

Water Resources Research®

COMMENTARY

10.1029/2022WR034234

Key Points:

- New thinking is needed to improve the predictive skill of conceptual hydrological models as they are confronted by multi-faceted change
- Next generation conceptual models need to increase their “hydrological fidelity” founded upon multiple hydrological lines of evidence
- Virtual Hydrological Laboratories accelerate this development through controlled testing for future changes yet to be observed

Correspondence to:

M. Thyer,
mark.thyer@adelaide.edu.au












Citation:

Thyer, M., Gupta, H., Westra, S., McInerney, D., Maier, H. R., Kavetski, D., et al. (2024). Virtual Hydrological Laboratories: Developing the next generation of conceptual models to support decision making under change. *Water Resources Research*, 60, e2022WR034234. <https://doi.org/10.1029/2022WR034234>

Received 18 OCT 2023

Accepted 7 FEB 2024

Virtual Hydrological Laboratories: Developing the Next Generation of Conceptual Models to Support Decision Making Under Change

Mark Thyer¹ , Hoshin Gupta², Seth Westra¹ , David McInerney¹ , Holger R. Maier¹ , Dmitri Kavetski¹ , Anthony Jakeman³ , Barry Croke³ , Craig Simmons⁴ , Daniel Partington⁵ , Margaret Shanfield⁵ , and Christina Tague⁶ 

¹School of Architecture and Civil Engineering, University of Adelaide, Adelaide, SA, Australia, ²Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA, ³Mathematical Sciences Institute and Fenner School of Environment and Society, The Australian National University, Canberra, ACT, Australia, ⁴Engineering, Science and Environment, University of Newcastle, Newcastle, NSW, Australia, ⁵National Centre for Groundwater Research and Training, College of Science and Engineering, Flinders University, Adelaide, SA, Australia, ⁶Bren School of Environmental Science and Management, University of California, Santa Barbara, Santa Barbara, CA, USA

Abstract As hydrological systems are pushed outside the envelope of historical experience, the ability of current hydrological models to serve as a basis for credible prediction and decision making is increasingly challenged. Conceptual models are the most common type of surface water hydrological model used for decision support due to reasonable performance in the absence of change, ease of use and computational speed that facilitate scenario, sensitivity and uncertainty analysis. Hence, conceptual models in effect represent the current “shopfront” of hydrological science as seen by practitioners. However, these models have notable limitations in their ability to resolve internal catchment processes and subsequently capture hydrological change. New thinking is needed to confront the challenges faced by the current generation of conceptual models in dealing with a changing environment. We argue the next generation of conceptual models should combine the parsimony of conceptual models with our best available scientific understanding. We propose a strategy to develop such models using multiple hydrological lines of evidence. This strategy includes using appropriately selected physically resolved models as “Virtual Hydrological Laboratories” to test and refine the simpler models’ ability to predict future hydrological changes. This approach moves beyond the sole focus on “predictive skill” measured using metrics of historical performance, facilitating the development of the next generation of conceptual models with hydrological fidelity (i.e., models that “get the right answers for the right reasons”). This quest is more than a scientific curiosity; it is expected by policy makers who need to know what to plan for.

1. Introduction

“...the sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work—that is, correctly describe phenomena from a reasonably wide area. Furthermore, it must satisfy certain aesthetic criteria—that is, in relation to how much it describes, it must be rather simple.” (von Neumann, 1955)

1.1. Hydrological Science Is Limited by the Inability to Perform Controlled Experiments

Catchments represent the outcome of physical, biological and social processes that play out across multiple spatial and temporal scales (Gunderson & Holling, 2002). Whilst hydrological systems globally are subject to the same fundamental underlying physical (if not necessarily biological or social) processes, their diverse local expressions mean that each catchment is generally regarded as unique (Beven, 2000). Moreover, observational limitations mean that it is impossible to fully characterize internal functioning due to poor identifiability of the system (e.g., Guillaume et al., 2019), which is often referred to as equifinality in hydrology (Beven, 2006) or under-determination in the philosophy of science (Rosenberg & McIntyre, 2019). Moreover, catchment responses are both highly non-linear (with small changes potentially having large consequences) and constantly changing at a

© 2024. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

range of spatial and temporal scales (Fowler et al., 2022; Montanari et al., 2013) leading to complex causal chains with multiple interactions and feedbacks.

These statements form the backdrop for the hydrological science enterprise, differentiating it from fields that can rely heavily on controlled laboratory experiments (e.g., physical sciences), and from fields where randomized control trials are viewed as the highest line of evidence (e.g., medical sciences). We simply do not have access to multiple identical (or even functionally equivalent) realizations of a hydrological system (such as a catchment) for which we can alter one or more physical characteristics and/or boundary conditions in one set while leaving the other set as controls. Likewise, for practical and ethical reasons, we cannot randomly select a statistical sample of catchments to modify (e.g., construct dams, modify land use, or some other sort of intervention) while using an equivalent statistical sample as a control. Hydrological studies that do involve the use of controlled experiments generally occur at the laboratory or small plot scale and focus on understanding component processes at scales that do not easily or directly translate to the scales relevant to water management.

1.2. Progress in Hydrological Science Can Be Considered a Qualified Success

Despite its challenges, the hydrological sciences have seen tremendous progress during the last century and a half, and our capacity to predict hydrological fluxes at a range of spatial and temporal scales is unprecedented. A key feature of the scientific discipline is a heavy reliance on models—where models are the main expression of hydrological theory and represent complex sets of hypotheses (always multiple), which can be tested on data from those elements of the hydrological system that lend themselves to observation. In keeping with common interpretations of the scientific method, we should expect such models (representing competing hypotheses) to undergo a type of scientific natural selection—gradually weeding out the weaker potential explanations and retaining the stronger ones—by confronting them with more and different kinds of data.

However, taking the pragmatic perspective that the quality of a scientific endeavor is ultimately measured through its predictive skill under varied conditions, we might call this partnership a *qualified* success. While our capacity to predict dominant hydrological quantities generally continues to increase, our capacity to verify the proposed models remains limited by inadequate access to relevant data (e.g., Maier et al., 2023), and by the unavoidable inability to conduct controlled experiments. Put simply, our capacity to test and refine hydrological models at the catchment scale, for a wide range of catchments, is severely limited.

1.3. Simpler Conceptual Models Are More Commonly Used in Practical Contexts

If we accept the assertion that hydrological science is largely the discipline of building and testing models of hydrological processes, then what do hydrological models consist of? There indeed exists a wide spectrum of hydrological models, with varying degrees of simplification in space, time, and process representation (Peel & McMahon, 2020). For the purposes of this discussion, we (somewhat crudely) categorize hydrological models into two end-member model types—“conceptual models” (CMs) and “physically resolved models” (PRMs). The first type, CMs, refers to simplified, lumped or semi-distributed, hydrological representations based on an aggregate perceptual-conceptual physical understanding of the dominant processes that give rise to the catchment-scale behaviors we observe under typical climatic conditions (Gupta et al., 2012). Classic examples of CMs include GR4J (Perrin et al., 2003), HBV (Bergström, 1995) and IHACRES (Croke & Jakeman, 2004). In contrast, the second model type, PRMs, refers to space-time-and-process-resolved spatially discretized hydrological representations based on detailed physics-based interpretation of subcatchment scale storages, flows and physical transformations of water within the system. This physics-based understanding is typically obtained from detailed laboratory/plot scale and experimental catchment studies. Examples of PRMs include HydroGeoSphere (Brunner & Simmons, 2012), HM (VanderKwaak, 1999), and ParFlow (Kollet & Maxwell, 2006). We prefer to use the term “physically resolved” here, instead of “physically based” and/or “process-based,” because the real difference between conceptual and physically resolved models is not that the former have no basis in physical reality (they clearly do, in that attributes of reality, such as the processes of catchment storage, routing and ET fluxes are mirrored onto model attributes), but that the essential difference between them is the level of detail with (and form of representation by) which physical attributes of space, time and hydrologically relevant processes are *resolved* within them.

In categorizing hydrological models into these two types, we recognize several drawbacks (as outlined by Partington et al., 2022, and others), including: (a) there have been extensive debates in the hydrological literature on the relative merits of categorizing models (e.g., Abbott et al., 2001; Beven, 1989; Fatichi et al., 2016; Grayson et al., 1992 and many others); (b) definitions are not “sharp” but form a continuum (e.g., Mount et al., 2016); (c) there are always exceptions to the model types summarized in the preceding paragraph; (d) this categorization applies to models which predominantly focus on predicting surface-water quantities rather than groundwater quantities—noting of course there can be significant groundwater-surface water interactions.

Of these two model types, CMs are generally the more common choice for practical management applications. This includes eWater SOURCE, Australia’s hydrological modeling platform (Welsh et al., 2013), the Sacramento model in the US (Burnash et al., 1973), and GR4J in France (Perrin et al., 2003). A recent review of hydrological models in Switzerland by Horton et al. (2022) (based on 157 journal articles) found that simpler conceptual models (PREVAH and HBV-light) dominate usage with over 45% of applications, and are the most widely used in climate change impact studies. Similarly, Peel and McMahon (2020), who reviewed 279 rainfall-runoff models, found that 74% of models were conceptual.

One can postulate likely reasons for the popularity of CMs over PRMs. In terms of performance, model inter-comparison studies have shown that CMs provide similar or better streamflow predictions than PRMs (e.g., Reed et al., 2004; Refsgaard & Knudsen, 1996). Another obvious advantage of CMs over PRMs is their ease of use and low computational cost, which makes it possible to explore many hundreds to thousands of scenarios; thereby supporting evaluation of risk when making decisions to manage the potential impacts of climate variability and change (e.g., Partington et al., 2022). Faster computational speed also permits both sensitivity (Razavi et al., 2021) and comprehensive uncertainty analysis (Renard et al., 2010)—typically requiring parameter samples of thousands or more for convergence (the number depending on complexity of the model parameter space). There is also a strong social component to hydrological model selection, with Addor and Melsen (2019) referring to “practicality, convenience, experience, and habit” as key factors, and Horton et al. (2022) establishing, via surveys of model developers and users in Switzerland, that institutional knowledge is a key factor in model choice.

1.4. Current Conceptual Models Are Unlikely to Be Able to Deal With Hydrological Change

Recent signs suggest that increasingly rapid hydrological change will have profound implications, not just for our modeling efforts, but for hydrological science in general. Natural and anthropogenic forces act to alter water balances, vegetation dynamics, soil porosities and longer-term geomorphologies, and these changes interact to drive further change along complex, interconnected and multi-scale causal chains (Section 2.1 will provide a more detailed description of the types of “hydrological change”). It has become increasingly difficult to accept that these changes can be adequately represented using highly conceptualized model abstractions, whose behaviors are controlled by static parameters that cannot easily be related to observed geometrical and physical properties of the changing system. This is not mere speculation—many recent studies have highlighted the challenges that conventional CMs face when attempting to represent the catchment-scale changes that are taking place (e.g., Clarke, 2007; Fowler et al., 2020; Gibbs et al., 2018; Saft et al., 2016).

Despite their popularity, CMs are known to often produce unrealistic internal hydrologic fluxes (Bouaziz et al., 2021; L. Li et al., 2015; Peel & McMahon, 2020), as these models do not provide detailed “physically resolved” representations of internal catchment processes. In addition, the reliance of these models on calibration to tune parameter values can result in parameter interactions that mask structural inadequacies. Observational limitations make it impossible to properly assess whether the “accuracy” of a CM’s predictions is due to the correct partitioning of fluxes and proper assignment of state variables within the model (see Partington et al., 2013). Indeed, it is a major conceptual leap to view catchments as series and parallel arrangements of large “buckets” influenced by aggregated processes, and to routinely assume that individual physical processes such as groundwater/surface interactions, capillary drive, preferential pathways, and spatial patterns of ponding or re-infiltration can be ignored at catchment scales and under operational conditions.

Put simply, it is difficult to know whether the CMs are “right for the right reasons.” This concept, known as hydrological fidelity, describes the extent to which a model faithfully simulates dominant hydrological processes (Clark et al., 2015). Unsurprisingly, in order for CMs to represent hydrological change, their internal process representations will likely need greater hydrological fidelity and more direct connection to physical catchment attributes (e.g., see Duethmann et al., 2020, and case study therein).

1.5. Will PRMs Be the Savior to Overcome the Challenge of Predicting Change?

In contrast to CMs, PRMs provide a physically resolved representation of a catchment. These models are typically derived from laboratory/plot scale studies and experimental catchments; their mathematical representation typically employs partial differential equations derived from continuity of mass and momentum in combination with initial and boundary conditions (Freeze & Harlan, 1969). These models are appealing because they aim to capture diverse catchment processes at scales that are far more detailed than those used in CMs. As noted by Fatichi et al. (2016), “when applied spatially, from hillslope to continental scales, such a model [PRM] can incorporate the space–time variability of the primary forcings, such as precipitation and radiation, and variations of land-surface properties (e.g., topography, soils, vegetation) at the sub-hillslope scale, while resolving the subsurface domain in horizontal and vertical directions in a way to describe catchment heterogeneity.” As such, PRMs provide the best opportunity to capture internal catchment processes (Peel & McMahon, 2020). In turn, as stated by Fatichi et al. (2016), it is believed this helps provide predictions under hydrological change, such as changes in land use/land cover or a non-stationary climate (e.g., Pierini et al., 2014; van Roosmalen et al., 2009; Wang et al., 2008).

The challenges with using PRMs in practical contexts are twofold: (a) observational constraints limit the ability to establish reliable model constitutive functions and may result in over-parametrization, providing a risk to model accuracy, and (b) there is a high cost related to their practical implementation.

In regard to the first challenge, hydrological systems exist in the so-called realm of “organized complexity” (Dooge, 1986; Weinberg, 2001), where it is difficult to (a) acquire detailed process-relevant data at space-time resolutions required to study the fundamental scale-dependent nature of hydrologically relevant processes, (b) develop and verify accurate mathematical representations of those processes at those various scales, and (c) constrain the values of the extremely large number of parameters that must be specified when applying PRMs to common real-world catchments. This makes it difficult to test the fidelity of PRMs, except for a relatively small number of “data-rich” experimental catchments (e.g., Camporese et al., 2014, 2015; Thyer et al., 2004). Thus, calibration of PRMs is susceptible to over-parameterization problems. It is generally recognized that only a small number of parameters can be identified from typically available catchment data; see Jakeman and Hornberger (1993) and Shin et al. (2015) for further discussion.

In regard to the second challenge, PRMs have a significant cost associated with model development, as field observations, data preparation and parameter calibrations are very expensive (Ampadu et al., 2013) and often unavailable at the intended model scale. The computational costs of PRMs often preclude comprehensive scenario, sensitivity and/or uncertainty analysis. The computational challenges are being gradually overcome by massively parallel super-computers (Kollet et al., 2010) (and to some extent by model emulation), though such resources are not readily available to the majority of hydrological modelers.

These considerations, in our view, suggest that PRMs, while being attractive in terms of potentially more appealing process presentation, are on their own unlikely to provide a reliable and practical basis for representing environmental changes across a wide range of scenarios and locations.

1.6. The Need for a New Approach to Conceptual Models to Represent Change

The evidence presented in Section 1.4 suggests that CMs are the most common type of model used for practical applications by hydrological model users (e.g., engineers/planners/designers/operators of water resource systems) for decision support. In effect, CMs and their predictions represent the “shopfront” of hydrological science. Given the pressing and profound implications of hydrological change, it is critical that the hydrological community develop models with a higher hydrological fidelity than current CMs.

This does not mean that we should reject CMs in favor of PRMs, or vice versa. It is unlikely that the challenges associated with PRMs related to observational limitations, practical implementation and institutional convenience will be alleviated any time soon, hence there is unlikely to be widespread adoption of PRMs in practical contexts in the foreseeable future. Similarly, it seems wasteful for the development of CMs to not take advantage of the tremendous investment and hydrological knowledge generated by PRMs. We need to find a “sweet spot” where we can take advantage of strengths and overcome the weaknesses of both model types.

Key questions: “Can the HLE evaluate the model for robustness...”	Hydrological Lines of Evidence (HLEs)				
	Lab/Plot Scale Experiments	Experimental Catchments	Paired Catchments	Large Sample Hydrology	Emerging HLE: Virtual Hydrological Laboratories
1a. To hydrological changes due to climatic phenomena?	Controlled Experimental	Observational	Observational	Observational	Controlled Experimental
1b. To hydrological changes due to human intervention?	Controlled Experimental	Controlled Experimental	Controlled Experimental	Observational	Controlled Experimental
2. On practically-relevant catchment sizes?	No	Limited	Yes	Yes	Yes
3. On diverse range of catchment properties?	Yes	Limited	Limited	Yes	Yes
4. For a wide range of processes/predictions?	Yes	Yes	No	No	Yes
5. Using real data?	Yes	Yes	Yes	Yes	No

Figure 1. Evaluation of hydrological lines of evidence against five key questions to accelerate the development of the next generation of conceptual models. The questions are listed as row headings on left hand side of the figure (see description in Section 2.2). The four existing HLEs are listed as column headings (light blue background). The “cells” in the figure provide the outcomes of the evaluation (Section 2.3 for further information). For question 1, regarding “hydrological change,” we assess whether the HLEs can evaluate the model for robustness to hydrological change using either a controlled experimental or observational approach. For questions 2–5 the HLEs are classified into three categories (“yes”, “no”, “limited”), based on the authors’ experience. The right hand column (dark blue background) evaluates an emerging HLE, Virtual Hydrological Laboratories (see Section 3 for detailed discussion).

It is time to re-think how we assemble our hydrological lines of evidence when constructing CMs so that our evolving best available understanding can be used to provide the next generation of CMs with better hydrological fidelity, and therefore greater confidence in their predictions in the face of hydrological change.

2. Accelerating Development of the Next Generation of Conceptual Models

As an empirical science, hydrology can be characterized as the process of developing hypotheses (through models) and testing them against available evidence (in the form of observations/process understanding) (Section 1.2). A number of “hydrological lines of evidence” (HLEs) are used in the hydrological literature to inform process understanding and provide the evidence to change our hydrological models (i.e., update our hypotheses). These HLEs range from laboratory scale experiments to large sample hydrology (described below in Section 2.3). This section evaluates HLEs in terms of their ability to accelerate the development of the next generation of conceptual models that are robust in terms of both predictive ability and hydrological fidelity when faced with hydrological change.

The outcomes of this evaluation of the HLEs are summarized in Figure 1. An overview of the evaluation process is outlined as follows. Section 2.1 starts by defining “hydrological change” and discussing the consequences of hydrological science’s limitations in controlling these changes (which impacts our ability to conduct controlled experiments). Section 2.2 identifies five key questions that need to be asked of each HLE to identify models that provide robust predictions of hydrological change for a broad range of catchments and/or changing conditions. Section 2.3 defines and evaluates four existing HLEs against these key questions. Section 3 then provides a comprehensive definition and evaluation of an emerging fifth HLE, Virtual Hydrological Laboratories (VHLs), which are increasingly being used to provide virtual observations for experimentation (see Fatichi et al., 2016). This includes a discussion of the unique opportunities and challenges provided by the VHL approach.

2.1. Hydrological Change Is Itself a Complex Issue

Hydrological changes occurring within a catchment can be complex. Building upon Thirel et al. (2015), we define hydrological change as a *significant change* in *physical characteristics* (e.g., land use), or *boundary conditions* (e.g., rainfall/temperature change). “Significant change” refers to systematic deviations of the system behaviors from the historical record. Though this definition is subjective, it is important to distinguish from trivial changes.

The inability of hydrological science to undertake controlled experiments (see Section 1.1) provides the motivation to classify the drivers of change into two groups, based on level of control that we have as hydrological scientists and practitioners.

- (a) “Climatic phenomena” refers to changes that we cannot control (at the scale of the catchment) due to climate variability or anthropogenic climate change. Climatic phenomena can alter boundary conditions (e.g., precipitation/temperature), which can change flow dynamics by activating or deactivating hydrological processes as the system adapts to the changes. Physical characteristics can also change as ecosystems adapt in response to climate and/or are subject to disturbances such as bushfires/vegetation mortality, which have a substantial impact on a catchment’s hydrology (Anderegg et al., 2015; Partington et al., 2022).
- (b) “Human interventions” refers to changes that we can have some degree of control over. This includes changes to land cover (e.g., agriculture/urbanization), water management (e.g., dam construction, or streambed mining) or other interventions (e.g., water withdrawals/pumping/diversions) that will have significant impact on flow dynamics.

2.2. Key Questions to Evaluate Hydrological Lines of Evidence

Regardless of their origin, hydrological changes to catchment characteristics or boundary conditions lead to changes in hydrological processes and manifest as changes in the input-state-output dynamics of the system. The altered system dynamics will challenge our assumptions regarding catchment structure and function as embodied by CMs. This is a key challenge faced by catchment models—will they be robust to these changes? In this study we build upon the definition of Mathevet et al. (2020) and use the term “robustness” to refer to the capability of a model to hold a certain level of performance for a broad range of different or changing conditions.

This definition provides the motivation to propose the following key questions to evaluate the HLEs, starting with a focus on evaluating hydrological change:

1. *How can the HLE evaluate the model for robustness to hydrological change due to (a) climatic phenomena, or (b) human intervention?*

This question has two possible answers, depending on our ability to control the real-world system of interest:

- *Controlled experimental approach.* The real-world system can be manipulated and the dynamical response of the system to this change can be observed. The benefits are that confounding factors can be controlled so that the impact of changes in the experimental factors on the system can definitively be determined. The challenge is that it is not always practical, feasible or ethical to manipulate real-world systems.
- *Observational approach.* The system cannot be manipulated and only its response to changes (e.g., climatic phenomena) can be observed. The benefits are that existing data from real-world systems can be used for studying the impact of changes that happened in the past, or phenomena/processes that are too vast or complex to control easily. The challenges are that uncontrolled confounding factors make it difficult to determine causation and one may have to wait a long time to obtain observations of rare, naturally occurring phenomena and/or detect trends due to slow moving gradual changes.

The next key questions (2–5) are motivated by the need for predictions to be robust to change for a range of realistic catchment types (size, properties etc.) and processes (moving beyond streamflow-only evaluation):

2. *Can the HLE evaluate the model for robustness on catchments of practically relevant size?* Although catchment management decisions can occur at a range of sizes, it is common for practitioners to be interested in impacts at larger catchment with sizes of the order of 10–10,000 km² or greater.
3. *Can the HLE evaluate the model for robustness on a diverse range of catchment properties?* Hydrological model users require robust predictions on catchments with a wide variety of properties (sizes, slopes, elevations, soil vegetation types, etc.).

4. *Can the HLE evaluate the model for robustness on a wide range of processes/predictions?* This question refers to moving beyond evaluating only streamflow prediction at the outlet, and could include streamflow at internal catchment locations, different streamflow generating mechanisms (infiltration excess vs. saturation excess), ET processes, snow/ice related physical phenomena, and groundwater/surface water exchanges.
5. *Can the HLE evaluate the model for robustness using real data?* To ensure realistic predictions, it is important to evaluate models against observations from real catchments. This consideration becomes important when using synthetic data generated in a Virtual Hydrological Laboratory for experimentation—see Section 3 for detailed discussion.

2.3. Evaluating Existing Hydrological Lines of Evidence

This section evaluates each of the existing hydrological lines of evidence against the five key questions outlined above, with outcomes summarized in Figure 1.

To answer question 1, regarding hydrological change, each HLE is categorized by whether the change can be evaluated using either a controlled experimental approach or an observational approach. In Figure 1, the answer to this question is split into two components as it depends on whether the change is due to (a) climatic phenomena or (b) human interventions (see Section 2.1). Key questions 2–5 are answered by classifying the HLEs into three different categories (“yes”, “no”, “limited”), based on the authors' experience, and with justification provided below. We recognize this classification is subjective, but the analysis has value because it starts a conversation on the pathway we should take to develop the next generation of CMs.

The principal lines of evidence that have been used in hydrology to inform process understanding and its incorporation into hydrological models are categorized and evaluated as follows:

- *Lab/plot scale experiments.* These focus on using detailed observations at small spatial scales ($\sim 5 \text{ m}^2$) at a plot or in a laboratory to enable the use of controlled experiments (Glaser et al., 2019; Nanda et al., 2018). Evidence of this kind is used to develop small scale process understanding, for a diverse range of hydrological properties, but has limited applicability for modeling catchments of practically relevant size.
- *Experimental catchments.* Heavily instrumented experimental catchments provide high-resolution observations of multiple hydrological variables and catchment properties. Evidence of this kind facilitates understanding of key hydrological processes (e.g., McGlynn et al., 2002) and can test the ability of models to represent dominant processes (e.g., Fenicia et al., 2014). However, experimental catchments are somewhat rare, typically small in size, and usually provide data for only short time periods (a few years). While these catchments enable experimental studies of the impacts of human intervention (e.g., land-use change), evaluation of changes related to climate is necessarily limited to the observational approach.
- *Paired catchments.* These are typically catchments selected to experience the same hydroclimate, with similar properties such as slope, aspect, soils, area, climate, and vegetation. This experimental setup reduces the influence of confounding factors. Evidence of this kind can be useful when one catchment undergoes change (e.g., alteration of land-use) so that the other catchment can serve as a control against which to assess the impacts of that change (Brown et al., 2005; McDonnell et al., 2018). Paired catchments can provide valuable evidence at practically relevant catchment sizes; however, it can be difficult to find suitably paired catchments that span the wide range of hydrological diversity of practical relevance, and such pairs typically are not heavily instrumented and thus cannot be used to test over a wide range of processes/predictions. While hydrological change is controlled for, any evidence pertaining to changing climate will be observational.
- *Large sample hydrology.* This involves the compilation of historical data for a common set of hydrologically relevant variables and catchment attributes across a large number of catchments (100–1000s) with widely varying catchment properties. Evidence of this kind can be used to assess model adequacy and performance for a wide range of catchments in a statistical sense and to study hydroclimatic and hydrogeological variability. To achieve relative consistency of input/output data used across catchments, the data sets used in large sample hydrology studies are typically limited to a few key variables (e.g., streamflow at catchment outlets, interpolated catchment rainfall and PET). For example, see data sets such as CAMELS (Addor et al., 2017; Fowler et al., 2021) and studies that use this HLE (Coron et al., 2012; Fowler et al., 2016; Gupta et al., 2014). Hence, this HLE only supports testing on a narrow range of processes/predictions, using an observational approach that is limited to historical changes in climate and/or land-use. It is noted that remotely sensed data,

across a range of scales from local to global, is increasingly used to derive new and even larger scale data sets for other hydrological variables for use in large sample hydrology studies (McCabe et al., 2017; Sheffield et al., 2018), and can overcome some of these limitations.

Figure 1 shows that the four existing HLEs are unable to simultaneously address all five key questions. They do not provide sufficient information to support the development of a new generation of CMs that can provide robust support for decision making in the face of hydrological change across a wide range of situations of practical relevance.

However, there exists another complementary approach that provides the opportunity to integrate multiple lines of evidence, while taking advantage of their respective strengths and overcoming their respective weaknesses—the *Virtual Hydrological Laboratory*—as described in the next section.

3. The Value of “Virtual Hydrological Laboratories” as Another Emerging Line of Evidence

3.1. Overview of a Virtual Hydrological Laboratory Approach

We envisage the *Virtual Hydrological Laboratory* (VHL; Figure 2) to be a conceptual framework consisting of a computational component that can generate a wide range of “virtual catchments” and an evaluation component that enables the use of controlled experiments to test conceptual models for a range of hydrological change scenarios. Each virtual catchment (VC) will be constructed using the best available PRMs that provide detailed physics-based understanding of sub-catchment scale storages, flows and physical transformations of water within the system. Being virtual, these VCs can differ systematically (and widely) in their climate and/or catchment characteristics, and can be subjected to a range of change scenarios to determine how their hydrological behaviors respond to such changes. Being based in process understanding, the VHL should be constructed to be consistent with the four other HLEs discussed above. The VHL can then provide a framework within which the next generation of CMs can be developed, by comprehensively evaluating competing catchment-scale model structures (systems architectures and process parameterizations) against these virtual catchments. We emphasize that while VCs are constructed from PRMs, they are distinct from traditional applications where a PRM is implemented for a specific catchment.

As discussed above, broad application of PRMs within a decision making context is limited by data availability and computation time. In a VHL, catchment characteristics can be assigned to cover a range of likely values, and VHL simulations used to create and test CMs that can then be more broadly applied. Testing the VHL could focus on behavioral modeling (Schaeffli et al., 2011; Yang & Chui, 2021) to ensure that relationships between drivers of change and hydrologic responses are consistent with the four lines of evidence rather than using catchment specific performance metrics for single hydrologic variables (e.g., Nash Sutcliffe efficiency between observed and modeled flow) that rely heavily on having complete and accurate data sets to parameterize the model and define boundary conditions. Note that in this VHL approach, selection of suitable “trustworthy” PRMs for the particular change experiment of interest will be crucial—see Section 3.3 for further discussion.

The VHL approach will facilitate and accelerate CM development because:

- (1) In a virtual catchment all hydrological components can be observed (albeit virtually), and thus CMs can be subjected to a more comprehensive level of scrutiny than current observational data sets allow;
- (2) The capability to systematically change catchment characteristics will provide the ability to conduct controlled experiments that isolate the key changes in hydrological processes for different catchments types—this is currently not possible with real-world experiments; and
- (3) The ability to systematically change climate and land cover/use characteristics will provide the opportunity to undertake hydrological change experiments and evaluate CMs on a wide range of future hydrological change scenarios that are outside the envelope of observations. This strategy provides real potential to proactively “future-proof” CMs to be able to support decision making in the face of changes that are yet to be observed.

The concept of a “virtual laboratory” or “digital twins,” although not (yet) fully utilized in hydrology, is extensively applied in other scientific and engineering fields. For instance, virtual simulators are used in military, aviation, and medical training to prepare for rare events that don't often occur during short training periods (Lateef, 2010). Hydrology faces similar challenges, particularly in predicting rare events like extreme floods due

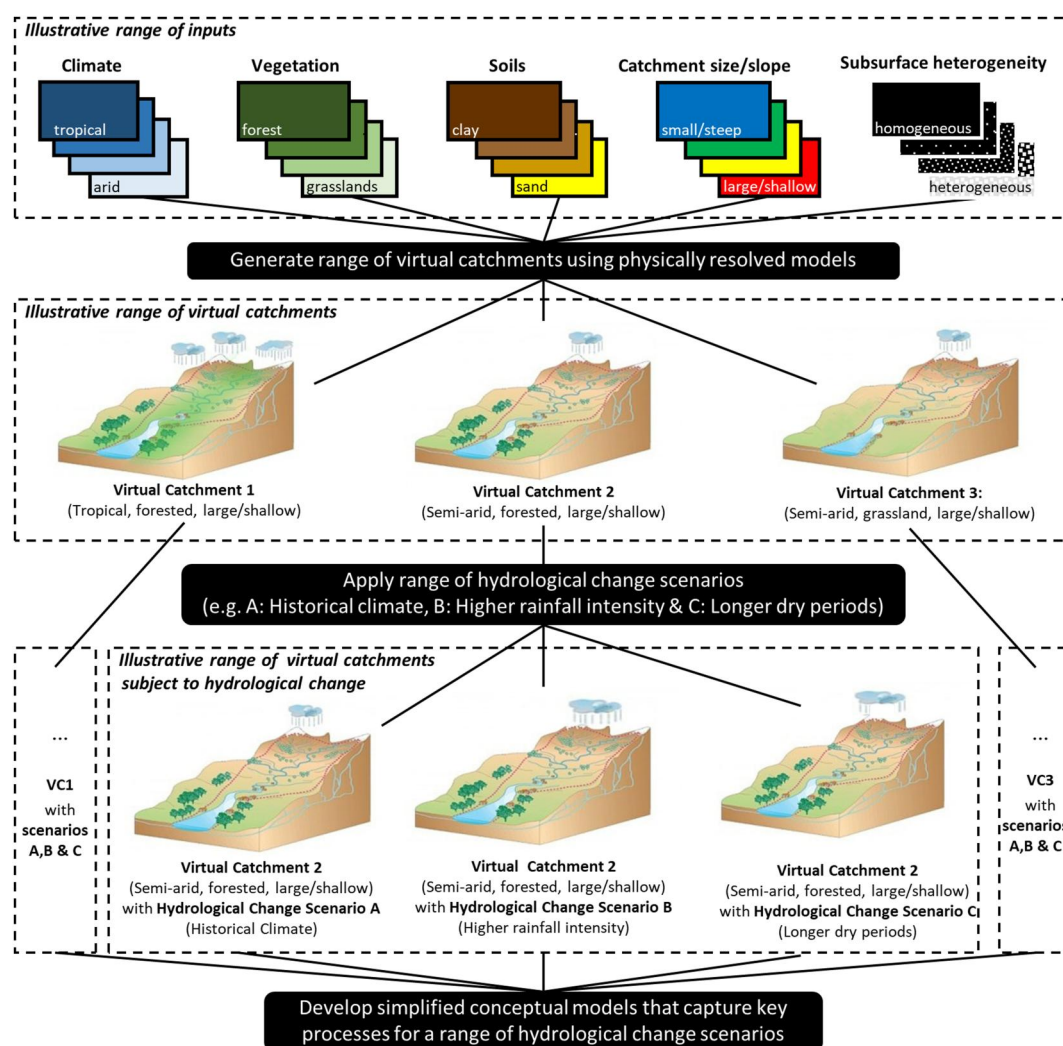


Figure 2. Schematic of the Virtual Hydrological Laboratory. (Catchment image adapted from <https://sage-loir.fr/wp-content/uploads/2022/03/SchemaBV.png>).

to changing climate. Climate modeling also employs virtual approaches, using complex, highly resolved models to guide the parameterization of less-resolved but computationally viable global climate models, which are too coarse to model cloud processes (O’Gorman & Dwyer, 2018). This parallels the hydrological challenge of capturing aggregate effects of surface and sub-surface heterogeneity in catchment-scale CMs. In biology, “virtual simulators” enhance the range of data for machine learning (ML) model training (Deist et al., 2019). Similarly, in hydrology, we often have limited observations for a catchment (typically areal rainfall/PET and outlet stream-flow), hence a VHL could generate a wider range of synthetic data to enhance system knowledge for better process evaluation and prediction.

The VHL concept has been gradually emerging in the hydrological sciences. Fatichi et al. (2016) provided several examples of how virtual labs could be integrated with natural ones (e.g., experimental catchments) to advance theory and process understanding by coordinating model development with field observation activities. Weiler and McDonnell (2004) used a range of VHL-like experiments to develop perceptual models of hillslope processes. Tague and Moritz (2019) used an eco-hydrological model as a VHL to investigate how different assumptions about plant root characteristics are likely to impact hydrologic responses to forest thinning. L. Li et al. (2014) used a VHL approach to enhance the performance of recursive digital baseflow filters and evaluate a conceptual model’s (AWBM) ability to simulate quick-flow and slow flow responses. J. Li et al. (2014) used a VHL to assess the ability of event-based methods to estimate flood frequency in different Australian climates.

Bennett et al. (2018) used a VHL to evaluate the ability of stochastic rainfall models to provide reliable streamflow predictions. Stephens et al. (2020) used a VHL to show historical periods with equivalent precipitation statistics cannot necessarily be used as proxies for future climate change when examining catchment runoff or model performance.

Inspired by the past use of VHLs, we argue that VHLs need to become a key hydrological line of evidence to inform development of the next generation of CMs.

3.2. Arguments for Development of the VHL Approach

To accelerate the development of the next generation of CMs, the use of VHL-type approaches needs to be expanded to become the backbone of the workflow in CM development and evaluation. Three key arguments why this transformation is required are outlined below:

Argument 1: *The VHL approach enables construction of conceptual models as simplified models guided by scientific understanding of the key dominant processes under hydrological change.*

To provide greater confidence that the next generation of CMs can provide robust predictions of hydrological change, these models will need to faithfully represent key processes—that is, they will need to improve their hydrological fidelity.

Initially—starting as far back as Linsley and Kohler (1958)—catchment CM development relied substantially on hydrologists' personal assessments regarding model structure and parameterization. Hence, the CMs reflected the knowledge, experience and prejudices of their developers (Gharari et al., 2021; Gupta et al., 2012). Contemporary CM development has evolved considerably since those early days. Notably, conceptual models of the GR family are derived from extensive testing using large sample hydrology (e.g., Mathevet et al., 2020; Perrin et al., 2003), while models such as SuperFlex have incorporated process understanding from the experimental catchments HLE (Fenicia et al., 2014, 2016). However, while current CMs share many common principles—for example, the conceptualization of water flowing through series and parallel assemblages of storage tanks—their system architectures, process parameterization equations, and parameter specifications vary widely (Clark et al., 2008).

One of the key challenges with the HLEs of large sample hydrology and experimental catchments for CM development is that these approaches are typically “backward” looking. The use of past historical data sets can only evaluate the ability of the CMs to represent past hydrological change (and the report card on those evaluations is generally not good—see Section 1.4). Large sample hydrology and experimental catchments will continue to play a role in hydrological model development as there is still significant scope to improve CMs by learning from the past. However, for the future, as we face unprecedented hydrological change, how do we provide greater confidence in our models to provide predictions of future change? If we continue to rely solely on the use of HLEs such as large sample hydrology and experimental catchments, we will have to wait until the hydrological change has occurred before we can evaluate and improve CMs. That is too late for supporting decisions in the interim. We cannot wait. We need new thinking to accelerate the development of CMs that can capture the dominant processes under hydrological change.

To capitalize on current scientific knowledge, one strategy is to treat CM development as “reduced-order” modeling, using detailed PRMs to guide simpler CMs. The CM should aim to retain key input-state-output dynamics from PRMs, particularly those that are relevant for the hydro-climatic region of interest and application objective (e.g., capturing future changes in catchment yield or flood peaks). This approach recognizes that the best CM for a particular hydro-climatic region, and for a specific hydrological purpose, may be different (Watson et al., 2013). A VHL-like approach can be followed, starting with a highly realistic model for the target environment, and conducting virtual experiments to design a minimal-information-loss, reduced-order CM. This creates a continuum of progressively simpler model representations, minimizing information loss with respect to a model's intended purpose (Tishby et al., 2000). By doing so, it could be argued that the resulting CMs reflect current best available relevant scientific understanding, at least for the given scale and context. This reasoning aligns with the “organizing principles” of the behavioral modeling approach proposed by Schaeffli et al. (2011). As such, the “reduced-order” approach aims to get “the right answers for the right reasons” and achieve hydrological fidelity under climate change.

The VHL could be designed to quickly compare different CMs—and help the modeler choose the appropriate CM for the particular application—similar to frameworks such as FUSE (Clark et al., 2008), SuperFlex (Kavetski & Fenicia, 2011), MaRRMoT (Knoben et al., 2019), Raven (Craig et al., 2020), and others. An approach that could be helpful in terms of identifying the appropriate reduced-order models here is surrogate or emulation modeling (e.g., Asher et al., 2015; Yang et al., 2018), where simplified models that capture the behavior of specific quantities of interest are developed from PRMs. These surrogate approaches could be explored to help identify (or used instead of) suitable CMs. The VHL could also provide improved visualization of model structure and underlying hydrologic concepts, which would contribute to more reliable, and defensible use of CMs, particularly in decision-making contexts (e.g., Partington et al., 2013; Tague & Frew, 2021).

Inevitably, when moving from a complex PRM to a simpler CM, there will be some information loss and there is a risk that simplified CMs are unable to capture the entire range of streamflow regimes. To overcome this deficiency, probabilistic error models (e.g., Hunter et al., 2021; McInerney et al., 2017, 2020) could be used in combination with these simpler CMs to capture streamflow uncertainty that is important for management applications such as streamflow forecasting.

Argument 2: *A VHL enables the use of controlled experiments to evaluate hydrological change.*

Clark et al. (2011) considered the set of available CMs to be “Multiple Working Hypotheses” of catchment process representations. Since we must account for processes that are relevant to hydrological change of various kinds, our “hypotheses” of catchment processes clearly need to be improved. However, there is a potentially large variety of processes that may need to be included in these new CMs to enable investigation of the impacts of changes to boundary conditions and/or physical properties of the catchment of interest.

A key challenge to determining whether a process should be included in a CM is the existence of “confounding factors” that make it difficult to test hypotheses. As an example, a change in catchment vegetation due to bushfire and/or land clearing may lead to a change in catchment yield. However, yield is also dependent on climate variability (annual rainfall, temperature), which can act as a confounding factor. If we try to correlate changes in catchment yield with changes in vegetation pre- and post-bushfire, the different climates pre- and post-bushfire can confound the development of a suitable relationship. Such confounding effects of climate variability could be reduced by computing yields over longer periods (i.e., 20 years instead of 10 years), but then as the vegetation grows back, the change in catchment yield can become very difficult to determine.

A major advantage of the VHL, which is not easily provided by the existing HLEs, is the powerful ability to conduct controlled experiments to test for changes in either climatic phenomena (via virtual simulation of climate) or human interventions (via virtual simulation of water management/land-use change) at practically relevant catchment scales (Figure 1). In a VHL, we can (more or less) independently vary, and therefore control for, a variety of important factors, such as catchment size and shape, geology, topographic gradient (aspect and slope), channel network form, vegetation type, soil layering and structure, material properties, sub-surface boundary conditions, and climate drivers, among many more. Accordingly, we can represent a very diverse range of catchments (limited only by hydrological understanding and computational power). As the VHL can comprehensively embody our current state of knowledge regarding the catchment system (albeit, virtually), it can be used as a benchmark against which any simpler CM hypothesis can be tested over a wide range of potential applications. Being able to conduct controlled experiments provides the VHL with the unique ability to evaluate the impact of hydrological change in a controlled way for both human intervention (e.g., land-use change) and climatic phenomena (e.g., climate variability/change). As Figure 1 shows, no other HLE provides this capability at practically relevant scales of interest. This translates into the opportunity to proactively build CMs that are robust under hydrological change, rather than passively waiting (potentially years or, more likely, decades) for the observational data sets to reveal the flaws in existing CMs.

Argument 3: *VHLs provide the vehicle that enables the integrating of multiple lines of evidence into the next generation of conceptual models.*

The first two arguments outline the particular strengths of the VHL approach that overcome the limitations of existing HLEs. However, rather than replacing existing HLEs, we envision that the VHL approach represents a *fifth HLE* (Figure 1).

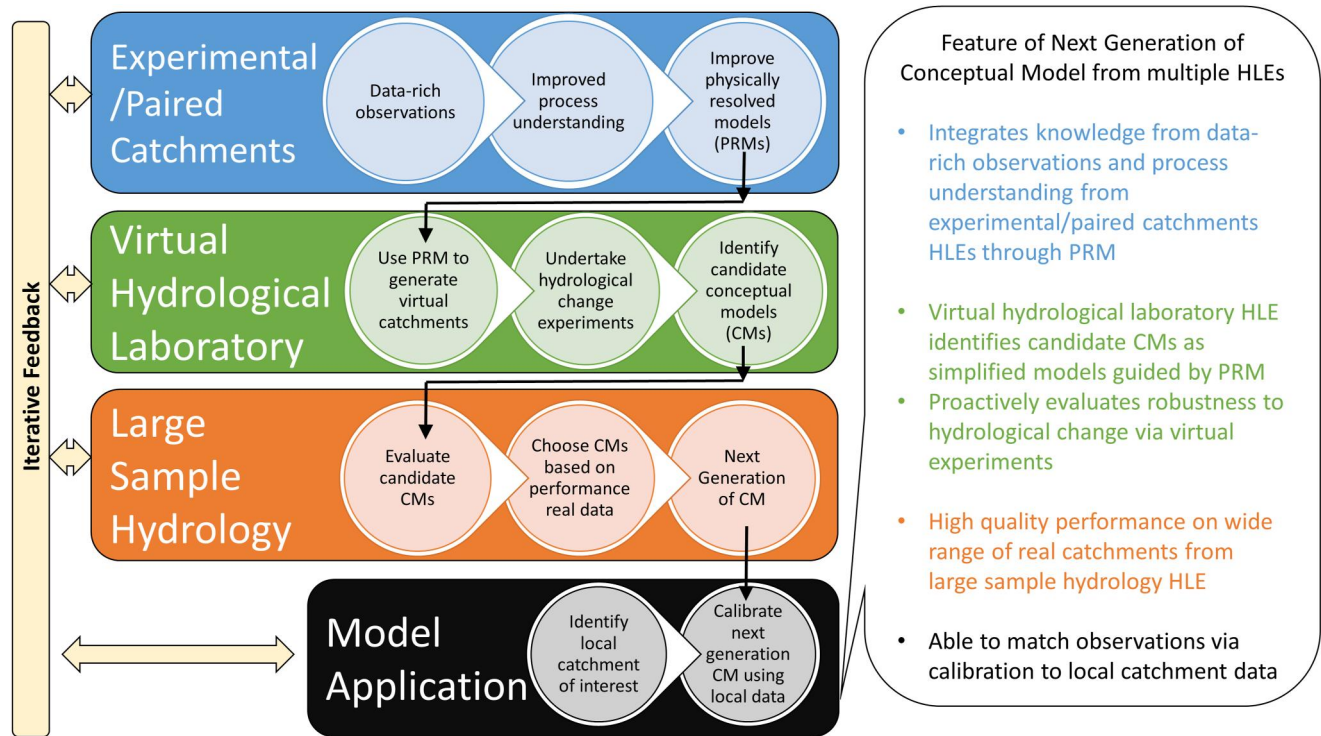


Figure 3. Framework that integrates multiple hydrological lines of evidence, including a VHL, to provide the best available scientific understanding and develop the next generation of conceptual models.

The VHL approach enables us to integrate multiple lines of evidence, complementing and integrating existing HLEs (Figure 3). The resulting framework uses data-rich observations from experimental catchments and/or paired catchments to improve process understanding, and subsequently improve PRMs (blue section, Figure 3). The VHL can be used to develop virtual catchments and undertake hydrological change experiments, enabling the development of candidate CMs as simpler reduced-order models and proactive evaluation of robustness to hydrological change (green section, Figure 3). Large-sample hydrological studies can be used to test these candidate CMs and ensure high quality performance on a wide range of catchments (orange section, Figure 3), weeding out any candidate CMs that are unable to match real observations. Finally, calibration with local data remains part of the conceptual model application step (black section, Figure 3). While each step uses different lines of evidence, it will almost certainly benefit from iterative feedback from the other steps to identify and address limitations at each stage.

A key benefit of the integrated multiple HLE framework (Figure 3) is that it does not rely solely on a single line of evidence to develop the next generation of CMs. We are not relying exclusively on PRMs to provide the basis for all CM future model development. We are not building CMs using only Large Sample Hydrology, which has provided a solid foundation for CM development, but can only look at past change using an observational approach. This integrated approach based on multiple HLEs will enable the cross-pollination of ideas across different sub-disciplines within hydrological sciences. This is why the iterative feedback loop is so important—it provides opportunity for ideas such as a reverse coupling, where the use of existing well-performing CMs could serve as a check on the PRMs, thus yielding a coherent combination of PRMs and CMs. We argue, based on these benefits, that this integrated multiple HLE framework should become a standard part of the CM development workflow, rather than the exception.

3.3. Limitations of the VHL Approach: Selection of a Trustworthy PRM Is Crucial

It is important to acknowledge the limitations of using the VHL approach as a fifth HLE. These include:

- (1) The data used in the VHL approach to test the CMs for hydrological change are not “real” observations (see Figures 1 and 2). The ability to conduct hydrological change experiments using the VHL (see argument 2)

relies on the assumption that the PRMs incorporate not only relevant-scale hydrological processes (such as soil infiltration and evapotranspiration from vegetation), but that they also provide realistic representations of emergent properties at the catchment scale, such as flood peaks (e.g., Farmer et al., 2015) and/or Budyko curves.

- (2) The VHL approach assumes a level of “trust” that the PRMs capture future hydrological change. This is a thorny issue because the ability to capture future change cannot be guaranteed for any environmental model. If we accept the premise that PRMs reflect our best current understanding of environmental physics, then it is logical to propose that such models represent our best current chance to simulate future changes. The review by Fatichi et al. (2016) cites several studies where this assumption was made and PRMs were used to estimate the impact of future climates (e.g., Huntington & Niswonger, 2012; Piras et al., 2014; Sulis et al., 2012) and land-use/cover change (e.g., Ebel & Mirus, 2014; Ogden & Stallard, 2013; Pierini et al., 2014; van Roosmalen et al., 2009). Rodriguez and Tomasella (2016) used a PRM to examine the impact of land-use change in Amazonian basins. More recently, Gelfan and Millionshchikova (2018) and Gelfan et al. (2020) proposed comprehensive evaluation/validation tests of PRMs for climate change impacts. Stephens et al. (2020) used a PRM in a series of virtual change experiments to evaluate the impact of changing climate and its interaction with vegetation dynamics on catchment runoff response.

For the VHL approach to be adopted, it is important to be aware of these assumptions/limitations and take steps to mitigate them. The aim is to avoid the situation, where, as eloquently stated by a reviewer “we encircle ourselves by developing models with different models that we have already built.” A key step is the requirement for careful scrutiny and selection of PRMs. This selection process needs to consider the strength of evidence for the use of a given PRM for the chosen hydrological change experiment in a given catchment type (hereafter in this paragraph referred to as a “scenario”). The aim of this selection process is to identify PRMs that are sufficiently “trustworthy” to be used in the VHL for a given scenario. For example, if the scenario of interest is to capture changes in vegetation dynamics, the ecohydrological model RHESSys could be considered a “trustworthy” PRM because its representation of processes related to vegetation dynamics have been developed, tested and refined based on large number of studies and using multiple HLEs, namely experimental and/or paired catchments (Son et al., 2016; Tsamir et al., 2019; Zierl et al., 2007). Indeed, RHESSys has already been used in virtual change experiments (Stephens et al., 2020). For other scenarios, a “trustworthy” PRM may not currently exist, especially considering the wide range of potentially different hydrological changes (see Section 2.1), wide range of catchment types and potential processes of interest. In these scenarios, the lack of trustworthy PRM will hopefully motivate further research to develop a sufficiently trustworthy PRM. This research may involve comparisons with CMs and observed data in order to question and improve existing PRMs. An example of a scenario where a trustworthy PRM does not currently exist is the shifts in streamflow found by Fowler et al. (2020), and references therein, for which there is yet to be a study that evaluates whether a PRM can capture this change.

If the VHL approach is to be adopted, there is a clear need to develop an objective process for determining if a PRM is trustworthy. This could include adapting the criteria used for multiple lines and levels of evidence in ecology (e.g., Norris et al., 2012), while recent developments of validation techniques by Gelfan and Millionshchikova (2018) and Gelfan et al. (2020) are also promising in terms of developing and establishing trustworthy PRMs for climate change.

One could argue these limitations preclude the VHL approach from helping guide CM development. However, this argument misses several important points.

- (1) In order to predict future behavior of environmental systems, we must unavoidably use our best available current understanding of these systems to make assumptions about the future.
- (2) The integration of multiple HLEs in an iterative framework (as outlined in Figure 3) will likely lead to explicitly confronting, questioning and ultimately improving the assumptions underlying both CMs and PRMs and the impact of those assumptions on predicting future change. Indeed, we hope this framework will further encourage the ongoing improvement and refinement of PRMs themselves, to make them more “trustworthy.”
- (3) The integration of multiple HLEs, notably large sample hydrology, also mitigates these limitations by avoiding exclusive reliance on the VHL. By constraining CM development to be consistent with the understanding encoded into VHLs and emergent catchment behavior from real catchment data used in large

sample hydrology, we gain a potentially very powerful way to develop scientifically informed predictions of future change.

Finally, we note that the modification of CMs to simulate past change has already been successful in certain contexts. For example, Duethmann et al. (2020) modified the HBV model to account for vegetation dynamics in 156 Austrian catchments. These types of studies give more confidence that CMs can be modified to provide more reliable predictions of change. The VHL approach provides a mechanism to inform these modifications for future changes (as yet unseen) by reconciling the strengths of multiple HLEs.

4. Summary and Call to Action

The need for new thinking becomes ever more pressing as the predictive skill of current hydrological models is confronted by rapid, multi-faceted change and catchments are pushed outside of the envelope of historical experience. To continue to advance our scientific enterprise, we should revisit the role of conceptual models as our most important workhorse, ensuring their use not merely as tools for prediction, but as agents that represent our best available scientific understanding (applied in a given context).

Conceptual models are the model type most commonly used for supporting decision making in surface water hydrology. As such, they and their predictions essentially represent the “shopfront” of hydrological science. Our next generation of conceptual models will need to be firmly founded upon the multiple lines of evidence available in hydrology. This paper argues that the key to integrating these lines of evidence is the use of Virtual Hydrological Laboratories (VHLs). The VHL approach enables the integration of knowledge from experimental/paired catchments through physically resolved models and provides the ability to conduct controlled experiments using appropriately selected physically resolved models in a virtual lab-type environment. This VHL approach will accelerate the development of the next generation of conceptual models for providing robust predictions of hydrological change through a process of hypothesis falsification—rather than waiting decades or even centuries for the observations to do the falsification for us. This VHL approach also requires us, as hydrologists, to confront and enhance the assumptions regarding future change that are embodied in both conceptual and physically resolved models.

The VHL approach aims to spearhead the proactive development of the next generation of conceptual models that can better support decision making, and will shift the focus toward hydrological fidelity (i.e., “getting the right answers for the right reasons”). This proactive development of conceptual models that can reliably predict future outcomes before they occur is expected by policy makers who need to know what to plan for.

Acknowledgments

This research was supported by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (project DP190102138).

References

- Abbott, M. B., Babovic, V. M., & Cunge, J. A. (2001). Towards the hydraulics of the hydroinformatics era. *Journal of Hydraulic Research*, 39(4), 339–349. <https://doi.org/10.1080/0022168010949839>
- Addor, N., & Melsen, L. A. (2019). Legacy, rather than adequacy, drives the selection of hydrological models. *Water Resources Research*, 55(1), 378–390. <https://doi.org/10.1029/2018wr022958>
- Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21(10), 5293–5313. <https://doi.org/10.5194/hess-21-5293-2017>
- Ampadu, B., Chappell, N. A., & Kasei, R. A. (2013). Rainfall-riverflow modelling approaches: Making a choice of data-based mechanistic modelling approach for data limited catchments: A review. *Canadian Journal of Pure and Applied Sciences*, 7, 2571–2580.
- Anderegg, W. R., Hicke, J. A., Fisher, R. A., Allen, C. D., Aukema, J., Bentz, B., et al. (2015). Tree mortality from drought, insects, and their interactions in a changing climate. *New Phytologist*, 208(3), 674–683. <https://doi.org/10.1111/nph.13477>
- Asher, M. J., Croke, B. F. W., Jakeman, A. J., & Peeters, L. J. M. (2015). A review of surrogate models and their application to groundwater modeling. *Water Resources Research*, 51(8), 5957–5973. <https://doi.org/10.1002/2015wr016967>
- Bennett, B., Thyer, M., Leonard, M., Lambert, M., & Bates, B. (2018). A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model. *Journal of Hydrology*, 556, 1123–1138. <https://doi.org/10.1016/j.jhydrol.2016.12.043>
- Bergström, S. (1995). The HBV model. In V. P. Singh (Ed.), *Computer models of watershed hydrology*. Water Resources Publications, Highlands Ranch, CO.
- Beven, K. (1989). Changing ideas in hydrology—The case of physically-based models. *Journal of Hydrology*, 105(1–2), 157–172. [https://doi.org/10.1016/0022-1694\(89\)90101-7](https://doi.org/10.1016/0022-1694(89)90101-7)
- Beven, K. J. (2000). Uniqueness of place and process representations in hydrological modelling. *Hydrology and Earth System Sciences*, 4(2), 203–213. <https://doi.org/10.5194/hess-4-203-2000>
- Beven, K. J. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1–2), 18–36. <https://doi.org/10.1016/j.jhydrol.2005.07.007>
- Bouaziz, L. J. E., Fenicia, F., Thirel, G., de Boer-Euser, T., Buitink, J., Brauer, C. C., et al. (2021). Behind the scenes of streamflow model performance. *Hydrology and Earth System Sciences*, 25(2), 1069–1095. <https://doi.org/10.5194/hess-25-1069-2021>
- Brown, A. E., Zhang, L., McMahon, T. A., Western, A. W., & Vertessy, R. A. (2005). A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. *Journal of Hydrology*, 310(1–4), 28–61. <https://doi.org/10.1016/j.jhydrol.2004.12.010>

- Brunner, P., & Simmons, C. T. (2012). HydroGeoSphere: A fully integrated, physically based hydrological model. *Ground Water*, 50(2), 170–176. <https://doi.org/10.1111/j.1745-6584.2011.00882.x>
- Burnash, R. J., Ferral, R. L., & McGuire, R. A. (1973). *A generalized streamflow simulation system: Conceptual modeling for digital computers*. US Department of Commerce, National Weather Service, and State of California, Department of Water Resources.
- Camporese, M., Daly, E., & Paniconi, C. (2015). Catchment-scale Richards equation-based modeling of evapotranspiration via boundary condition switching and root water uptake schemes. *Water Resources Research*, 51(7), 5756–5771. <https://doi.org/10.1002/2015wr017139>
- Camporese, M., Penna, D., Borga, M., & Paniconi, C. (2014). A field and modeling study of nonlinear storage-discharge dynamics for an Alpine headwater catchment. *Water Resources Research*, 50(2), 806–822. <https://doi.org/10.1002/2013wr013604>
- Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, 47(9), W09301. <https://doi.org/10.1029/2010wr009827>
- Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., et al. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, 51(4), 2498–2514. <https://doi.org/10.1002/2015wr017198>
- Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., et al. (2008). Framework for understanding structural errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, 44(12), W00B02. <https://doi.org/10.1029/2007WR006735>
- Clarke, R. T. (2007). Hydrological prediction in a non-stationary world. *Hydrology and Earth System Sciences*, 11(1), 408–414. <https://doi.org/10.5194/hess-11-408-2007>
- Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., & Hendrickx, F. (2012). Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. *Water Resources Research*, 48(5), W05552. <https://doi.org/10.1029/2011wr011721>
- Craig, J. R., Brown, G., Chlumsky, R., Jenkinson, R. W., Jost, G., Lee, K., et al. (2020). Flexible watershed simulation with the Raven hydrological modelling framework. *Environmental Modelling & Software*, 129, 104728. <https://doi.org/10.1016/j.envsoft.2020.104728>
- Croke, B. F. W., & Jakeman, A. J. (2004). A catchment moisture deficit module for the IHACRES rainfall-runoff model. *Environmental Modelling & Software*, 19, 1–5. <https://doi.org/10.1016/j.envsoft.2003.09.001>
- Deist, T. M., Patti, A., Wang, Z., Krane, D., Sorenson, T., & Craft, D. (2019). Simulation-assisted machine learning. *Bioinformatics*, 35(20), 4072–4080. <https://doi.org/10.1093/bioinformatics/btz199>
- Dooge, J. C. I. (1986). Looking for hydrologic laws. *Water Resources Research*, 22(9S), 546–558. <https://doi.org/10.1029/wr022i09sp0046s>
- Duethmann, D., Blöschl, G., & Parajka, J. (2020). Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? *Hydrology and Earth System Sciences*, 24(7), 3493–3511. <https://doi.org/10.5194/hess-24-3493-2020>
- Ebel, B. A., & Mirus, B. B. (2014). Disturbance hydrology: Challenges and opportunities. *Hydrological Processes*, 28(19), 5140–5148. <https://doi.org/10.1002/hyp.10256>
- Farmer, W. H., Over, T. M., & Vogel, R. M. (2015). Multiple regression and inverse moments improve the characterization of the spatial scaling behavior of daily streamflows in the southeast United States. *Water Resources Research*, 51(3), 1775–1796. <https://doi.org/10.1002/2014wr015924>
- Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., et al. (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, 537, 45–60. <https://doi.org/10.1016/j.jhydrol.2016.03.026>
- Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., & Freer, J. (2014). Catchment properties, function, and conceptual model representation: Is there a correspondence? *Hydrological Processes*, 28(4), 2451–2467. <https://doi.org/10.1002/hyp.9726>
- Fenicia, F., Kavetski, D., Savenije, H. H. G., & Pfister, L. (2016). From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. *Water Resources Research*, 52(2), 954–989. <https://doi.org/10.1002/2015wr017398>
- Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., et al. (2020). Many commonly used rainfall-runoff models lack long, slow dynamics: Implications for runoff projections. *Water Resources Research*, 56(5), e2019WR025286. <https://doi.org/10.1029/2019wr025286>
- Fowler, K., Peel, M., Saft, M., Nathan, R., Horne, A., Wilby, R., et al. (2022). Hydrological shifts threaten water resources. *Water Resources Research*, 58(8), e2021WR031210. <https://doi.org/10.1029/2021wr031210>
- Fowler, K. J., Peel, M. C., Western, A. W., Zhang, L., & Peterson, T. J. (2016). Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models. *Water Resources Research*, 52(3), 1820–1846. <https://doi.org/10.1002/2015wr018068>
- Fowler, K. J. A., Acharya, S. C., Addor, N., Chou, C., & Peel, M. C. (2021). CAMELS-AUS: Hydrometeorological time series and landscape attributes for 222 catchments in Australia. *Earth System Science Data*, 13(8), 3847–3867. <https://doi.org/10.5194/essd-13-3847-2021>
- Freeze, R. A., & Harlan, R. L. (1969). Blueprint for a physically-based, digitally-simulated hydrologic response model. *Journal of Hydrology*, 9(3), 237–258. [https://doi.org/10.1016/0022-1694\(69\)90020-1](https://doi.org/10.1016/0022-1694(69)90020-1)
- Gelfan, A., Kalugin, A., Krylenko, I., Nasonova, O., Gusev, Y., & Kovalev, E. (2020). Does a successful comprehensive evaluation increase confidence in a hydrological model intended for climate impact assessment? *Climatic Change*, 163(3), 1165–1185. <https://doi.org/10.1007/s10584-020-02930-z>
- Gelfan, A. N., & Millionshchikova, T. D. (2018). Validation of a hydrological model intended for impact study: Problem statement and solution example for Selenga River Basin. *Water Resources*, 45(S1), 90–101. <https://doi.org/10.1134/s0097807818050354>
- Gharari, S., Gupta, H. V., Clark, M. P., Hrachowitz, M., Fenicia, F., Matgen, P., & Savenije, H. H. G. (2021). Understanding the information content in the hierarchy of model development decisions: Learning from data. *Water Resources Research*, 57(6), e2020WR027948. <https://doi.org/10.1029/2020wr027948>
- Gibbs, M. S., McInerney, D., Humphrey, G., Thyer, M. A., Maier, H. R., Dandy, G. C., & Kavetski, D. (2018). State updating and calibration period selection to improve dynamic monthly streamflow forecasts for an environmental flow management application. *Hydrology and Earth System Sciences*, 22(1), 871–887. <https://doi.org/10.5194/hess-22-871-2018>
- Glaser, B., Jackisch, C., Hopp, L., & Klaus, J. (2019). How meaningful are plot-scale observations and simulations of preferential flow for catchment models? *Vadose Zone Journal*, 18(1), 1–18. <https://doi.org/10.2136/vzj2018.08.0146>
- Grayson, R. B., Moore, I. D., & McMahon, T. A. (1992). Physically based hydrologic modeling: 2. Is the concept realistic? *Water Resources Research*, 28(10), 2659–2666. <https://doi.org/10.1029/92wr01259>
- Guillaume, J. H. A., Jakeman, J. D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., et al. (2019). Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose. *Environmental Modelling & Software*, 119, 418–432. <https://doi.org/10.1016/j.envsoft.2019.07.007>
- Gunderson, L. H., & Holling, C. S. (2002). *Panarchy: Understanding transformations in human and natural systems*. Island press.

- Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., & Ye, M. (2012). Towards a comprehensive assessment of model structural adequacy. *Water Resources Research*, 48(8), W08301. <https://doi.org/10.1029/2011wr011044>
- Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., & Andréassian, V. (2014). Large-sample hydrology: A need to balance depth with breadth. *Hydrology and Earth System Sciences*, 18(2), 463–477. <https://doi.org/10.5194/hess-18-463-2014>
- Horton, P., Schaeffli, B., & Kauzlaric, M. (2022). Why do we have so many different hydrological models? A review based on the case of Switzerland. *Wiley Interdisciplinary Reviews: Water*, 9(1), e1574. <https://doi.org/10.1002/wat2.1574>
- Hunter, J., Thyer, M., McInerney, D., & Kavetski, D. (2021). Achieving high-quality probabilistic predictions from hydrological models calibrated with a wide range of objective functions. *Journal of Hydrology*, 603, 126578. <https://doi.org/10.1016/j.jhydrol.2021.126578>
- Huntington, J. L., & Niswonger, R. G. (2012). Role of surface-water and groundwater interactions on projected summertime streamflow in snow dominated regions: An integrated modeling approach. *Water Resources Research*, 48(11), W11524. <https://doi.org/10.1029/2012wr012319>
- Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, 29(8), 2637–2649. <https://doi.org/10.1029/93wr00877>
- Kavetski, D., & Fenicia, F. (2011). Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. *Water Resources Research*, 47(11), W11511. <https://doi.org/10.1029/2011wr010748>
- Knoben, W. J. M., Freer, J. E., Fowler, K. J. A., Peel, M. C., & Woods, R. A. (2019). Modular assessment of rainfall-runoff models toolbox (MARRMoT) v1.2: An open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations. *Geoscientific Model Development*, 12(6), 2463–2480. <https://doi.org/10.5194/gmd-12-2463-2019>
- Kollet, S. J., & Maxwell, R. M. (2006). Integrated surface-groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, 29(7), 945–958. <https://doi.org/10.1016/j.advwatres.2005.08.006>
- Kollet, S. J., Maxwell, R. M., Woodward, C. S., Smith, S., Vanderborght, J., Vereecken, H., & Simmer, C. (2010). Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources. *Water Resources Research*, 46(4), W04201. <https://doi.org/10.1029/2009wr008730>
- Lateef, F. (2010). Simulation-based learning: Just like the real thing. *Journal of Emergencies, Trauma, and Shock*, 3(4), 348–352. <https://doi.org/10.4103/0974-2700.70743>
- Li, J., Thyer, M., Lambert, M., Kuczera, G., & Metcalfe, A. (2014). An efficient causative event-based approach for deriving the annual flood frequency distribution. *Journal of Hydrology*, 510, 412–423. <https://doi.org/10.1016/j.jhydrol.2013.12.035>
- Li, L., Lambert, M. F., Maier, H. R., Partington, D., & Simmons, C. T. (2015). Assessment of the internal dynamics of the Australian water balance model under different calibration regimes. *Environmental Modelling & Software*, 66, 57–68. <https://doi.org/10.1016/j.envsoft.2014.12.015>
- Li, L., Maier, H. R., Partington, D., Lambert, M. F., & Simmons, C. T. (2014). Performance assessment and improvement of recursive digital baseflow filters for catchments with different physical characteristics and hydrological inputs. *Environmental Modelling & Software*, 54, 39–52. <https://doi.org/10.1016/j.envsoft.2013.12.011>
- Linsley, R. K., & Kohler, M. A. (1958). *Hydrology for engineers*. McGraw Hill.
- Maier, H. R., Zheng, F., Gupta, H., Chen, J., Mai, J., Savic, D., et al. (2023). On how data are partitioned in model development and evaluation: Confronting the elephant in the room to enhance model generalization. *Environmental Modelling & Software*, 167, 105779. <https://doi.org/10.1016/j.envsoft.2023.105779>
- Mathevet, T., Gupta, H., Perrin, C., Andréassian, V., & Le Moine, N. (2020). Assessing the performance and robustness of two conceptual rainfall-runoff models on a worldwide sample of watersheds. *Journal of Hydrology*, 585, 124698. <https://doi.org/10.1016/j.jhydrol.2020.124698>
- McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., et al. (2017). The future of Earth observation in hydrology. *Hydrology and Earth System Sciences*, 21(7), 3879–3914. <https://doi.org/10.5194/hess-21-3879-2017>
- McDonnell, J. J., Evaristo, J., Bladon, K. D., Buttle, J., Creed, I. F., Dymond, S. F., et al. (2018). Water sustainability and watershed storage. *Nature Sustainability*, 1(8), 378–379. <https://doi.org/10.1038/s41893-018-0099-8>
- McGlynn, B. L., McDonnell, J. J., & Brammer, D. D. (2002). A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchments, New Zealand. *Journal of Hydrology*, 257(1–4), 1–26. [https://doi.org/10.1016/s0022-1694\(01\)00559-5](https://doi.org/10.1016/s0022-1694(01)00559-5)
- McInerney, D., Thyer, M., Kavetski, D., Laugesen, R., Tuteja, N., & Kuczera, G. (2020). Multi-temporal hydrological residual error modelling for seamless sub-seasonal streamflow forecasting. *Water Resources Research*, 56(11), e2019WR026979. <https://doi.org/10.1029/2019wr026979>
- McInerney, D., Thyer, M., Kavetski, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*, 53(3), 2199–2239. <https://doi.org/10.1002/2016wr019168>
- Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., et al. (2013). “Panta Rhei—Everything flows”: Change in hydrology and society—The IAHS scientific decade 2013–2022. *Hydrological Sciences Journal*, 58(6), 1256–1275. <https://doi.org/10.1080/02626667.2013.809088>
- Mount, N. J., Maier, H. R., Toth, E., Elshorbagy, A., Solomatine, D., Chang, F. J., & Abrahart, R. J. (2016). Data-driven modelling approaches for socio-hydrology: Opportunities and challenges within the Panta Rhei Science Plan. *Hydrological Sciences Journal*, 61, 1192–1208. <https://doi.org/10.1080/02626667.2016.1159683>
- Nanda, A., Sen, S., Jirwan, V., Sharma, A., & Kumar, V. (2018). Understanding plot-scale hydrology of Lesser Himalayan watershed—A field study and HYDRUS-2D modelling approach. *Hydrological Processes*, 32(9), 1254–1266. <https://doi.org/10.1002/hyp.11499>
- Norris, R. H., Webb, J. A., Nichols, S. J., Stewardson, M. J., & Harrison, E. T. (2012). Analyzing cause and effect in environmental assessments: Using weighted evidence from the literature. *Freshwater Science*, 31(1), 5–21. <https://doi.org/10.1899/11-027.1>
- Ogden, F. L., & Stallard, R. F. (2013). Land use effects on ecosystem service provisioning in tropical watersheds, still an important unsolved problem. *Proceedings of the National Academy of Sciences of the United States of America*, 110(52), E5037. <https://doi.org/10.1073/pnas.1314747111>
- O’Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, 10, 2548–2563. <https://doi.org/10.1029/2018ms001351>
- Partington, D., Brunner, P., Frei, S., Simmons, C. T., Werner, A. D., Therrien, R., et al. (2013). Interpreting streamflow generation mechanisms from integrated surface-subsurface flow models of a riparian wetland and catchment. *Water Resources Research*, 49(9), 5501–5519. <https://doi.org/10.1002/wrcr.20405>
- Partington, D., Thyer, M., Shanafield, M., McInerney, D., Westra, S., Maier, H., et al. (2022). Predicting wildfire induced changes to runoff: A review and synthesis of modeling approaches. *Wiley Interdisciplinary Reviews Water*, 9(5), e1599. <https://doi.org/10.1002/wat2.1599>
- Peel, M. C., & McMahon, T. A. (2020). Historical development of rainfall-runoff modeling. *Wiley Interdisciplinary Reviews Water*, 7(5), e1471. <https://doi.org/10.1002/wat2.1471>

- Perrin, C., Michel, C., & Andreassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1–4), 275–289. [https://doi.org/10.1016/S0022-1694\(03\)00225-7](https://doi.org/10.1016/S0022-1694(03)00225-7)
- Pierini, N. A., Vivoni, E. R., Robles-Morua, A., Scott, R. L., & Nearing, M. A. (2014). Using observations and a distributed hydrologic model to explore runoff thresholds linked with mesquite encroachment in the Sonoran Desert. *Water Resources Research*, 50(10), 8191–8215. <https://doi.org/10.1002/2014wr015781>
- Piras, M., Mascaro, G., Deidda, R., & Vivoni, E. R. (2014). Quantification of hydrologic impacts of climate change in a Mediterranean basin in Sardinia, Italy, through high-resolution simulations. *Hydrology and Earth System Sciences*, 18(12), 5201–5217. <https://doi.org/10.5194/hess-18-5201-2014>
- Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., et al. (2021). The future of sensitivity analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software*, 137, 104954. <https://doi.org/10.1016/j.envsoft.2020.104954>
- Reed, S., Koren, V., Smith, M., Zhang, Z., Morea, F., Seo, D.-J., & Dmip Participants, a. (2004). Overall distributed model intercomparison project results. *Journal of Hydrology*, 298(1–4), 27–60. <https://doi.org/10.1016/j.jhydrol.2004.03.031>
- Refsgaard, J. C., & Knudsen, J. (1996). Operational validation and intercomparison of different types of hydrological models. *Water Resources Research*, 32(7), 2189–2202. <https://doi.org/10.1029/96wr00896>
- Renard, B., Kavetski, D., Thyer, M., Kuczera, G., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, 46(5), W05521. <https://doi.org/10.1029/2009WR008328>
- Rodriguez, D. A., & Tomasella, J. (2016). On the ability of large-scale hydrological models to simulate land use and land cover change impacts in Amazonian basins. *Hydrological Sciences Journal*, 61, 1831–1846. <https://doi.org/10.1080/02626667.2015.1051979>
- Rosenberg, A., & McIntyre, L. (2019). *Philosophy of science: A contemporary introduction* (4th ed.). Routledge.
- Saft, M., Peel, M. C., Western, A. W., Perraud, J. M., & Zhang, L. (2016). Bias in streamflow projections due to climate-induced shifts in catchment response. *Geophysical Research Letters*, 43(4), 1574–1581. <https://doi.org/10.1002/2015gl067326>
- Schaeffli, B., Harman, C. J., Sivapalan, M., & Schymanski, S. J. (2011). HESS opinions: Hydrologic predictions in a changing environment: Behavioral modeling. *Hydrology and Earth System Sciences*, 15(2), 635–646. <https://doi.org/10.5194/hess-15-635-2011>
- Sheffield, J., Wood, E. F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A., & Verbist, K. (2018). Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions. *Water Resources Research*, 54(12), 9724–9758. <https://doi.org/10.1029/2017wr022437>
- Shin, M.-J., Guillaume, J. H. A., Croke, B. F. W., & Jakeman, A. J. (2015). A review of foundational methods for checking the structural identifiability of models: Results for rainfall-runoff. *Journal of Hydrology*, 520, 1–16. <https://doi.org/10.1016/j.jhydrol.2014.11.040>
- Son, K., Tague, C., & Hunsaker, C. (2016). Effects of model spatial resolution on ecohydrologic predictions and their sensitivity to inter-annual climate variability. *Water*, 8, 321. <https://doi.org/10.3390/w8080321>
- Stephens, C. M., Marshall, L. A., Johnson, F. M., Lin, L., Band, L. E., & Ajami, H. (2020). Is past variability a suitable proxy for future change? A virtual catchment experiment. *Water Resources Research*, 56(2), e2019WR026275. <https://doi.org/10.1029/2019wr026275>
- Sulis, M., Paniconi, C., Marrocu, M., Huard, D., & Chaumont, D. (2012). Hydrologic response to multimodel climate output using a physically based model of groundwater/surface water interactions. *Water Resources Research*, 48(12), W12510. <https://doi.org/10.1029/2012wr012304>
- Tague, C., & Frew, J. (2021). Visualization and ecohydrologic models: Opening the box. *Hydrological Processes*, 35(1), e13991. <https://doi.org/10.1002/hyp.13991>
- Tague, C. L., & Moritz, M. A. (2019). Plant accessible water storage capacity and tree-scale root interactions determine how forest density reductions alter forest water use and productivity. *Frontiers in Forests and Global Change*, 2. <https://doi.org/10.3389/ffgc.2019.00036>
- Thirel, G., Andréassian, V., Perrin, C., Audouy, J. N., Berthet, L., Edwards, P., et al. (2015). Hydrology under change: An evaluation protocol to investigate how hydrological models deal with changing catchments. *Hydrological Sciences Journal*, 60(7–8), 1184–1199. <https://doi.org/10.1080/02626667.2014.967248>
- Thyer, M., Beckers, J., Spittlehouse, D., Alila, Y., & Winkler, R. (2004). Diagnosing a distributed hydrologic model for two high-elevation forested catchments based on detailed stand- and basin-scale data. *Water Resources Research*, 40(1), W01103. <https://doi.org/10.1029/2003wr002414>
- Tishby, N., Pereira, F. C., & Bialek, W. (2000). The information bottleneck method. arXiv preprint physics/0004057.
- Tsamir, M., Gottlieb, S., Preisler, Y., Rotenberg, E., Tatarinov, F., Yakir, D., et al. (2019). Stand density effects on carbon and water fluxes in a semi-arid forest, from leaf to stand-scale. *Forest Ecology and Management*, 453, 117573. <https://doi.org/10.1016/j.foreco.2019.117573>
- VanderKwaak, J. E. (1999). Numerical simulation of flow and chemical transport in integrated surface-subsurface hydrologic systems. PhD thesis. University of Waterloo.
- van Roosmalen, L., Sonnenborg, T. O., & Jensen, K. H. (2009). Impact of climate and land use change on the hydrology of a large-scale agricultural catchment. *Water Resources Research*, 45(7), W00A15. <https://doi.org/10.1029/2007wr006760>
- von Neumann, J. (1955). Method in the physical sciences. In L. Leary (Ed.), *Collected works, The unity of knowledge*, (Vol. 6, pp. 157–164). Doubleday. Retrieved from <https://www.docdroid.net/fInvX04/methodinthephysicalsciences-pdf>
- Wang, S., Kang, S., Zhang, L., & Li, F. (2008). Modelling hydrological response to different land-use and climate change scenarios in the Zamu River basin of northwest China. *Hydrological Processes*, 22(14), 2502–2510. <https://doi.org/10.1002/hyp.6846>
- Watson, T. A., Doherty, J. E., & Christensen, S. (2013). Parameter and predictive outcomes of model simplification. *Water Resources Research*, 49(7), 3952–3977. <https://doi.org/10.1002/wrcr.20145>
- Weiler, M., & McDonnell, J. (2004). Virtual experiments: A new approach for improving process conceptualization in hillslope hydrology. *Journal of Hydrology*, 285(1–4), 3–18. [https://doi.org/10.1016/S0022-1694\(03\)00271-3](https://doi.org/10.1016/S0022-1694(03)00271-3)
- Weinberg, G. M. (2001). *An introduction to general systems thinking (silver anniversary ed.)*. Dorset House Publishing Co., Inc.
- Welsh, W. D., Vaze, J., Dutta, D., Rassam, D., Rahman, J. M., Jolly, I. D., et al. (2013). An integrated modelling framework for regulated river systems. *Environmental Modelling & Software*, 39, 81–102. <https://doi.org/10.1016/j.envsoft.2012.02.022>
- Yang, J., Jakeman, A., Fang, G., & Chen, X. (2018). Uncertainty analysis of a semi-distributed hydrologic model based on a Gaussian Process emulator. *Environmental Modelling & Software*, 101, 289–300. <https://doi.org/10.1016/j.envsoft.2017.11.037>
- Yang, Y., & Chui, T. F. M. (2021). Reliability assessment of machine learning models in hydrological predictions through metamorphic testing. *Water Resources Research*, 57(9), e2020WR029471. <https://doi.org/10.1029/2020wr029471>
- Zierl, B., Bugmann, H., & Tague, C. L. (2007). Water and carbon fluxes of European ecosystems: An evaluation of the ecohydrological model RHESSys. *Hydrological Processes*, 21(24), 3328–3339. <https://doi.org/10.1002/hyp.6540>