydrologic models through better use of existing theory.
- We seek to improve the use of models as learning tools to evaluate different process explanations.
- We define critical issues to address that will help narrow the gap between theory and models.

1. Motivation

The discipline of hydrology continues to be an exciting field, with ongoing advances in field observational techniques, availability of global data products and increasing computational power. Now, perhaps more than ever before, we are rising to the challenge of building models of everywhere [Beven 2007]. Key efforts include building continental-domain hydrologic models for water security assessments [Schewe et al. 2014; Mizukami et al. 2015] and improving the representation of hydrologic processes in Earth System Models [Clark et al. 2015a]. These efforts require moving beyond the traditional tactics used in hydrology such as detailed analysis and modeling of individual catchments. Instead, hydrologic synthesis is needed, across space and across many elements of hydrologic theory, in order to improve the physical realism and general applicability of hydrologic models, i.e., to improve hydrologic process representations across a large range of catchments [Gupta et al. 2014]. Advances in modern hydrologic modeling efforts are possible through progress on the following fundamental research challenges: identifying consistently observed behaviors across research watersheds; formulating the laws that govern macroscale hydrologic behavior; and unifying process explanations across watersheds in order to develop theory of hydrology at the catchment scale [Dooge 1986; Sivapalan 2005; McDonnell et al. 2007].

The needs of the hydrologic modeling community, articulated in this way, are admittedly sizeable and potentially insurmountable. Is it even possible to generalize about hydrologic behaviors given the unique character of individual basins [Beven 2000]? Do we now, and/or will we always, lack the necessary information on climate, topography, vegetation, soils, and subsurface structure required to develop powerful and exceptionless explanations? Put differently, are the problems of underdetermination so pronounced that we cannot move beyond explanatory pluralism [Kleinhans et al. 2005; Beven 2006a, 2006b]? These difficulties are shared across multiple disciplines, and are described very well in Nancy Cartwright's book "How the laws of physics lie": "*Covering-law theorists tend to think that nature is well regulated; in the extreme, that there is a law to cover every case. I do not. I imagine that natural objects are much like people in societies. Their behavior is constrained by some specific laws and by a handful of general principles, but it is not determined in detail, even statistically. What happens on most occasions is dictated by no law at all [...] God may have written just a few laws and grown tired. We do not know whether we are in a tidy universe or an untidy one*." [Cartwright 1983, p. 49].

The purpose of this paper is to bridge these two perspectives. On one hand we recognize that developing a unified hydrologic theory will be incredibly useful, and, on the other hand, we also recognize that the "messy" nature of the Universe makes theory development incredibly difficult. Nevertheless, we accept that elements of hydrologic theory exist now, and it is critical to reconcile hydrologic models with existing and emerging theory. While acknowledging uncertainty, underdetermination, and the difficulty to generalize, we contend that the hydrologic community has made tremendous advances over the past few decades in our capability to explain and predict individual processes, process interactions, patterns and scaling behavior. However process explanations (theories) are currently scattered across research groups and not yet widely incorporated in hydrologic models. Consequently, we propose that hydrologic science would substantially benefit from a model-based synthesis effort to systematically formulate, organize, formalize, encode, and evaluate hydrologic theories, i.e., to use models as a means to summarize, integrate and test many different, sometimes competing explanations of hydrologic behavior. The idea is that such models would be used to synthesize current process understanding and provide a controlled avenue to find where that understanding falls short.

The central thesis of this paper is as follows: It is possible to increase the physical realism and general applicability of hydrologic models by making better use of the elements of hydrologic theory that exist now. To this end, we explore the following three questions:

1) What are the key elements of current hydrologic theory? This requires research to reconcile consistently observed behavior in research watersheds with explanations of hydrologic processes, process interactions and scaling behavior, and includes algorithmic implementations of explanations as encoded in models.

2) How should we incorporate the elements of existing hydrologic theory in models? This requires developing multiple parameterizations and numerical approximations of process explanations of a given theory, within a common modeling framework, implemented as falsifiable (testable) hypotheses. In this context a community-based hydrologic modeling endeavor is needed, one like those implemented successfully in the atmospheric science and land-atmosphere interactions communities [Lawrence et al. 2011; Hurrell et al. 2013]

3) How should we evaluate competing hydrologic theories across scales and locations (while explicitly recognizing uncertainty)? To address this question we argue that research is needed to design and implement a suite of diagnostic metrics to evaluate model hypotheses (using incomplete and inexact information), and to test the utility of models for prediction/extrapolation.

In addressing these questions we follow a Popperian approach for discovery and learning via formulation of testable (falsifiable) hypotheses [Popper 1959]. As highlighted in Figure 1, we adopt the modeler's perspective with a focus on the iterative refinement of models and theory via systematic testing of multiple hypotheses. Our main contribution is to define a key set of research challenges, and methods for addressing them, in order to improve the link between theory, models and data.

A key facet of our approach is that we seek to improve the theoretical underpinnings of process-based hydrologic models, regardless of their complexity and intended purpose. We consider models of varying process complexity (i.e., models with a different number of processes explicitly represented), as well as models of different spatial complexity (i.e., spatially explicit models with different degrees of spatial discretization and connectivity, and spatially implicit lumped hydrologic models). Our primary considerations include both the underlying theories used to explain hydrologic behavior and how process explanations are represented in models. We do accept that different models answer different questions; yet we argue that all process-based models should be as deeply rooted as feasible in the available hydrologic theory. The purpose of the model defines the simplifications of the theory that the modeler is willing to tolerate. She may, for example, use a rougher estimate for a quantity of interest. By considering a broad range of process-based models, our desired outcome is to encourage more widespread adoption and scrutiny of hydrologic theory as part of model development.

The remainder of this paper is organized as follows. In Section 2 we define what we mean by hypotheses, laws, and theories, and place typical pragmatic approaches to hydrologic model development in the context of discovery and learning. In Section 3 we discuss exemplary areas in hydrology where the community has made progress in understanding hydrologic processes and developing mathematical representations of the process understanding. Our intent is to demonstrate how current hydrologic theory can be used to improve the scientific rigor of hydrologic models. In Section 4 we briefly discuss how current hydrologic theories can be incorporated within a modeling framework, and in Section 5 we discuss how the theories can be tested and refined. We close with concluding remarks in Section 6.

2. The gap between theory and models

Theory means different things to different people. To some, theory defines a concept that is unproven – a guess, or an educated guess – rather than a systematized understanding with explanatory power [Corneliussen 2015]. To others, theory is an antonym of application, where efforts may be described as "very theoretical" even in the absence of explanation. More precise definitions are available in recent papers in hydrology. For example, Sivalapan [2005] defines theory as "*the set of ideas or concepts that is best able to describe or explain the system of interest, the catchment, its presence in the landscape, its behavior, and its function in relation to other systems*". Similarly, Ehret et al. [2014] define theory as the "*Explanation of some aspect of the natural world, established by following the scientific method and confirmed by observation and experiment (empirical evidence). A theory has* *explanatory and predictive power; its strength is related to the parsimony of its principles, the diversity of phenomena it can explain and the quality of its falsifiable predictions […]*". A theory is distinct from a scientific law, which describes, but does not explain, consistently observed phenomena (e.g., Fourier's Law, Fick's Law, Ohm's Law, or even Darcy's Law, all of which are used in hydrologic models). A theory is also distinct from a hypothesis, which is a falsifiable statement (usually a quantifiable corollary under specific conditions) used to test a given theory (Figure 1).

For us, hydrologic theories are the stories that we tell to explain observed hydrologic processes. In this simple and general definition, we permit theories of varying strength (i.e., of varying explanatory and predictive power), we permit theories that explain and predict only a subset of hydrologic processes (i.e., the theories need not be comprehensive), and we do not require that theories be accepted as an accurate explanation by a broad cross-section of the scientific community. In this sense we define hydrologic theory as our explanations of individual processes, process interactions, patterns and scaling behavior. Our definition of theory is deliberately permissive – we all strive for theories that are strong, unified, and well accepted, but at this stage we do not impose such restrictions so that we can focus on testing, refining, and reconciling the widest set of theories that already exist.

To expand on what we mean by theory we provide some examples of general process explanations. First, consider snowmelt. Snowmelt is driven by the net fluxes of solar and long wave radiation, sensible and latent heat, the heat advected by precipitation, and the diffusion of heat throughout the snow-soil system [Clark et al. 2015c; equation 11]. This general understanding of snowmelt energetics is well established and incorporated into process-based models [Slater et al. 2001; Etchevers et al. 2004]. A common algorithmic simplification is that snowmelt can be parameterized as a function of air temperature [Hock 2003], which could be put in "law" form as "snow melts faster on warm days", or "the amount of snow that melts each day varies linearly with air temperature". This temperature-index approach to snow modeling has some relationship with energy balance theory. For example, several components of the energy balance, including sensible heat flux and incident longwave radiation (which dominates the energy balance in many settings [Ohmura, 2001]), are explicit functions of air temperature. However, generalizing using temperature-index snow models is limited because strong spatial variations in temperature-melt relationships make it difficult to extrapolate the model parameters across space. Moreover, these simplifications are likely to fail for extreme events or under climate change where the correlation between air temperature and snow-atmosphere energy fluxes is non-stationary [Huss et al. 2009]. Similar issues may arise for other physical processes – for example, parameterizing potential evapotranspiration as an empirical function of air temperature, i.e., neglecting energy balance theory, can exaggerate the hydrologic sensitivity to climate change [Milly and Dunne 2011; Sheffield et al. 2012].

More generally, consider explanations (theory) for a suite of interacting hydrologic processes. For example, with theory encoded in a model we can (to some extent) explain and predict the area-average infiltration due to spatial variability in water table depth [Beven and Kirkby 1979], spatial variability in soil moisture [Moore and Clarke 1981; Wood et al. 1992], or spatial variability in hydraulic conductivity [Hawkins and Cundy 1987]. We have critiques and comparisons of these process theories [Beven 1997; Clark and Gedney 2008; Clark et al. 2008]. Similarly, we can explain and predict area-average transpiration related to spatial variability in vegetation phenology [Koster and Suarez 1992; Liang et al. 1994; Bonan et al. 2002] or related to spatial variability in plant-available water [Famiglietti and Wood 1994; Koster et al. 2000]. We can explain and predict non-linear recession behavior based on spatial heterogeneity in hydraulic conductivity [Clark et al. 2009; Harman et al. 2009]. We also can explain and predict the non-linearity in runoff generation associated with thresholds, hydrologic connectivity, and hydrologic hotspots [Tromp‐van Meerveld and McDonnell 2006; Lehmann et al. 2007; Seyfried et al. 2009; Zehe and Sivapalan 2009; Jencso and McGlynn 2011]. The point of highlighting these few example research areas is that many process explanations already exist; the issue is that many important process explanations are not widely implemented as falsifiable hypotheses. Most commonly, the explanations (theory) for a particular behavior (formulated as a law rather than a hypothesis) are accepted within a given model and applied outside of the basins where the theory may have originally been developed and tested. This represents a missed opportunity to generalize and further test the same theory in different basins.

The gap between theory and models becomes evident when we consider that, in practice, a pragmatic rather than a process-based approach to hydrologic model development is generally followed. The pragmatic approach uses spatial discretizations, process parameterizations, and time stepping schemes borrowed from other extant models (e.g., reliance on the 1D moisture-based form of Richards' equation in land models [Clark et al. 2015a]). The pragmatic approach is often quite effective in generating predictions – multiple processes can lead to similar behavior, and hence multiple processes can be represented by the same law [McDonnell 2013]; however, this often comes at the expense of poor explanatory power and poor parameter transferability. The process-based approach, by contrast, is the classical approach described in recent textbooks and papers [Beven 2011; Gupta et al. 2012]: to first develop a conceptual representation of our understanding of how the world works based on inductive reasoning from observations, i.e., the theories we use to explain hydrologic behavior, and then encode algorithmic simplifications of our conceptualizations in a numerical model. Only a handful of hydrologic studies have followed the process-based approach to hydrologic models, by encoding theories as testable hypotheses in order to challenge and refine our understanding of hydrologic behavior [Freer et al. 2004; Lehmann et al. 2007; McMillan et al. 2012; Euser et al. 2013; Fenicia et al. 2014].

The pragmatic modeling approach most often applied in practice tends to sever the link between the models and the body of theory, thereby impeding continued refinement of our process understanding. Specifically, the pragmatic approach focuses attention on a model's predictive competence rather than its explanatory power. This limits our ability to generalize about hydrologic behaviors, leading to model "tuning" for particular basins, giving the impression that every basin is unique [Beven 2000; McDonnell et al. 2007]. If we cannot trust these models to generalize across observed space now, how can we trust them to predict historically unseen conditions? There is, at present, only a thin theoretical foundation to support applying models in new settings. Even worse, when models fail in new settings it is difficult to know which body of theory requires updating, particularly when it is faster and easier to update the parameters and move on with the immediate task at hand – generating predictions.

3. Towards a model-based synthesis of hydrologic theory

We now return to the primary concern of this paper: to reconcile hydrologic models with existing hydrologic theory. The first question is then "what theory"? Do the elements of hydrologic theory already exist, or is theory something that the hydrologic research community has yet to discover? An examination of the relevant literature – e.g., *Searching* *for the Holy Grail of scientific hydrology* [Beven 2006b] – suggests that our quest for explanations and model parameterizations of large-scale fluxes has not yet been successful. Here we take a more positive view. We accept that there is a tremendous amount of work to do, but, crucially, we recognize that we do already have many existing process explanations and model parameterizations that can be much better exploited in models than has been done to date.

Key questions that need to be addressed are

1) What existing hydrologic theories are included in models and what aspects of theory are ignored or not well assimilated?

2) What are the most important aspects of hydrologic theory that are not yet incorporated in models?

3) In what parts of extant models do existing theories have the most (and the least) explanatory power?

The first issue at hand is therefore to identify some useful elements of existing hydrologic theory. We consider advances in both the explanatory and predictive capabilities of models in three main areas: (a) Developing ways for the structure of the landscape to be better represented in the structure of models; (b) Advancing understanding of how small-scale processes combine to produce large-scale fluxes (emergent behavior) and the development of ways to parameterize this effect in models; and (c) Advancing understanding of how the principles of optimality (or ecological and landscape evolution) can be used to constrain model behavior. The following subsections expand on these topics.

3.1 Reflecting the structure of the landscape in the structure of models

The modeling community has pursued multiple methods to reflect the structure of the landscape in hydrologic models. An interesting example is Keith Beven's "alternative blueprint" [Beven 2002], which provides a substitute for the Freeze-Harlan blueprint for physics-based hydrologic modeling [Freeze and Harlan 1969]. Beven's idea is that the structure of hydrologic models should reflect the structure of the landscape (e.g., topography, vegetation, soils, geology), and he emphasizes the need to extensively experiment with different model structures and parameter sets in order to identify an ensemble of "behavioral" hydrologic models [Beven 2002]. However, applications of this alternative blueprint typically use models of lower state dimension, i.e., models with extensive lumping of physical processes and of the physical landscape, which can obscure the connection between the model structure and the landscape structure [although see Peters et al. 2003; Rinaldo et al. 2006; Fenicia et al. 2014]. The key question here is as follows: To what extent do models reflect our explanations of landscape controls on the space-time variability in hydrologic states and fluxes?

To develop and test theories that relate landscape properties to hydrologic behavior, we propose that the following tasks should be systematically dealt with:

a) Investigate how available theories can enable information on geomorphology, topography, vegetation, soils, and geology to be better used for defining model structure/parameters in different landscapes [Samaniego et al. 2010; Schaefli et al. 2014; Zehe et al. 2014].

b) Investigate the challenges in model-landscape mapping when hydrologic models are the basis for water quality and stream ecosystem models. Typical challenges include how to incorporate representations of the dynamics of surface flow connectivity between sediment sources and the stream channel [Bracken et al. 2015], the distinct thermal and biogeochemical signatures associated with different flow paths and network topology [Kurylyk et al. 2014; Leach and Moore 2015], as well as the behaviors of in-stream algae, invertebrates and fish [Power et al. 1995; Ceola et al. 2014];

c) Develop approaches for model-landscape mapping that can be applied in models of varying complexity, and account for landscape heterogeneity; and

d) Investigate to what extent it is possible, with typically available information, to discriminate among competing models to define alternative model structures in different landscapes [Jakeman and Hornberger 1993; Gupta and Nearing 2014].

These issues dig into to the heart of different philosophical approaches to hydrologic modeling [Harman and Troch 2014], especially the extent to which the details of the landscape are included in models, and the extent to which modelers pursue the quest for explanation versus prediction. For example, does the lumping of processes and the landscape in spatially lumped models limit the extent to which the structure of the landscape can be reflected in the structure of models? Put differently, is the structure of the landscape actually better reflected in spatially explicit models, where the higher granularity of process representations and the higher granularity of the landscape discretization enables examination of how geomorphology and spatial variability in topography, vegetation, soils, and geology affect the space-time variability in hydrologic states and fluxes? To what extent are spatially explicit models limited by the available data? Are models with detailed spatial representations extensible to other watersheds that are very different from where they were developed? Focused attention on these issues will help with the model implementation and testing of theories that map patterns to processes [Sivapalan 2005; McDonnell et al. 2007], and will help improve how the details of the landscape are represented in models [Wigmosta et al. 1994; Beven and Freer 2001; Bonan et al. 2002; Tague and Band 2004; Vivoni et al. 2005; Clark et al. 2015b].

3.2 Scale-Emergent behavior

A key challenge in hydrologic model development is to explain and predict how small-scale processes combine to affect large-scale fluxes [Reggiani et al. 1998; Reggiani et al. 1999; Beven 2006b; McDonnell et al. 2007; Troch et al. 2009]. This typically involves (a) formulating conservation equations for physically meaningful control volumes within the model domain; and (b) parameterizing fluxes at the boundaries of model control volumes in a way that represents the impact of sub-grid scale heterogeneities on grid-average fluxes. A major model development challenge is parameterizing grid-average fluxes, termed the “closure problem” [Reggiani et al. 1998; Reggiani et al. 1999; Reggiani and Schellekens 2003; Beven 2006b]. Solutions to the closure problem have proved to be rather difficult [Zehe et al. 2006; Harman and Sivapalan 2009].

To synthesize current hydrologic theory and modeling approaches, and to advance scale-appropriate flux parameterizations, the following tasks should receive immediate attention:

a) Identify which theories can explain and predict the impacts of structural and process heterogeneity on large-scale fluxes; and

b) Investigate the relative advantages of the different methods used to represent how small-scale process interactions affect large-scale behavior.

In addressing these tasks we recognize that emergent behavior has been represented in many different ways in many different models, providing an existing theoretical backbone to hydrologic models. The main approaches are (a) spatial integration of the small-scale equations [Maxwell and Kollet 2008; Kollet et al. 2010]; (b) development of "scale-appropriate" flux parameterizations, such as sub-grid probability distributions [Beven and Kirkby 1979; Moore and Clarke 1981; Liang et al. 1996; Koren et al. 1999; Luce et al. 1999], and new (upscaled) model equations [Mahrt 1987; Essery et al. 2008], including empirically-derived storage-discharge relationships [Ambroise et al. 1996; Clark et al. 2008; Fenicia et al. 2011]; (c) representing the role of thresholds and connectivity in defining larger-scale responses (e.g., the need to fulfill depression storage as in wetlands and bedrock topography) [Freer et al. 2002; Tromp - van Meerveld and McDonnell 2006; Clark et al. 2009; Jencso et al. 2009; Zehe and Sivapalan 2009; Spence et al. 2010; Shook et al. 2013]; and (d) formulation of macroscopic principles acting at the scale of interest [Rodríguez - Iturbe et al. 1992; Caylor et al. 2009; Schymanski et al. 2009; Schymanski et al. 2010]. These modeling approaches are not mutually exclusive, indicating the lack of a unifying theory in hydrology [Sivapalan 2005]. Most models include some mix of methods to parameterize the impact of sub-grid scale heterogeneities on large-scale fluxes, and it is necessary to synthesize, evaluate, and compare these methods, and most particularly the theory that they encode, in order to improve explanations of hydrologic processes and improve the physical realism of hydrologic model structures.

3.3 The use of optimality principles to constrain model behavior

Optimality is an interesting idea popularized in the fields of eco-hydrology and geomorphology – for example, that vegetation makes optimal use of the available resources, or that the landscape evolves to dissipate energy gradients in response to external forcings [Rodríguez - Iturbe et al. 1992; Rigon et al. 1993; Eagleson 2002; Schymanski et al. 2008; Schymanski et al. 2009; Zehe et al. 2013; Bonan et al. 2014]. As opposed to predicting large scale behavior emerging from small-scale processes, optimality seeks to predict behavior at the scale of interest by assuming that the system of interest self-organizes following some macroscopic extremum principle, such as the maximization of net carbon profit for vegetation, or the maximization of energy dissipation or entropy production for physical and biological systems. The beauty of extremum principles is that they can potentially dramatically reduce the number of unknowns and hence degrees of freedom in a system, pre-constraining calibration and therefore facilitating generalization and consequently testing/falsification [e.g., see Schymanski et al. 2008; Schymanski et al. 2009]. The theory of optimality can therefore be a key organizing principle for hydrologic model development [Schaefli et al. 2011].

The following tasks regarding application of the theory of optimality in support of hydrologic model development should be carried out in order to assess the utility of such principles in hydrological models:

a) Investigate how optimality principles can be implemented and tested in hydrological models in a comparative way, and to what extent they improve predictions at the scale of interest;

b) Investigate how (and to what extent) the principles of optimality can be used to constrain model behavior – specifically, what processes can be constrained by optimality principles, and at what spatial and temporal scales; and

c) Investigate how optimality principles can be applied in highly heterogeneous landscapes and identify the relevant scales of different optimality principles.

These issues relate to the limits to which optimality principles can be applied – for example, how optimality principles that govern energy fluxes and root dynamics differ in water-limited and energy-limited environments [Schymanski et al. 2008]; whether a thermodynamically optimal hillslope structure can explain the spatial organization of flow networks and serve as a first guess for uncalibrated predictions of rainfall-runoff processes [Zehe et al. 2013]; whether optimality principles are applicable in basins subject to substantial human modification. More generally, how can the theory of optimality be used to constrain interactions among processes and across scales? In our opinion, further exploration of optimality principles may yield useful constraints that will improve the fidelity and general applicability of hydrologic models.

3.4 Summary: Capitalizing on existing theory

The intent of our discussion here is two-fold: (1) to focus attention on some key areas where hydrologic theory already exists; and (2) to define a set of issues that need to be addressed in order to better represent this existing theory in models. For landscape structure, we recognize the opportunities to improve model representations of how landscape structure affects the space-time variability in hydrologic states and fluxes; and also that there are substantial challenges associated with data limitations and model identification in order to incorporate landscape structure in models with different complexity and with different intended purposes. For scale-emergent behavior, we recognize large advances in our capabilities to explain and predict fluxes of water and energy at larger scales; but note that we still lack information on the general applicability and relative merit of these different explanations and model parameterizations. For optimality, we recognize its potential for greatly reducing the number of unknown parameters; and also that we do not yet understand the limits of this theory or the extent to which it may be useful in different physical settings. These issues bring us to the next two challenges: How can theories be incorporated in models, and how can the theories be evaluated?

4. Model construction: Implementing theories in models

To define a path forward for model construction we return to our original premise: Modern hydrologic models do not reflect the current understanding of hydrologic processes, i.e., theory. Hydrologic models are too often based on empirical postulates (e.g., parsimonious storage-discharge relationships that describe the aggregate response of a catchment to external forcing). Or, at the other end of the spectrum, hydrologic models are based on physically motivated partial-differential equations that do not represent the impact of small-scale heterogeneities on large-scale fluxes (e.g., many physically motivated models neglect the importance of hillslope-scale connectivity and preferential flow in shaping catchment-scale fluxes). Such weak theoretical underpinnings lead several commentators to criticize the current generation of models for "getting the right answers for the wrong reasons" [Kirchner 2006] or for being "process weak" [McDonnell et al. 2007].

Improving the theoretical underpinnings of hydrologic models requires a modeling system to systematically evaluate different numerical implementations of current hydrologic theories. As a next step we propose that the following tasks need to be carried out:

a) Find ways to best encode different theories in our models, to allow for hypothesis generation, testing, grading, selection, and structured model improvement; and

b) Investigate how (and where) we can best incorporate algorithmic simplifications of varying complexity, to represent limited knowledge of hydrologic processes and catchment characteristics.

Addressing these issues is possible through recent model development efforts that pursue the method of multiple working hypotheses [Chamberlin 1890; Clark et al. 2011]. Modern model implementations of the method of the multiple working hypotheses now exist, offering a "master template" from which it is possible to incorporate different modeling decisions, process parameterizations and spatial organization [Kraft et al. 2011; Niu et al. 2011; Essery et al. 2013; Clark et al. 2015b; Clark et al. 2015c]. Recent advances in the development of multiple hypothesis modeling frameworks include: 1) The capability to represent all the biophysical and hydrologic processes thought to be relevant, extending beyond traditional land surface models as well as traditional hydrology models, and including options for model simplification (e.g., ignore or implicitly represent specific state variables and fluxes); 2) Implementation of modeling approaches in a clear and modular fashion, in order to incorporate multiple competing hypotheses of hydrologic behavior; 3) Flexible and hierarchical spatial organization, in order to experiment with different model representations of spatial variability and hydrologic connectivity; and 4) Incorporation of different strategies to estimate and adjust model parameters [Clark et al. 2015b]. These advances notwithstanding, further developments to multiple hypothesis modeling frameworks are required to better incorporate existing hydrologic theory (as proposed in this paper).

A key research priority is to define a community-based approach to incorporate hydrologic theories in models, building on the successful implementation of community models in the atmospheric science and land-atmosphere interactions communities [Lawrence et al. 2011; Hurrell et al. 2013]. This issue has received some attention in the hydrologic literature; most recently where Weiler and Beven [2015] consider the need for a community hydrologic model. Weiler and Beven offer an interesting and wide-ranging discussion on the challenges of agreeing on the modeling concepts, of adequate support and effective governance, and, critically, in the context of this paper, of the need to evaluate alternative formulations of sub-element parameterizations at different spatial scales and hydrologic regimes. Weiler and Beven argue that "*the most important aspect of a Community Modeling Initiative is to instigate a discussion* [on what process parameterizations should look like], *test the potential alternatives, understand their domain of applicability, and agree on a formulation, before such a model is released for general use*." Weiler and Beven [2015] deliberately avoid defining what a model should look like and how a model can be tested, and they leave as an open question whether such a community model could be programmed in a way that is agile enough to be used as an effective learning tool.

Here we propose a specific path forward for community modeling that is more focused than the path proposed by Weiler and Beven [2015]: Our primary aim is to evaluate alternative hydrologic theories and associated process parameterizations as well as alternative modeling concepts. We impose no requirement that we attain *agreement* on modeling concepts, and we hence deliberately take a model agnostic position to implement and test multiple theories and associated process parameterizations. Our proposed approach is the unified approach to hydrologic modeling defined by Clark et al. [2015b]. This modeling approach cleanly separates the conservation equations from the flux parameterizations, providing the flexibility to incorporate multiple modeling options to calculate the flux across the boundaries of model control volumes. The modeling approach employs hierarchal data structures, providing the flexibility to define multiple representations of spatial variability and hydrologic connectivity, including models with different spatial architecture and complexity. This flexibility enables users to isolate and evaluate individual modeling decisions, enabling the use of models as virtual laboratories [Weiler and McDonnell 2004; Sivapalan 2005; Blöschl 2006; Wagener et al. 2010] to help formalize and evaluate alternative hydrologic theories.

An important point here is the need for a community modeling *process* rather than a community hydrologic model. Given the diverse range of questions that the discipline of Hydrology seeks to answer, it is unreasonable (and unwise) to formulate a single community model for all purposes. The critical need is to further develop the "community of practice" of hydrologic modeling to consistently test and compare competing hypotheses and algorithms, i.e., to test and compare competing modeling approaches. This requires strong community engagement in formulating and evaluating multiple competing hypotheses. Such community efforts should be conducted within modeling frameworks that recognize the similarities among extant models, and control for their differences, and hence help form general conclusions of widespread relevance across models with very different objectives. A key metric of success is to incorporate our best theory in a wider range of multi-disciplinary modeling efforts, such as improving the representation of hydrologic processes in Earth System Models [Clark et al. 2015a], to ensure robust predictions of global environmental change.

5. Model evaluation: Developing a rigorous approach to evaluate and select among competing theories

A key issue in hydrologic model development, and also in achieving the solidification of theory, is the rationale used to select among competing alternatives. Our principles for model development are often based on individual philosophical penchants for either physics or parsimony [Ebel and Loague 2006; McDonnell et al. 2007], but neither is fully supported by data or model analysis [Smith et al. 2013; Mendoza et al. 2015]. Here we argue for a more systematic and robust approach to discriminate among model alternatives.

The issue at hand is the process for theoretical development outlined in Figure 1. At the stage where the data confronts the model, there is an option to detour on a side-loop where we calibrate model parameters – perhaps parameters derived from the theory in question, or perhaps parameters in other parts of the model. It is well understood that we can take that side-loop many times to avoid having a beautiful theory slayed by ugly facts [Huxley 1894]. Put differently, the process of model calibration can render model hypotheses unfalsifiable. This leads us to wonder whether we are actually in a situation where the current state of hydrology is too accepting of competing theories, and where hydrologic applications rely on calibration at the expense of understanding because explanation cannot be established with sufficient confidence.

Addressing the following questions will help to challenge and refine our hydrologic theories:

a) How can we best distinguish among competing theories? How can we best balance quantitative and qualitative insights to challenge and refine theories [Seibert and McDonnell 2002; Freer et al. 2004; Winsemius et al. 2009; Euser et al. 2013; Seibert and McDonnell 2013; Birkel et al. 2014; Wrede et al. 2014], especially given limitations of information on internal model states and fluxes;

b) What are the best model application practices for testing theories as opposed to continuing to increase model complexity;

c) What does a falsification framework look like? How can we improve understanding of the worth of data and the sensitivity of model rejection to assumptions and experimental designs? How can we meaningfully discriminate among competing hypotheses in the presence of incomplete and inexact information; and

d) What are the applications (and limits) of information theory to select among competing theories [Gupta and Nearing 2014; Nearing and Gupta 2014]?

We suggest performing at least the following tests to evaluate a hydrologic model or a land surface model: 1) Evaluate model simulations at internal locations within a given model element (e.g., eddy covariance stations, cosmic ray probes, streamflow gauging stations, snow depth measurements) not used during parameter estimation [Freer et al. 2004; Smith et al. 2013; Rakovec et al. 2015]; 2) Evaluate model simulations at many locations, especially those with climatic regimes different from that used for parameter estimation [Nijssen et al. 2001; Seibert 2003; Wenger et al. 2010; Coron et al. 2012]; 3) Evaluate internal model states across multiple spatial scales [Kumar et al. 2012]; 4) Test the flux matching condition between simulated fluxes across scales [Samaniego et al. 2010; Kumar et al. 2013]; and 5) Assess comparability and reproducibility of model results [Ceola et al. 2015]. The fundamental goal is to evaluate energy fluxes at the native scales at which observations can be made, for example from control volumes varying from 10^2 m (cosmic ray probe) to 10^5 m (GRACE satellite footprint). The common practice in hydrology of using univariate signatures to infer model parameters provides only weak constraints on model simulations of the terrestrial hydrologic cycle [Gupta et al. 2008; Rakovec et al. 2015].

The path toward meaningful model evaluation must embrace underdeterminism [Kleinhans et al. 2005; Beven 2006a, 2006b]. One path forward is to conduct controlled experiments, e.g., through targeted collection of the necessary data to test specific model constructs and hypotheses. This path was proposed by Zehe et al. [2014] to test their ideas of using functional units to represent the spatial organization of hydrologic processes. While critical, resource constraints invariably mean that such controlled experiments are limited in extent, constraining our capabilities to generalize. A parallel path forward is to evaluate individual model hypotheses in isolation [Clark et al. 2011]. This involves first decomposing a high-dimensional model into the individual decisions made during model development (as discussed in Section 4), and then making better use of the data that we do have to evaluate different model development decisions (which are ideally formulated as falsifiable concepts; see Figure 1). In this context underdeteminism can be reduced by defining metrics, or diagnostic signatures, that provide insight into the internal states and fluxes [Kirchner et al. 1996; Gupta et al. 2008; Euser et al. 2013; Birkel et al. 2014], using new measurement technologies that provide information at higher spatial and temporal resolution or that cover larger spatial areas [Tyler et al. 2009; Zreda et al. 2012], and qualitative insights [Winsemius et al. 2009; Wrede et al. 2014]. Underlying both of these paths is uncertainty in myriad sources. Uncertainty in model inputs, in the details of landscape structure, and in evaluation data are all important factors limiting the extent to which it is possible to discriminate among competing model alternatives; hence characterizing these uncertainties in a meaningful way is crucial to avoid incorrectly rejecting behavioral model structures [Beven et al. 2012; Clark et al. 2012]. Ultimately, if the outcome of the evaluation procedure is the inability to test a given hypothesis with our current observation capabilities, this would indicate a need for additional theory development, the need to identify priorities for future observing capabilities, or both.

As with model construction, evaluation should be a community effort: That is, where the community actively compares and debates the merits of alternative evaluation approaches using a framework that helps minimize the differences among models and model configurations [Ceola et al. 2015; Clark et al. 2015b]. This enables the community to move forward from developing models for particular basins to models maintained by a community and tested everywhere. The key to progress is to find cases around the world in which community models (sets of hypotheses) do not work well, and also where data exist with sufficient quality and density to evaluate why. These cases will provide the hints on how to move forward – failures are therefore the key to improve our theories. The ultimate goal is to have open source community models that are in principle applicable worldwide, and have open source multivariate and multi-scale data available for comprehensive model evaluation (recognizing data paucity and uncertainty). This requires substantial breadth of information across a diverse range of watershed types along with demonstrated depth of observing capabilities in specific locations [Gupta et al. 2014]. With that, we anticipate that the community will drastically advance model evaluation frameworks, gather and bring together relevant data, extensively test hypotheses, and accelerate progress for the discipline of hydrology as a whole.

6. Concluding remarks

Many have argued that there is a need to “discover” new laws and theories in hydrology. These discussions have tended to focus on particular problem areas (e.g., floods) or processes (e.g., hillslope storage). An underlying common theme has emerged where laws and theories are lacking to address these challenges in a common way – that there are no general principles, only separate applications to unique catchments [Beven 2000; McDonnell et al. 2007]. In this Commentary we depart from earlier narratives by arguing that substantial bodies of theory already exist for hydrology, but are rarely recognized as such; moreover, important elements and insights drawn from existing theories are not widely or consistently implemented and tested in hydrologic models, particularly for regular applications outside of watersheds where individual models have been developed and tested. More generally, we argue that the growing gap between models and theory is impeding the progress of hydrologic science.

We propose here that it is possible to improve the theoretical underpinnings of hydrologic models by focusing attention on three related issues. First, we propose that a useful starting point is the synthesis of our understanding of hydrologic processes (hydrologic theory), based on commonly observed behavior in research watersheds (formulated as hydrologic laws). Ultimately this synthesis will result in multiple algorithmic simplifications of the components of hydrologic theory, including algorithms of varying complexity. Second, we propose that these multiple theory-based conceptualizations be systematically incorporated into community models, encoding theory into models as multiple testable hypotheses, to enable systematic scrutiny of competing hypotheses. Third, we propose that comprehensive, multi-scale, diagnostic, model evaluation be designed and systematically carried out to apply, challenge, and subsequently refine current hydrologic theory and its instantiations in hydrology models. Our proposed synthesis effort requires research to systematically formulate, organize, encode, and evaluate hydrologic theories, so that our models synthesize the best process understanding and are used as an avenue to evaluate and refine hydrologic theories. A key challenge is to develop methods that use incomplete and inexact information to effectively evaluate competing hypotheses, and to improve the extent to which we can scrutinize and refine hydrologic theories. Such a synthesis will strengthen the link between algorithms, theory and observations, improving our understanding of the impact of model simplifications, increasing the fidelity of model simulations, and, ultimately, increasing our confidence in model predictions.

Pursuing the questions defined in this paper will be challenging, and requires strong community engagement. The questions we pose require a broad range of interdisciplinary expertise; the quest for generality requires synthesis across a broad range of hydroclimatic regimes and geological settings, and an enhanced model-based synthesis and evaluation procedure requires developing creative and effective methods for model construction and analysis. We therefore welcome collaborations from scientists interested in the synthesis of process explanations and modeling approaches across diverse physical environments, in constructing models to encode the components of hydrologic theory as testable hypotheses, and in advancing model evaluation efforts to provide meaningful and comprehensive evaluation of model alternatives (i.e., model evaluation under uncertainty). Such strong community engagement will enable the community to move forward from developing models for particular basins to theoretically grounded models maintained by the community and tested everywhere, which will accelerate the continuing refinement of hydrologic models and the grounding theory they encode.

Acknowledgements

We appreciate comments from Bart Nijssen, David Rupp, Nans Addor, Uwe Ehret, Hoshin Gupta, and from two anonymous reviewers. This commentary is related to the Panta Rhei Research Initiative of the International Association of Hydrological Sciences (IAHS).

7. References

Ambroise, B., K. Beven, and J. Freer, 1996: Toward a generalization of the TOPMODEL concepts: Topographic indices of hydrological similarity. *Water Resources Research*, **32,** 2135-2145, doi: 10.1029/95wr03716.

Beven, K., and J. Freer, 2001: A dynamic TOPMODEL. *Hydrological Processes*, **15,** 1993-2011, doi: 10.1002/hyp.252.

Beven, K., 2006a: A manifesto for the equifinality thesis. *Journal of Hydrology*, **320,** 18-36, doi: 10.1016/j.jhydrol.2005.07.007.

Beven, K., 2006b: Searching for the Holy Grail of scientific hydrology: Q(t) = H((S)under-left-arrow, (R)under-left-arrow, Delta t)A as closure. *Hydrology and Earth System Sciences*, **10,** 609-618, doi: 10.5194/hess-10-609-2006.

Beven, K., 2007: Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process. *Hydrology and Earth System Sciences*, **11,** 460-467
Beven, K. J., and M. J. Kirkby, 1979: A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal* **24.1,** 43-69, doi: 10.1080/02626667909491834.
Beven, K. J., 1997: TOPMODEL: A critique. *Hydrological Processes*, **11,** 1069-1085
Beven, K. J., 2000: Uniqueness of place and process representations in hydrological modelling. *Hydrology and Earth System Sciences*, **4,** 203-213
Beven, K. J., 2002: Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrological Processes*, **16,** 189-206, doi: 10.1002/hyp.343.
Beven, K. J., 2011: *Rainfall-runoff modelling: the primer.* John Wiley & Sons.
Beven, K. J., P. Smith, I. Westerberg, and J. Freer, 2012: Comment on "Pursuing the method of multiple working hypotheses for hydrological modeling" by M. P. Clark et al. *Water Resources Research*, **48**, doi: W11801, 10.1029/2012wr012282.
Birkel, C., C. Soulsby, and D. Tetzlaff, 2014: Developing a consistent process- based conceptualization of catchment functioning using measurements of internal state variables. *Water Resources Research*, **50,** 3481-3501, doi: 10.1002/2013wr014925.
Blöschl, G., 2006: Hydrologic synthesis: Across processes, places, and scales. *Water Resources Research*, **42**
Bonan, G., M. Williams, R. Fisher, and K. Oleson, 2014: Modeling stomatal conductance in the earth system: linking leaf water-use efficiency and water transport along the soil–plant–atmosphere continuum. *Geoscientific Model Development*, **7,** 2193-2222
Bonan, G. B., S. Levis, L. Kergoat, and K. W. Oleson, 2002: Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochemical Cycles*, **16**, doi: 10.1029/2000gb001360.
Bracken, L. J., L. Turnbull, J. Wainwright, and P. Bogaart, 2015: Sediment connectivity: a framework for understanding sediment transfer at multiple scales. *Earth Surface Processes and Landforms*, **40,** 177-188
Cartwright, N., 1983: *How the laws of physics lie.* Cambridge Univ Press.
Caylor, K. K., T. M. Scanlon, and I. Rodriguez-Iturbe, 2009: Ecohydrological optimization of pattern and processes in water-limited ecosystems: A trade-off-based hypothesis *Water Resources Research*, **45,** W08407 doi: 10.1029/2008WR007230
Ceola, S., E. Bertuzzo, G. Singer, T. J. Battin, A. Montanari, and A. Rinaldo, 2014: Hydrologic controls on basin‐scale distribution of benthic invertebrates. *Water Resources Research*, **50,** 2903-2920
Ceola, S., B. Arheimer, E. Baratti, G. Blöschl, R. Capell, A. Castellarin, J. Freer, D. Han, M. Hrachowitz, and Y. Hundecha, 2015: Virtual laboratories: new opportunities for collaborative water science. *Hydrology and Earth System Sciences*, **19,** 2101-2117
Chamberlin, T. C., 1890: The method of multiple working hypotheses. *Science*, **15,** 92-96
Clark, D. B., and N. Gedney, 2008: Representing the effects of subgrid variability of soil moisture on runoff generation in a land surface model. *Journal of Geophysical Research-Atmospheres*, **113**, doi: 10.1029/2007jd008940.
Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay, 2008: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, **44**, doi: 1029/2007WR006735.
Clark, M. P., D. E. Rupp, R. A. Woods, H. J. Tromp-van Meerveld, N. E. Peters, and J. E. Freer, 2009: Consistency between hydrological models and field observations: linking processes at the hillslope scale to hydrological responses at the watershed scale. *Hydrological Processes*, **23,** 311-319, doi: 10.1002/hyp.7154.
Clark, M. P., D. Kavetski, and F. Fenicia, 2011: Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, **47**, doi: 10.1029/2010WR009827.
Clark, M. P., D. Kavetski, and F. Fenicia, 2012: Reply to comment by K. Beven et al. on "Pursuing the method of multiple working hypotheses for hydrological modeling". *Water Resources Research*, **48**, doi: W11802, 10.1029/2012wr012547.

Clark, M. P., Y. Fan, D. L. Lawrence, J. C. Adam, D. Bolster, D. Gochis, R. L. Hooper, M. Kumar, L. R. Leung, D. S. Mackay, R. M. Maxwell, C. Shen, S. C. Swenson, and X. Zeng, 2015a: Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research*, doi: 10.1002/2015WR017096.

Clark, M. P., B. Nijssen, J. Lundquist, D. Kavetski, D. Rupp, R. Woods, E. Gutmann, A. Wood, L. Brekke, J. Arnold, D. Gochis, and R. Rasmussen, 2015b: A unified approach to process-based hydrologic modeling. Part 1: Modeling concept. *Water Resources Research*, **51**, doi: 10.1002/2015WR017198.

Clark, M. P., B. Nijssen, J. Lundquist, D. Kavetski, D. Rupp, R. Woods, E. Gutmann, A. Wood, D. Gochis, R. Rasmussen, D. Tarboton, V. Mahat, G. Flerchinger, and D. Marks, 2015c: A unified approach for process-based hydrologic modeling: Part 2. Model implementation and example applications. *Water Resources Research*, **51**, doi: 10.1002/2015WR017200.

Corneliussen, S. T., 2015: Should scientists think harder about explaining the concept "theory"? *Physics Today, Daily Edition*

Coron, L., V. Andréassian, C. Perrin, J. Lerat, J. Vaze, M. Bourqui, and F. Hendrickx, 2012: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. *Water Resour. Res.*, **48,** W05552, doi: 10.1029/2011wr011721.

Dooge, J. C. I., 1986: Looking for hydrologic laws. *Water Resources Research*, **22,** 46S-58S, doi: 10.1029/WR022i09Sp0046S.

Eagleson, P. S., 2002: *Ecohydrology: Darwinian expression of vegetation form and function.* Cambridge University Press.

Ebel, B. A., and K. Loague, 2006: Physics based hydrologic response simulation: Seeing through the fog of equifinality. *Hydrological Processes*, **20,** 2887-2900, doi: 10.1002/hyp.6388.

Ehret, U., and Coauthors, 2014: Advancing catchment hydrology to deal with predictions under change. *Hydrol. Earth Syst. Sci.*, **18,** 649-671, doi: 10.5194/hess-18-649-2014.

Essery, R., P. Bunting, J. Hardy, T. Link, D. Marks, R. Melloh, J. Pomeroy, A. Rowlands, and N. Rutter, 2008: Radiative transfer modeling of a coniferous canopy characterized by airborne remote sensing. *Journal of Hydrometeorology*, **9,** 228-241, doi: 10.1175/2007jhm870.1.

Essery, R., S. Morin, Y. Lejeune, and C. B. Menard, 2013: A comparison of 1701 snow models using observations from an alpine site. *Advances in Water Resources*, **55,** 131-148, doi: 10.1016/j.advwatres.2012.07.013.

Etchevers, P., and Coauthors, 2004: Validation of the energy budget of an alpine snowpack simulated by several snow models (SnowMIP project). *Annals of Glaciology, Vol 38, 2004*, P. M. B. Fohn, Ed., 150-158.

Euser, T., H. C. Winsemius, M. Hrachowitz, F. Fenicia, S. Uhlenbrook, and H. H. G. Savenije, 2013: A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences*, **17,** 1893-1912, doi: 10.5194/hess-17-1893-2013.

Famiglietti, J. S., and E. F. Wood, 1994: Multiscale modeling of spatially variable water and energy balance processes. *Water Resources Research*, **30,** 3061-3078, doi: 10.1029/94wr01498.

Fenicia, F., D. Kavetski, and H. H. G. Savenije, 2011: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, **47**, doi: W11510, 10.1029/2010wr010174.

Fenicia, F., D. Kavetski, H. H. G. Savenije, M. P. Clark, G. Schoups, L. Pfister, and J. Freer, 2014: Catchment properties, function, and conceptual model representation: is there a correspondence? *Hydrological Processes*, **28,** 2451-2467, doi: 10.1002/hyp.9726.

Freer, J., J. J. McDonnell, K. J. Beven, N. E. Peters, D. A. Burns, R. P. Hooper, B. Aulenbach, and C. Kendall, 2002: The role of bedrock topography on subsurface storm flow. *Water Resources Research*, **38**, doi: 10.1029/2001wr000872.

Freer, J. E., H. McMillan, J. J. McDonnell, and K. J. Beven, 2004: Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. *Journal of Hydrology*, **291,** 254-277, doi: 10.1016/j.jhydrol.2003.12.037.

Freeze, R. A., and R. Harlan, 1969: Blueprint for a physically-based, digitally-simulated hydrologic response model. *Journal of Hydrology*, **9,** 237-258, doi: 10.1016/0022-1694(69)90020-1.

Garland, T., 2015: The scientific method as an ongoing process. *Presentation available at* http://idea.ucr.edu/documents/flash/scientific_method/story.htm, **accessed 20 July 2015**
Gupta, H., C. Perrin, G. Bloschl, A. Montanari, R. Kumar, M. Clark, and V. Andréassian, 2014: Large-sample hydrology: a need to balance depth with breadth. *Hydrology and Earth System Sciences*, **18,** p. 463-p. 477
Gupta, H. V., T. Wagener, and Y. Q. Liu, 2008: Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes*, **22,** 3802-3813, doi: 10.1002/hyp.6989.
Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye, 2012: Towards a comprehensive assessment of model structural adequacy. *Water Resources Research*, **48**, doi: 10.1029/2011wr011044.
Gupta, H. V., and G. S. Nearing, 2014: Debates—The future of hydrological sciences: A (common) path forward? Using models and data to learn: A systems theoretic perspective on the future of hydrological science. *Water Resources Research*, **50,** 5351-5359
Harman, C., and M. Sivapalan, 2009: Effects of hydraulic conductivity variability on hillslope‐scale shallow subsurface flow response and storage‐discharge relations. *Water resources research*, **45**
Harman, C., M. Sivapalan, and P. Kumar, 2009: Power law catchment‐scale recessions arising from heterogeneous linear small‐scale dynamics. *Water resources research*, **45**
Harman, C., and P. Troch, 2014: What makes Darwinian hydrology" Darwinian"? Asking a different kind of question about landscapes. *Hydrology and Earth System Sciences*, **18,** 417-433
Hawkins, R., and T. Cundy, 1987: STEADY‐STATE ANALYSIS OF INFILTRATION AND OVERLAND FLOW FOR SPATIALLY‐VARIED HILLSLOPES1. *JAWRA Journal of the American Water Resources Association*, **23,** 251-256
Hock, R., 2003: Temperature index melt modelling in mountain areas. *Journal of Hydrology*, **282,** 104-115
Hurrell, J. W., M. Holland, P. Gent, S. Ghan, J. E. Kay, P. Kushner, J.-F. Lamarque, W. Large, D. Lawrence, and K. Lindsay, 2013: The community earth system model: a framework for collaborative research. *Bulletin of the American Meteorological Society*, **94,** 1339-1360
Huss, M., M. Funk, and A. Ohmura, 2009: Strong Alpine glacier melt in the 1940s due to enhanced solar radiation. *Geophysical Research Letters*, **36**, doi: 10.1029/2009GL040789.
Huxley, T. H., 1894: *Collected Essays: Darwiniana.* Vol. 2, Appleton.
Jakeman, A., and G. Hornberger, 1993: How much complexity is warranted in a rainfall‐runoff model? *Water Resources Research*, **29,** 2637-2649
Jencso, K. G., B. L. McGlynn, M. N. Gooseff, S. M. Wondzell, K. E. Bencala, and L. A. Marshall, 2009: Hydrologic connectivity between landscapes and streams: Transferring reach‐and plot‐scale understanding to the catchment scale. *Water Resources Research*, **45**
Jencso, K. G., and B. L. McGlynn, 2011: Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation. *Water Resources Research*, **47**
Kirchner, J., 2006: Getting the right answers for the wrong reasons. *Water Resources Research*, **42**, doi: 10.1029/2005wr004362.
Kirchner, J. W., R. P. Hooper, C. Kendall, C. Neal, and G. Leavesley, 1996: Testing and validating environmental models. *Science of the Total Environment*, **183,** 33-47
Kleinhans, M. G., C. J. Buskes, and H. W. de Regt, 2005: Terra Incognita: explanation and reduction in earth science. *International Studies in the Philosophy of Science*, **19,** 289-317
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 Resources Research*, **46**, doi: W04201, 10.1029/2009wr008730.
Koren, V., J. Schaake, K. Mitchell, Q. Y. Duan, F. Chen, and J. Baker, 1999: A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **104,** 19569-19585, doi: 10.1029/1999jd900232.

Koster, R. D., and M. J. Suarez, 1992: Modeling the land surface boundary in climate models as a composite of independent vegetation stands. *Journal of Geophysical Research-Atmospheres*, **97,** 2697-2715, doi: 10.1029/91jd01696.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model 1. Model structure. *Journal of Geophysical Research-Atmospheres*, **105,** 24809-24822, doi: 10.1029/2000jd900327.

Kraft, P., K. B. Vaché, H.-G. Frede, and L. Breuer, 2011: CMF: a hydrological programming language extension for integrated catchment models. *Environmental Modelling & Software*, **26,** 828-830, doi: 10.1016/j.envsoft.2010.12.009.

Kumar, R., L. Samaniego, and S. Attinger, 2013: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resources Research*, **49,** 360-379, doi: 10.1029/2012wr012195.

Kumar, S. V., C. D. Peters-Lidard, J. Santanello, K. Harrison, Y. Liu, and M. Shaw, 2012: Land surface Verification Toolkit (LVT) - a generalized framework for land surface model evaluation. *Geoscientific Model Development*, **5,** 869-886, doi: 10.5194/gmd-5-869-2012.

Kurylyk, B. L., K. T. MacQuarrie, and C. I. Voss, 2014: Climate change impacts on the temperature and magnitude of groundwater discharge from shallow, unconfined aquifers. *Water Resources Research*, **50,** 3253-3274

Lawrence, D. M., K. W. Oleson, M. G. Flanner, P. E. Thornton, S. C. Swenson, P. J. Lawrence, X. B. Zeng, Z. L. Yang, S. Levis, K. Sakaguchi, G. B. Bonan, and A. G. Slater, 2011: Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, **3**, doi: M03001, 10.1029/2011ms000045.

Leach, J., and R. Moore, 2015: Observations and modeling of hillslope throughflow temperatures in a coastal forested catchment. *Water Resources Research*

Lehmann, P., C. Hinz, G. McGrath, H. Tromp-van Meerveld, and J. McDonnell, 2007: Rainfall threshold for hillslope outflow: an emergent property of flow pathway connectivity. *Hydrology and Earth System Sciences Discussions*, **11,** 1047-1063

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land-surface water and energy fluxes for General Circulation Models. *Journal of Geophysical Research-Atmospheres*, **99,** 14415-14428, doi: 10.1029/94jd00483.

Liang, X., D. P. Lettenmaier, and E. F. Wood, 1996: One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model. *Journal of Geophysical Research-Atmospheres*, **101,** 21403-21422, doi: 10.1029/96jd01448.

Luce, C. H., D. G. Tarboton, and K. R. Cooley, 1999: Sub-grid parameterization of snow distribution for an energy and mass balance snow cover model. *Hydrological Processes*, **13,** 1921-1933, doi: 10.1002/(sici)1099-1085(199909)13:12/13<1921::aid-hyp867>3.3.co;2-j.

Mahrt, L., 1987: Grid-averaged surface fluxes. *Monthly Weather Review*, **115,** 1550-1560, doi: 10.1175/1520-0493(1987)115<1550:gasf>2.0.co;2.

Maxwell, R. M., and S. J. Kollet, 2008: Quantifying the effects of three-dimensional subsurface heterogeneity on Hortonian runoff processes using a coupled numerical, stochastic approach. *Advances in Water Resources*, **31,** 807-817, doi: 10.1016/j.advwatres.2008.01.020.

McDonnell, J., M. Sivapalan, K. Vaché, S. Dunn, G. Grant, R. Haggerty, C. Hinz, R. Hooper, J. Kirchner, and M. Roderick, 2007: Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research*, **43**, doi: 10.1029/2006wr005467.

McDonnell, J. J., 2013: Are all runoff processes the same? *Hydrological Processes*, **27,** 4103-4111

McMillan, H., D. Tetzlaff, M. Clark, and C. Soulsby, 2012: Do time - variable tracers aid the evaluation of hydrological model structure? A multimodel approach. *Water Resources Research*, **48**

Mendoza, P., M. Clark, M. Barlage, B. Rajagopalan, L. Samaniego, G. Abramowitz, and H. V. Gupta, 2015: Are we unnecessarily constraining the agility of complex process-based models? *Water Resources Research*, doi: 10.1002/2014WR015820.

Milly, P., and K. A. Dunne, 2011: On the hydrologic adjustment of climate-model projections: The potential pitfall of potential evapotranspiration. *Earth Interactions*, **15,** 1-14

Mizukami, N., M. P. Clark, E. Gutmann, P. A. Mendoza, A. J. Newman, B. Nijssen, B. Livneh, L. Hay, J. R. Arnold, and L. Brekke, 2015: Implications of the methodological choices for hydrologic portrayals of climate change over the Contiguous United States: statistically downscaled forcing data and hydrologic models. *Journal of Hydrometeorology*, doi: 10.1175/JHM-D-14-0187.1

Moore, R., and R. Clarke, 1981: A distribution function approach to rainfall runoff modeling. *Water Resources Research*, **17,** 1367-1382, doi: 10.1029/WR017i005p01367.

Nearing, G. S., and H. V. Gupta, 2014: The quantity and quality of information in hydrologic models. *Water Resources Research***,** n/a-n/a, doi: 10.1002/2014wr015895.

Nijssen, B., G. M. O'Donnell, D. P. Lettenmaier, D. Lohmann, and E. F. Wood, 2001: Predicting the discharge of global rivers. *Journal of Climate*, **14,** 3307-3323, doi: 10.1175/1520-0442(2001)014<3307:ptdogr>2.0.co;2.

Niu, G. Y., Z. L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, and Y. L. Xia, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research-Atmospheres*, **116**, doi: D12109, 10.1029/2010jd015139.

Peters, N. E., J. Freer, and K. Beven, 2003: Modelling hydrologic responses in a small forested catchment (Panola Mountain, Georgia, USA): a comparison of the original and a new dynamic TOPMODEL. *Hydrological Processes*, **17,** 345-362, doi: 10.1002/hyp.1128.

Popper, K. R., 1959: The logic of scientific discovery. *London: Hutchinson*

Power, M. E., A. Sun, G. Parker, W. E. Dietrich, and J. T. Wootton, 1995: Hydraulic food-chain models. *BioScience***,** 159-167

Rakovec, O., R. Kumar, J. Mai, M. Cuntz, S. Thober, M. Zink, S. Attinger, C. Schafer, M. Schron, and L. Samaniego, 2015: Multiscale and multivariate evaluation of water fluxes and states over European river basins. *Journal of Hydrometeorology*, doi: doi:10.1175/JHM-D-15-0054.1.

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. *Advances in Water Resources*, **22,** 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. *Advances in Water Resources*, **23,** 15-39, doi: 10.1016/s0309-1708(99)00005-6.

Reggiani, P., and J. Schellekens, 2003: Modelling of hydrological responses: the representative elementary watershed approach as an alternative blueprint for watershed modelling. *Hydrological Processes*, **17,** 3785-3789, doi: 10.1002/hyp.5167.

Rigon, R., A. Rinaldo, I. Rodriguez ‐ Iturbe, R. L. Bras, and E. Ijjasz ‐ Vasquez, 1993: Optimal channel networks: a framework for the study of river basin morphology. *Water Resources Research*, **29,** 1635-1646

Rinaldo, A., G. Botter, E. Bertuzzo, A. Uccelli, T. Settin, and M. Marani, 2006: Transport at basin scales: 1. Theoretical framework. *Hydrology and Earth System Sciences*, **10,** 19-29

Rodríguez ‐ Iturbe, I., A. Rinaldo, R. Rigon, R. L. Bras, A. Marani, and E. Ijjász ‐ Vásquez, 1992: Energy dissipation, runoff production, and the three ‐ dimensional structure of river basins. *Water Resources Research*, **28,** 1095-1103

Samaniego, L., R. Kumar, and S. Attinger, 2010: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, **46**, doi: W05523, 10.1029/2008wr007327.

Schaefli, B., C. Harman, M. Sivapalan, and S. Schymanski, 2011: HESS Opinions: Hydrologic predictions in a changing environment: behavioral modeling. *Hydrology and Earth System Sciences*, **15,** 635-646

Schaefli, B., L. Nicótina, C. Imfeld, P. Da Ronco, E. Bertuzzo, and A. Rinaldo, 2014: SEHR-ECHO v1. 0: a Spatially Explicit Hydrologic Response model for ecohydrologic applications. *Geoscientific Model Development*, **7,** 2733-2746

Schewe, J., and Coauthors, 2014: Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **111,** 3245-3250, doi: 10.1073/pnas.1222460110.

Schymanski, S., M. Sivapalan, M. Roderick, J. Beringer, and L. Hutley, 2008: An optimality-based model of the coupled soil moisture and root dynamics. *Hydrology and Earth System Sciences Discussions*, **12,** 913-932

Schymanski, S., M. Sivapalan, M. Roderick, L. Hutley, and J. Beringer, 2009: An optimality‐based model of the dynamic feedbacks between natural vegetation and the water balance. *Water Resources Research*, **45**

Schymanski, S. J., A. Kleidon, M. Stieglitz, and J. Narula, 2010: Maximum entropy production allows a simple representation of heterogeneity in semiarid ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365,** 1449-1455

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 Resources Research*, **38,** 23-21-23-14

Seibert, J., 2003: Reliability of model predictions outside calibration conditions. *Nordic Hydrology*, **34,** 477-492

Seibert, J., and J. McDonnell, 2013: Gauging the Ungauged Basin: The Relative Value of Soft and Hard Data. *Journal of Hydrologic Engineering***,** na, doi: 10.1061/(ASCE)HE.1943-5584.0000861.

Seyfried, M. S., L. E. Grant, D. Marks, A. Winstral, and J. McNamara, 2009: Simulated soil water storage effects on streamflow generation in a mountainous snowmelt environment, Idaho, USA. *Hydrological Processes*, **23,** 858-873, doi: 10.1002/hyp.7211.

Sheffield, J., E. F. Wood, and M. L. Roderick, 2012: Little change in global drought over the past 60 years. *Nature*, **491,** 435-438

Shook, K., J. W. Pomeroy, C. Spence, and L. Boychuk, 2013: Storage dynamics simulations in prairie wetland hydrology models: evaluation and parameterization. *Hydrological Processes*, **27,** 1875-1889, doi: 10.1002/hyp.9867.

Sivapalan, M., 2005: Pattern, process and function: elements of a unified theory of hydrology at the catchment scale. *Encyclopedia of Hydrological Sciences*

Slater, A. G., and Coauthors, 2001: The representation of snow in land surface schemes: Results from PILPS 2(d). *Journal of Hydrometeorology*, **2,** 7-25, doi: 10.1175/1525-7541(2001)002<0007:trosil>2.0.co;2.

Smith, M., and Coauthors, 2013: The distributed model intercomparison project - Phase 2: Experiment design and summary results of the western basin experiments. *Journal of Hydrology*, **507,** 300-329, doi: 10.1016/j.jhydrol.2013.08.040.

Spence, C., X. Guan, R. Phillips, N. Hedstrom, R. Granger, and B. Reid, 2010: Storage dynamics and streamflow in a catchment with a variable contributing area. *Hydrological Processes*, **24,** 2209-2221

Tague, C., and L. Band, 2004: RHESSys: regional hydro-ecologic simulation system-an object-oriented approach to spatially distributed modeling of carbon, water, and nutrient cycling. *Earth Interactions*, **8,** 1-42, doi: 10.1175/1087-3562(2004)8<1:RRHSSO>2.0.CO;2.

Troch, P. A., G. A. Carrillo, I. Heidbüchel, S. Rajagopal, M. Switanek, T. H. Volkmann, and M. Yaeger, 2009: Dealing with landscape heterogeneity in watershed hydrology: A review of recent progress toward new hydrological theory. *Geography Compass*, **3,** 375-392, doi: 10.1111/j.1749-8198.2008.00186.x.

Tromp‐van Meerveld, H., and J. McDonnell, 2006: Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research*, **42**

Tyler, S. W., J. S. Selker, M. B. Hausner, C. E. Hatch, T. Torgersen, C. E. Thodal, and S. G. Schladow, 2009: Environmental temperature sensing using Raman spectra DTS fiber‐optic methods. *Water Resources Research*, **45**

Vivoni, E. R., V. Teles, V. Y. Ivanov, R. L. Bras, and D. Entekhabi, 2005: Embedding landscape processes into triangulated terrain models. *International Journal of Geographical Information Science*, **19,** 429-457, doi: 10.1080/13658810512331325111.

Wagener, T., M. Sivapalan, P. A. Troch, B. L. McGlynn, C. J. Harman, H. V. Gupta, P. Kumar, P. S. C. Rao, N. B. Basu, and J. S. Wilson, 2010: The future of hydrology: An evolving science for a changing world. *Water Resources Research*, **46**, doi: 10.1029/2009wr008906.

Weiler, M., and J. McDonnell, 2004: Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology. *Journal of Hydrology*, **285,** 3-18

Weiler, M., and K. Beven, 2015: Do we need a Community Hydrological Model? *Water Resources Research*, doi: 10.1002/2014WR016731.

Wenger, S. J., C. H. Luce, A. F. Hamlet, D. J. Isaak, and H. M. Neville, 2010: Macroscale hydrologic modeling of ecologically relevant flow metrics. *Water Resources Research*, **46**

Wigmosta, M. S., L. W. Vail, and D. P. Lettenmaier, 1994: A distributed hydrology-vegetation model for complex terrain. *Water Resources Research*, **30,** 1665-1679, doi: 10.1029/94wr00436.

Winsemius, H., B. Schaefli, A. Montanari, and H. H. G. Savenije, 2009: On the calibration of hydrological models in ungauged basins: a framework for integrating hard and soft hydrological information. *Water Resources Research*, **45,** W12422, doi: 10.1029/2009WR007706.

Wood, E. F., D. P. Lettenmaier, and V. G. Zartarian, 1992: A land surface hydrology parameterization with sub grid variability for general circulation models. *Journal of Geophysical Research-Atmospheres*, **97,** 2717-2728, doi: 10.1029/91jd01786.

Wrede, S., F. Fenicia, N. Martínez - Carreras, J. Juilleret, C. Hissler, A. Krein, H. H. Savenije, S. Uhlenbrook, D. Kavetski, and L. Pfister, 2014: Towards more systematic perceptual model development: a case study using 3 Luxembourgish catchments. *Hydrological Processes*

Zehe, E., H. Lee, and M. Sivapalan, 2006: Dynamical process upscaling for deriving catchment scale state variables and constitutive relations for meso-scale process models. *Hydrology and Earth System Sciences*, **10,** 981-996, doi: 10.5194/hess-10-981-2006.

Zehe, E., and M. Sivapalan, 2009: Threshold behaviour in hydrological systems as (human) geo-ecosystems: manifestations, controls, implications. *Hydrology and Earth System Sciences*, **13,** 1273-1297

Zehe, E., T. Blume, A. Kleidon, U. Ehret, U. Scherer, and M. Westhoff, 2013: A thermodynamic approach to link self-organization, preferential flow and rainfall–runoff behaviour. *Hydrology and Earth System Sciences*, **17,** 4297-4322

Zehe, E., U. Ehret, L. Pfister, T. Blume, B. Schröder, M. Westhoff, C. Jackisch, S. Schymanski, M. Weiler, and K. Schulz, 2014: HESS Opinions: Functional units: a novel framework to explore the link between spatial organization and hydrological functioning of intermediate scale catchments. *Hydrology and Earth System Sciences Discussions*, **11,** 3249-3313

Zreda, M., W. Shuttleworth, X. Zeng, C. Zweck, D. Desilets, T. Franz, and R. Rosolem, 2012: COSMOS: the cosmic-ray soil moisture observing system. *Hydrology and Earth System Sciences*, **16,** 4079-4099

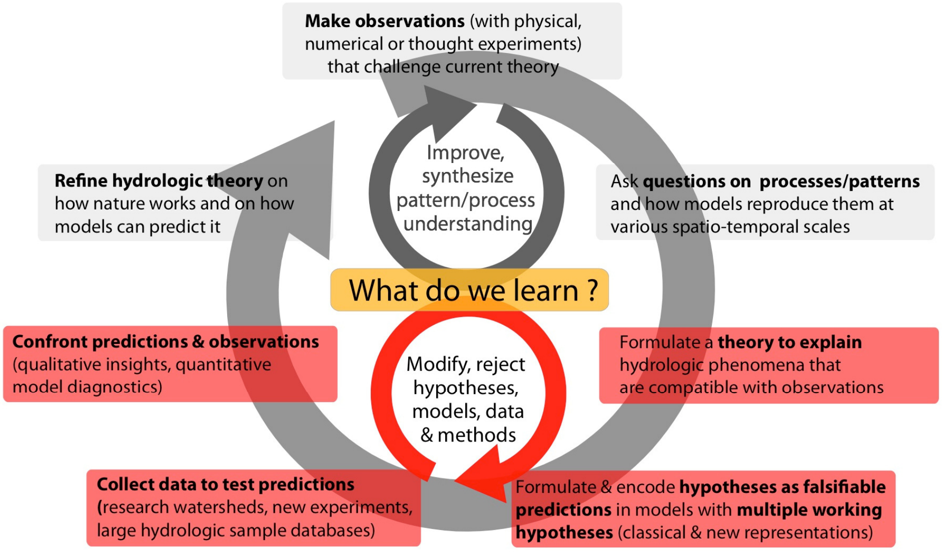

Figure 1. A theoretically grounded approach to hydrologic model development, following the scientific method as defined by Popper [1959]. This graphic is inspired by Garland [2015].