WIREs WATER | WILEY

On doing hydrology with dragons: Realizing the value of perceptual models and knowledge accumulation

Thorsten Wagener[1,2,3] | Tom Gleeson[4,5] | Gemma Coxon[2,6] |
Andreas Hartmann[1,7] | Nicholas Howden[1,2] | Francesca Pianosi[1,2] |
Mostaquimur Rahman[1,2] | Rafael Rosolem[1,2] | Lina Stein[1] | Ross Woods[1,2]

[1]Department of Civil Engineering, University of Bristol, Bristol, UK

[2]Cabot Institute for the Environment, University of Bristol, Bristol, UK

[3]Institute for Environmental Science and Geography, University of Potsdam, Potsdam, Germany

[4]Department of Civil Engineering, University of Victoria, Victoria, British Columbia, Canada

[5]School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, Canada

[6]School of Geographical Sciences, University of Bristol, Bristol, UK

[7]Institute of Earth and Environmental Sciences, University of Freiburg, Freiburg, Germany

Correspondence
Thorsten Wagener, Department of Civil Engineering, University of Bristol, Bristol, UK.
Email: thorsten.wagener@bristol.ac.uk

Abstract

Our ability to fully and reliably observe and simulate the terrestrial hydrologic cycle is limited, and in-depth experimental studies cover only a tiny fraction of our landscape. On medieval maps, unexplored regions were shown as images of dragons—displaying a fear of the unknown. With time, cartographers dared to leave such areas blank, thus inviting explorations of what lay beyond the edge of current knowledge. In hydrology, we are still in a phase where maps of variables more likely contain hydrologic dragons than blank areas, which would acknowledge a lack of knowledge. In which regions is our ability to extrapolate well developed, and where is it poor? Where are available data sets informative, and where are they just poor approximations of likely system properties? How do we best identify and acknowledge these gaps to better understand and reduce the uncertainty in characterizing hydrologic systems? The accumulation of knowledge has been postulated as a fundamental mark of scientific advancement. In hydrology, we lack an effective strategy for knowledge accumulation as a community, and insufficiently focus on highlighting knowledge gaps where they exist. We propose two strategies to rectify these deficiencies. Firstly, the use of open and shared perceptual models to develop, debate, and test hypotheses. Secondly, improved knowledge accumulation in hydrology through a stronger focus on knowledge extraction and integration from available peer-reviewed articles. The latter should include metadata to tag journal articles complemented by a

common hydro-meteorological database that would enable searching, organizing and analyzing previous studies in a hydrologically meaningful manner.

This article is categorized under:
    Engineering Water > Planning Water
    Science of Water > Hydrological Processes
    Science of Water > Methods

KEYWORDS
large-scale hydrology, uncertainty, perceptual model, metadata, machine learning

1 | INTRODUCTION

Humanity has always been uncomfortable with knowledge gaps. When the explorer John Cabot left Bristol harbor in 1497 to chart a new route to Asia, he was trying to fill a knowledge gap of European geographers (and filled a different one—the existence of North America—by accident). Such attempts to meet key knowledge gaps had not been common at the time, a circumstance nicely visualized in the example of two maps. In his macro-history of humanity, Harari (2015) discusses two world maps produced by European cartographers that, while being drawn within less than 100 years of each other, were separated by key events including the voyages of John Cabot and Christopher Columbus (Figure 1). The Fra

FIGURE 1 (a) The Fra Mauro world map (Italy, 1459) shows seemingly complete knowledge of the world. However, on closer inspection, maps like this one included statements like HIC SUNT LEONES and images of monsters in unexplored regions. (b) The Salviati Planisphere is a world map (Spain, 1525) without imagined representations, highlighting knowledge gaps. *Source*: Wikipedia, https://en. wikipedia.org/wiki/World_map (accessed March 2020)

Mauro world map from 1459 shows the world in great detail with hundreds of illustrations and thousands of descriptive texts (Figure 1a), which is surprising given the poor state of knowledge about the world at the time. Looking more closely, however, one finds that this one and other maps from the period included imagined representations in unexplored (unknown) areas. Indeed, it was customary practice at the time to show monsters rather than to leave spaces empty—suggesting that it is undesirable and dangerous to explore such places. Ancient Roman and Medieval mapmakers demarcated such unknown areas with the phrase *HIC SUNT LEONES* ("here are lions") or alternatively with HIC SUNT DRACONES ("here are dragons"), thus populating unknown areas with creatures that would instill fear in the reader (Agostinho et al., 2019). Fewer than 100 years later the Salviati Planisphere map from 1525 (Figure 1b) not only shows the newly found eastern coasts of North and South America, but also reveals empty space in thus far unexplored areas. The 16th-century map reveals where knowledge gaps exist by daring to leave such areas blank, thus inviting explorations to discover what lies beyond the edge of current knowledge. Rather than fearing ignorance, this step of acknowledging the unknown became a scientific goal in itself—an important scientific belief in modernity.

In hydrology, we have seen our knowledge and analysis domain expand from the catchment to continental and even global scales, increasingly with the help of large-scale hydrologic simulation models. Global hydrologic models have emerged, not just any more as the endeavors of particularly brave scientists (e.g., Manabe & Holloway, 1975), but rather as tools for regular scientific analysis and increasingly even as potential tools for water resource management (Archfield et al., 2015; Bierkens et al., 2015; Straatsma et al., 2016). Hydrology is indeed moving toward realizing elements of the "models of everywhere" idea of Beven—that is, modeling as a learning process (Beven, 2001). At the same time, there is the hope for more realistic process representations through hyper-resolution models (Clark et al., 2017; Maxwell & Condon, 2016; Wood et al., 2011), though the debate regarding the best way to represent the physics of hydrologic processes in our models continues (Beven et al., 2015). Global and continental-scale hydrologic models increasingly reveal human influence on global fluxes of terrestrial sediments to the oceans (Syvitski et al., 2005), risks to global river biodiversity (Vörösmarty et al., 2010), global depletion of groundwater resources (Wada et al., 2010), the relative impacts of groundwater pumping on our rivers (De Graaf et al., 2019), global drivers of flood risk (Winsemius et al., 2016), impacts of human activity on the global water cycle (Bosmans et al., 2017), and potential implications of climate change for the global freshwater system (Döll et al., 2018).

Equally, we see the emergence of large-scale and even global data sets which add new dimensions to our ability to analyze global hydrology (Beck, van Dijk, et al., 2019; Ghiggi et al., 2019; Lindersson et al., 2020). New global data sets of physical system properties, such as subsurface hydrogeological properties, have been accumulated in recent years (Gleeson et al., 2011; Huscroft et al., 2018). Various global precipitation data sets can be used to force hydrologic simulation models (Beck, van Dijk, et al., 2019), while land surface fluxes have been regionalized from flux tower networks to the global scale (Jung et al., 2011). Satellite data provide continuous time estimates of freshwater storage (Rodell et al., 2018; Famiglietti et al., 2015) and the topographic characteristics of our land surface in a hydrologically meaningful manner (Linke et al., 2019; Nardi et al., 2019; Yan et al., 2019), while remotely sensed observations of vertical fluxes are used to create land surface water balances everywhere (Miralles et al., 2011). Increasingly, new global data sets of land cover change and other human interventions in the water cycle are becoming available, such as artificial storage through reservoirs (Lehner et al., 2011; Mulligan et al., 2020). Widely used historical data such as river flows of the Global Runoff Data Centre (Grabs et al., 1996) are complemented with hydrologic response data compilations like karst spring hydrographs (Olarinoye et al., 2020). However, all of these data sets have their own issues and limitations as we stress further below.

The potential power of such large-scale data sets is maybe best exemplified by the tremendous success of machine learning and other data-based approaches, which often outperform simulation models that are based on our mechanistic understanding of how nature works (Kratzert, Klotz, Shalev, et al., 2019; Reichstein et al., 2019; Shen et al., 2018). For example, Boers et al. (2019) find teleconnection patterns in global extreme rainfall by analyzing high-resolution satellite data with the help of complex networks. Stolbova et al. (2016) managed to empirically predict the onset of the monsoon two weeks further in advance than previous methods (including predictions from dynamic models). Addor et al. (2018) used random forests to demonstrate the predictability of hydrologic signatures across the United States and showed that signatures predictable from descriptors which vary smoothly in space, such as those related to climate, regionalize particularly well. Machine learning tools have also been used widely to turn in-situ observations into global data sets. Jung et al. (2011) used a machine learning strategy—model tree ensembles—to upscale Fluxnet observations of carbon dioxide, water, and energy fluxes to the global scale. The ability of deep learning models to predict streamflow in gauged and ungauged catchments has already been demonstrated (Kratzert, Klotz, Herrnegger, et al., 2019; Nearing et al., 2021).

So, what is the link between dragons and these aspects of hydrology? Relevant hydrologic quantities and properties as well as their uncertainties tend to be poorly determined by scarce historical observations away from areas

where observations are concentrated (Beven et al., 2020). The Predictions in Ungauged Basins (PUB) initiative has given us many examples that demonstrate this problem in the context of streamflow predictions (Hrachowitz et al., 2013), and the study of hydrologic extremes has shown that much of the landscape we think we know is rather akin to a terra incognita—unmapped regions (B. Merz et al., 2015). However, we also find that our models sometimes seamingly work well in some regions without local calibration (Van Werkhoven et al., 2009), so it is not simply a question of well-studied regions versus poorly studied ones. Large-scale or even global hydrology brings this problem even more strongly to the forefront. In our endeavor to move to larger scales, we unavoidably move hydrologic investigation away from highly studied headwater catchments, hillslopes, aquifers, or Fluxnet sites to regions of poorly explored and poorly characterized landscapes—relying on our ability to extrapolate with models instead (Fan et al., 2019). But this presents a major challenge: How can we deal with such knowledge gaps or epistemic uncertainties apart from hoping for potential new future measurement techniques? Large-scale outputs of hydrologic simulation models or compiled data sets will unavoidably include hydrologic dragons rather than meaningful information in some places. So, similar to map makers who drew the Salviati Planisphere map from 1525, we need to start distinguishing these areas so that they can be highlighted as "blank," that is in need of further exploration. This is of course not a question of information versus no information (truly blank space), because the space of possible hydrologic behavior or system properties is not completely unconstrained anywhere in the world (Beven, 2001; Kirchner, 2006; Wagener & Montanari, 2011). It is rather a question of how much can we know, what is the basis for this knowledge, and how confident are we in our knowledge?

So, while the combination of global models (both data-based and mechanistic) and global data sets undoubtedly offers tremendous opportunities for scientific advancement and for new scales of management, it also contains hydrologic dragons, that is, knowledge gaps that are currently difficult to identify and address. Our observations of hydrologic fluxes and storages do not allow us to characterize the water balance (especially subsurface properties) at the above-mentioned hyper-resolution (Beven et al., 2015; Beven & Cloke, 2012). Different assumptions (poorly constrained by available information) still leave the potential for different conclusions when using the same data, for example, that catchment-scale water balance errors are due to precipitation error (Beck, Wood, et al., 2019) or due to subsurface losses (Liu et al., 2020). Global geological data show artifacts such as variability along administrative boundaries due to differences in processing underlying observations across administrative units (Gleeson et al., 2011). Pedotransfer functions based on soil texture classes, the basis for estimating soil hydraulic properties used in many hydrologic and other models, have been derived from very limited and biased empirical data, while ignoring structural soil characteristics (or 2019; Fatichi et al., 2020). So, it is not too surprising that Gutmann and Small (2007) find soil texture only explains a small fraction (5% in their case) of the expected variability of soil hydraulic properties. Similarly, Rosero et al. (2010) found that behavioral soil and vegetation parameters, derived as parameter sets through conditioning of the NOAH land surface model to observations from flux towers along a climatic gradient with varying soil and vegetation properties, correlated with the climatic gradient, but not with soil or vegetation properties. These examples put a critical focus on the transfer algorithms used to translate the actual measurement into hydrologically meaningful information. Most (if not all) of our data sets are based on measurements subsequently processed through models for interpretation, interpolation or extrapolation (Gupta & Nearing, 2014). They are therefore associated with their own (and often significant) uncertainties with potentially significant consequences for subsequent use as we will discuss further below (Kauffeldt et al., 2013; Yatheendradas et al., 2008). For example, J. Yang et al. (2013) provide a good discussion of the uncertainties stemming from remote sensing measurements, as well as from the subsequent algorithms needed to transform these measurements into the relevant variables for climate science. How often do we consider or even just acknowledge the uncertainties in this processing chain where possibly unknown corrections are made or poorly defined parameters are used in the algorithms? Are these data post-processing models sufficiently realistic and do we sufficiently acknowledge this lack of realism where and when it occurs?

So how do we identify hydrologic dragons, and, more importantly, how might we overcome them given that they originate from a lack of observations or even observational capability? In this brief commentary, we discuss two strategies to address this question.

- A focus on perceptual models to pool and test our knowledge.
- Improved knowledge accumulation in hydrology.

Below we discuss each of these strategies and what role they play.

2 | PERCEPTUAL MODELS TO POOL AND TEST OUR KNOWLEDGE AND EXPERIENCE

We propose that a currently underutilized strategy to collect and share information (defined here as "data seen in a particular context") as well as knowledge (defined here as "our understanding gained through experience") valuable for large-scale modeling and hydrology in general lies in openly shared and jointly evolved perceptual models. Perceptual models in hydrology are defined by the evolving understanding of real-world system based on the interpretation of all available information, influenced by each hydrologist's unique experience and training (Beven, 2001; Gupta et al., 2008; Gupta et al., 2012; Seibert & McDonnell, 2002; Tetzlaff et al., 2008). Sometimes, perceptual models are seen as one step in a modeling chain where they form the basis of more formal system conceptualizations, for example, in hydrology (Beven, 2001) and hydrogeology (Brassington & Younger, 2010). Here we use the term perceptual model as the (typically visual) representation of the hydrologist's understanding of the system, including her subjective understanding, speculation and opinion, but without any specific consideration of subsequent simulation model building efforts (e.g., whether the subsequent simulation model is spatially lumped or distributed). We do not, of course, claim that any modeler would build and apply large-scale hydrologic model without perceptual models as a baseline, or that an experimentalist does not have a perceptual model in mind when placing their instruments (Figure 2): the issue is rather one of publishing and sharing such models so that differences in the interpretation of available information about the hydrology of a place becomes visible, can be debated, and can be addressed.

Large-scale hydrologic simulation models depend on appropriate data sets to define their parameters (and potentially model structure). Workflows to integrate simulation models and data have become increasingly sophisticated and efficient (Turuncoglu et al., 2013; Leonard and Duffy, 2014; Leonard and Duffy, 2016; Blair et al., 2019). Leonard and Duffy (2013), for example, developed workflows through which 100 s of terabytes of US data sets organized at the United States Geological Survey Hydrological Unit Code Level 12 scale can be used to parameterize a version of their watershed model anywhere in the continental United States. Workflows like these facilitate the integration of models
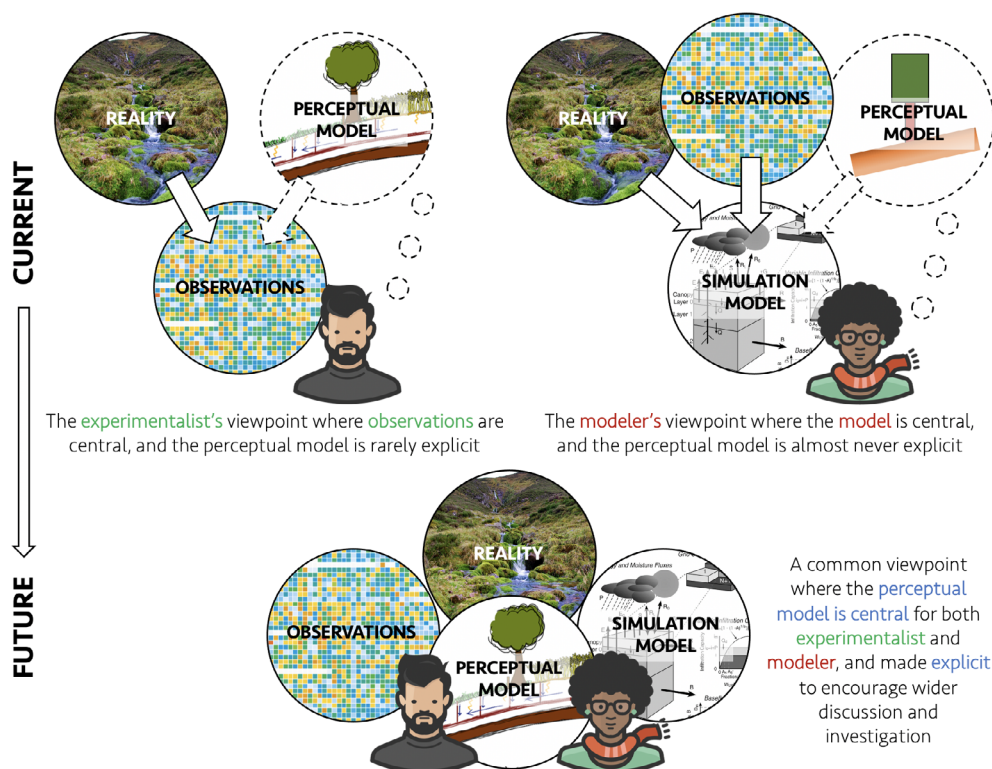

The experimentalist's viewpoint where observations are central, and the perceptual model is rarely explicit

The modeler's viewpoint where the model is central, and the perceptual model is almost never explicit

A common viewpoint where the perceptual model is central for both experimentalist and modeler, and made explicit to encourage wider discussion and investigation

FIGURE 2 We suggest moving the perceptual model from its often implicit side-role, to an explicit central role for both experimentalists and modelers, that is, for both deciding where and what to measure, as well as how to simplify reality in our simulation models (Perceptual model picture taken from Loritz et al., 2017; reality picture taken by Fabian Nippgen; model image from Cherkauer et al., 2003; character icons are by iconify obtained from iconfinder.com under the creative commons license)
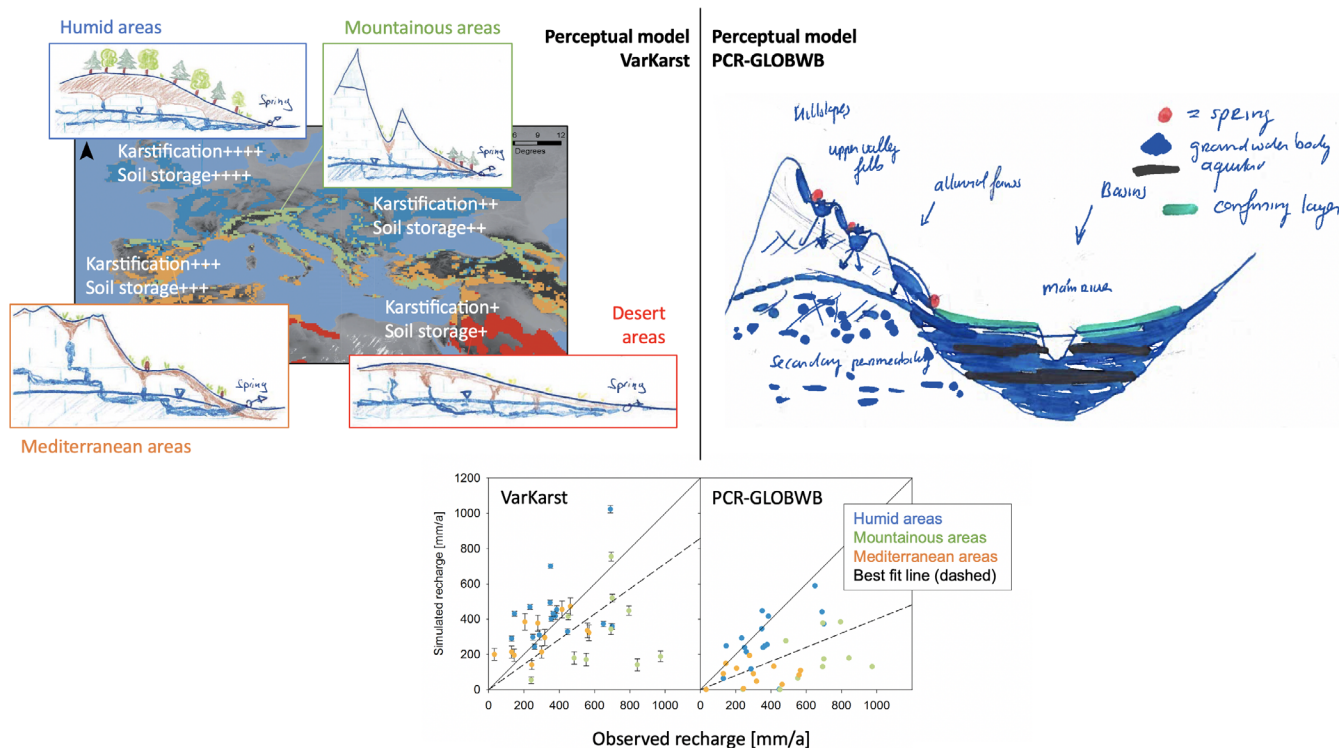
FIGURE 3 Two perceptual models of large-scale hydrologic models are shown at the top. On the left are four perceptual models of the carbonate rock regions across Europe/North Africa/Middle East by Hartmann et al. (2015). The different perceptual models are derived based on the expected differences between carbonate rock regions, including relative differences in the degree of karstification and the amount of storage present. The top right figure is the perceptual model underlying the global hydrological model PCR-GLOBWB (©Marc Bierkens). The bottom graphs shows how the differences in perceptual models propagate into differences in recharge predictions of the simulation models (Hartmann et al., 2015)

and data, leading to a strong focus on available data. A priori models (i.e., without subsequent calibration) built in this manner are the basis for much of our global hydrology (Y. Yang et al., 2019).

Relying heavily on the integration of data sets creates problems when and where currently available data sets are poor descriptors of the underlying hydrologic processes, that is, where they are mere hydrologic dragons. It also ignores knowledge (or experience) that has been gained but is not easily embedded in data sets, thus potentially ignoring knowledge which would significantly alter our predictions. Hartmann et al. (2017) compared groundwater recharge estimates of two large-scale models, PCR-GLOBWB and VarKarst-R across the carbonate rock regions of Europe, Northern Africa, and the Middle East (Figure 3). The former is a global integrated hydrologic model based on global data sets, while the latter is a parsimonious model tailored to regions with strongly focused recharge processes. The authors found that recharge estimates of the simpler model were more consistent with available observations and local model results. One reason for this result is that the underlying perceptual model for VarKarst-R is based on the expected dominant system characteristics of carbonate rock regions, derived from experience derived other in highly studied locations. The perceptual model of PCR-GLOBWB (at the time) assumed that the world consisted only of two systems: mountains and alluvial plains—the remaining tailoring was done through adjustment of its parameters (Prof Mark Bierkens, personal communication). The simpler model used in-depth knowledge gained through local studies to develop different perceptual models for a key hydrologic domain (carbonate rock regions), which were then further constrained using similarity principles expressed as behavioral rules. How can we formalize the integration of available knowledge through perceptual models?

While many journal papers, for example, those describing modeling studies, will include a schematic depiction of a simulation model, few include the underlying perceptual model of the system that the modeler had in mind. The schematic of the simulation model typically includes assumptions related to the implementation choices of the modeler, for example, they might select a spatially lumped or a grid-based model. However, this schematic might be quite far removed from the underlying system perception the modeler started with. Any perceptual model will

likely be much more complex than the subsequent simulation model we execute on a computer—how we simplify the former to reach the latter is one of the most exciting aspects of hydrological modeling (Beven & Chappell, 2021). We as a community have not, so far, systematically collected and used the knowledge provided in perceptual models for hydrologic modeling. Open perceptual models would provide a forum to discuss and challenge our current thinking about the dominant hydrologic processes of different places, about hydrologic connectivity, about boundary conditions, and so on. Even if we cannot agree on system properties and behavior in absolute terms, they might offer us an opportunity to discuss the relative difference between places (e.g., evapotranspiration rates should be higher in this place than that one, this catchment should respond faster than the next, or subsurface storage should be larger in this system than another one)—which might already significantly improve our understanding of dominant process controls and whose value is likely underrated in hydrological investigations (Rogger et al., 2012). Perceptual models also help us to make (unavoidable) subjective choices transparent by bringing them out into the open.

These perceptual models also directly relate to the wider problem of transferring knowledge that has been gained in a specific catchment or location to other (even seemingly similar) places. As McDonnell et al. (2007, p. 2) conclude that: "As a community, and as individuals, we have progressed along a philosophical path that 'if we characterize enough hillslopes and watersheds around the world through detailed experimentations, some new understanding is bound to emerge eventually.'" As the authors acknowledge, such a reductionist approach has not led to the transferrable knowledge we aim to find in hydrology (Dooge, 1986). How can we complement this focus on understanding individual places with one of understanding regional scale variability in a structured manner? Put another way, how can we construct our hydrologic knowledge landscape so that it transcends the uniqueness of place and we increase our chance of understanding whether hydrologic similarity might exists and at what scales—preferably utilizing advancements in computational science to capture and share knowledge landscapes (Gil et al., 2019)? Investigating individual catchments in depth unavoidably confronts us with high levels of complexity (Tetzlaff et al., 2008) and unique features that distinguish one particular catchment from another (Beven, 2000). However, at some higher level, we continue to assume that principles of hydrologic similarity apply and are helpful for regionalization, classification and thus organization of hydrologically relevant spatial units (McDonnell & Woods, 2004; Wagener et al., 2007). Part of the problem is a lack of hydrologically meaningful descriptors of catchments (or other hydrologic units of relevant size). Climatic and topographic catchment descriptors have been assessed widely (e.g., Seibert & McGlynn, 2007; Knoben et al., 2018), and they have been shown to be valuable predictors of some hydrological responses (e.g., Addor et al., 2018; Kuentz et al., 2017; Troy et al., 2008). Subsurface characteristics on the other hand are much harder to observe and characterize (Addor et al., 2018; Beven & Cloke, 2012; R. Merz et al., 2020), and therefore might have to be more strongly based on our expectations than just on directly observable properties. Few attempts to integrate (expected) system conceptualizations and data have been made thus far (Boorman et al., 1995; Enemark et al., 2019) and opportunities for improvement remain (Wagener et al., 2021). How we better characterize hydrologic units across scales meaningfully is unclear, though there is a danger that the subsequent stage of utilizing such data is the more appealing due to the advancements in machine learning and the ease with which we can build models today.

Open and structured discussions of perceptual models might reveal divergent expectations of the dominant hydrologic processes in particular places, potentially even before we have taken in situ measurements. Revealing such divergences would help us to identify hydrologic dragons and provide the basis for developing hypotheses to be tested, so that candidate perceptual models can be rejected. It could also support the development and testing of increasingly powerful deep learning models, thus connecting the process-based modeling community with the data-based modeling community (Nearing et al., 2021). A starting point for a global set of perceptual models could be the previously proposed simple perceptual models of comparative hydrology (Falkenmark & Chapman, 1989), which would nonetheless require much more tailoring to each location using top-down thinking already applied in many modeling studies (Sivapalan et al., 2003; Young, 2003). How much can we reduce model prediction uncertainty if we do not just constrain the expected hydrological behavior with available data, but also using our expectations across large scales (Hartmann et al., 2015; Sarrazin et al., 2018)? These ideas of using qualitative information more rigorously is of course building on previous suggestion at the catchment scale discussed, for example, by Seibert and McDonnell (2002), Savenije (2010), or Kelleher et al. (2013). Some studies have shown that simpler measurements, sensibly distributed in space and time, might provide insight that is more transferable than that produced by much more in-depth measurements that can only be made in very few places (Jencso & McGlynn, 2012).

3 | IMPROVED KNOWLEDGE ACCUMULATION IN HYDROLOGY

The philosophy of science underlying most of hydrology is based on the process of scientific evolution proposed by Popper (1959) where hypotheses are falsified through evidence (data) and remain conditionally valid only as long as they are consistent with all available evidence. The approach by Popper—one of hypothesis testing—is the strategy often utilized (or at least attempted) in hydrology (though some argue that this has not been done very well; Pfister & Kirchner, 2017; Beven, 2018; Beven & Chappell, 2021). Another idea, less frequently discussed, is that knowledge accumulation itself constitutes scientific advancement in its own right—rather than just being a component in the hypotheses falsification style scientific processes mentioned above. Does science inherently advance if we accumulate knowledge? Or only if this knowledge leads to the development of new ideas or theories? Bird (2007, 2008) suggests that: "Science (or some particular scientific field or theory) makes progress precisely when it shows the accumulation of scientific knowledge; an episode in science is progressive when at the end of the episode there is more knowledge than at the beginning." While Bird was not the first to suggest this concept for scientific progress, which can be traced back to Francis Bacon, he nonetheless reenergized the discussion (Mizrahi, 2013). Here, we do not want to answer the question whether knowledge accumulation is equal to scientific progress or not, but rather stress that effective knowledge accumulation is a fundamental element for scientific progress and for tackling hydrologic dragons.

We argued in Section 2 that perceptual models are one way toward reducing knowledge gaps (dragons) in large-scale hydrology. So how and where is knowledge—to build up these perceptual models—currently captured in the field of hydrology? And especially, how well does it accumulate? Given that we regularly ask what questions remain in hydrology (Sivapalan, 2009; Blöschl et al., 2019), it seems equally relevant to ask what we already know and how confident we are that all available knowledge has been captured. An exhaustive case study of how the hydrologic community has accumulated knowledge was the synthesis effort within the PUB initiative (Blöschl et al., 2019; Hrachowitz et al., 2013). Over 100 authors produced a compendium of what had been learned about the PUB problem—often rerunning analyses to make them consistent and comparable. While this work is a good example of what can be achieved through such a community effort, it used more resources than those normally available. Different—more sustainable and routine—strategies to achieve such a synthesis are needed to address our hydrologic questions (Blöschl et al., 2019).

Knowledge about the hydrology of different places is shared mainly through peer-reviewed journal papers. While this meant reading a few hundred papers per year in the 1970s, it now requires checking in excess of 3000 papers in 2020 alone (more than 8 per day), even if we only focus on the main hydrology journals (Figure 4). Many, if not most, of these papers will describe what has been learned by studying the hydrology of a particular place, or a collection of places, thus providing insight into the hydrologic variability found in our highly heterogeneous world (Beven, 2000).
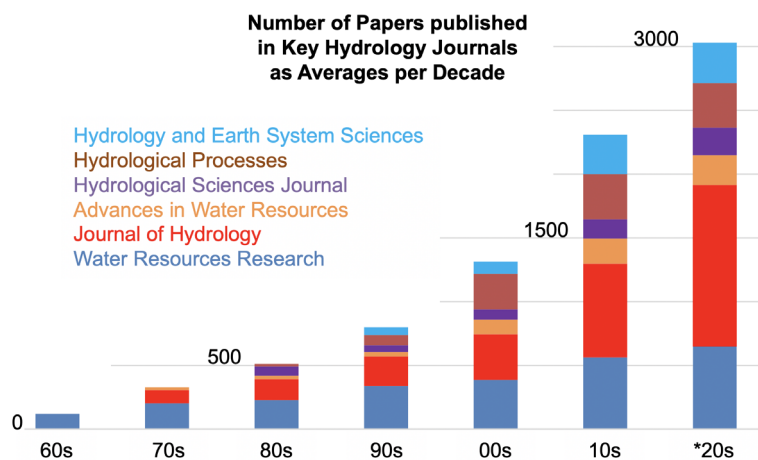

FIGURE 4 Number of papers listed in Web of Science (March 2021) for the different journals. Papers are accumulated by decade from the 1960s to the 2010s (even if the journal started sometime during a decade). Asterisk shows that the year 2020 only reflects that year. The journals and their individual decadal numbers are as follows: *Water Resources Research* (119, 202, 227, 338, 385, 563, 647), *Journal of Hydrology* (−, 103, 166, 231, 359, 735, 1266), *Advances in Water Resources* (−, 23, 26, 36, 115, 195, 234), *Hydrological Science Journal* (−, −, 72, 52, 79, 150, 217), *Hydrological Processes* (−, −, 21, 83, 281, 358, 306), *Hydrology and Earth System Sciences* (−, −, −, 60, 94, 306, 315). *Hydrological Sciences Journal* was searched as "Hydrological Sciences Journal – Journal des Sciences Hydrologiques"

The existing meta-analyses show the great potential for learning by reviewing and synthesizing the existing literature (e.g., Evaristo & McDonnell, 2017; Price, 2011). Review papers play an important role as well, regardless of whether they are published in our main journals or in journals which specialize on reviews. In either case, problems include that reviews regularly cannot consider a large fraction of the papers in a particular topic (given the sheer number), and thus more likely propose a new organization of available knowledge with a limited number of papers as examples, that is, they are more qualitative reviews (which of course does not mean that they are not useful!). In response to this issue, some journals (e.g., *Environmental Research Letters*) specifically advocate more quantitative, metadata-driven reviews. These quantitative, metadata-driven reviews are a useful complement to more qualitative reviews and require an easier way to identify and organize the existing literature for synthesis. Given the growing complexity of reviewing any particular area of hydrology, it is maybe surprising that hydrology has not yet adopted more formalized strategies to literature reviews as is common in other fields—for example, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses methodology (Moher et al., 2009). Requiring authors to develop and explicitly communicate a strategy for their reviews would likely also be useful in hydrology. While a full systematic review might be too complex for a standard paper, the current lack of strategy often leads to poorly grounded studies and therefore unclear contributions.

We might also look at textbooks, field trips, or graduate supervisors for this knowledge. In contrast to journal papers, most hydrology textbooks provide the underlying hydrologic theory, but the tailoring to a specific place remains difficult to encapsulate in general guidelines (Wagener & Montanari, 2011). Field trips are a tangible and impactful way of sharing knowledge, and new ways for sharing our experience with understanding the hydrology of specific catchments or regions are emerging. For example, Google Earth Engine and other virtual earth educational tools provide exciting opportunities for "virtual field visits" in combination with those field trips that can be done locally. All students will be offered some level of accumulated knowledge through their supervisors′ experience. However, each supervisor shares their subjective knowledge landscape and can only reach a small number of students, which creates a heterogeneous and dissatisfying baseline for our scientific community. It is important to have exceptional researchers who provide inspiration and creativity, but it would be beneficial to have a better general baseline for our accumulated knowledge.

We need a better way to find, extract, and accumulate the hydrological knowledge hidden in over 2500 papers published per year if we want to have a chance to tackle hydrologic dragons at the global scale. Hradec et al. (2019) call this the challenge of assessing information "trapped in the text". "The sheer volume of text means that, unassisted, we cannot hope to read all available sources, nor even to keep up to date with all advances in a particular field" (Hradec et al., 2019). For example: How many studies last year analyzed the water balance of catchments located in the subtropics? Or, have the Nash Sutcliffe Efficiency values of models applied to semi-arid catchments improved over the last decade? We can currently only answer these questions by manually looking through large numbers of papers in a tedious manner. In other fields, it has been recently proposed to approach this problem through machine learning, which has been shown to be able to extract scientific knowledge hidden in scientific papers (Hradec et al., 2019; Kumar et al., 2018; Tshitoyan et al., 2019). Tshitoyan et al. (2019) demonstrate what the automated mining of scientific literature can achieve by showing, for example, that structure–property relationships in materials can be derived from information gathered in this manner. Hradec et al. (2019) developed a software tool to perform a similar automated semantic analysis of a large number of documents to support European policy making. The recent need for fast progress in studying COVID-19 has further pushed the development and use of such methods (Khanday et al., 2020). While such approaches are not yet widely explored, hydrology, which offers a large number of historical studies on a very diverse hydrologic systems (at catchment or other scales), is exceptionally well placed to test its potential.

Regardless of the advancements made with text mining algorithms, the hydrological community needs to advance how knowledge synthesis is supported. The PUB effort (Blöschl et al., 2013), for example, has shown how tremendously difficult it is to simply identify everybody who has simulated a particular catchment in the past, and to ask how well their hydrologic model performed. A starting point to organize our articles in a hydrologically relevant manner would be the inclusion of mandatory metadata for each hydrologic study in each journal article as *Essential Hydrological Descriptors* in addition to the standard key words or subject tags. These could include geolocation, time period, spatial and temporal resolution, and extent, fluxes, and stores studied at the hydrologic study area. Requiring additional information that has to be provided by the authors has to be balanced with the effort it takes to provide this information. However, such metadata tagging is already done in some data journals such as *Scientific Data* which includes machine-readable metadata on location, time period etc. for every article. We suggest that this needs to be expanded to all journal articles published in hydrology (unless they are purely theoretical) so that the identification and synthesis of studies, for example, for a particular location and time period, is highly simplified. We would also need to add such machine-

readable metadata tagging retrospectively to the many articles already published, so that we do not lose the information stored there and to continue to utilize what has been learned.

Simply organizing all published hydrologic studies on a particular topic, for example, flooding or groundwater recharge rates, by geolocation, would enable us to see where time and space clusters of studies exist and which studies we can compare for consistency of conclusions. More importantly, it would highlight actual blank spaces on the map of global hydrology, showing which catchments or regions have never been studied locally. Clearly the regions of Europe and North America will be densely populated with study locations, but how many places in the developing world have never been studied in the peer-reviewed scientific literature, or have actually been studied, but their results are published in papers that are unfortunately often rather poorly cited (maybe because they are not published in the main hydrology journals or published in a Red Book of the International Association of Hydrological Science)?

Such metadata should describe characteristics that are unlikely to change—such as the geolocation and time period covered by a study. They could form the basis for developing suitable further descriptors to analyze such studies during syntheses. Imagine a *Web of Hydrology* (rather than a Web of Science) where the papers (identified by the DOI and their metadata) are connected with hydrologically relevant information that would be calculated from one or more common global data sets—check out the https://www.isipedia.org idea in this regard. Such descriptors could include, for example, climate descriptors to group the existing studies not just by location, but by the similarity of the climate they were performed in (e.g., all studies performed in cold arid regions). The community debate around what climate (or topographic or geologic or ...) descriptors should form the basis for organizing our hydrology by itself would be a very interesting and relevant study by itself (e.g., discussions by Winter, 2001; McDonnell & Woods, 2004; Buttle, 2006; Wagener et al., 2007). Including multiple data sets would enable at least a basic assessment of the uncertainty in how well we can characterize a place and a time period. This database would also slowly grow through efforts to extract and submit hydrologically relevant information from journal papers such as groundwater recharge estimates or the performance of a hydrologic model applied to a particular catchment. It would take a community-scale effort to make such a *Web of Hydrology* happen and widely utilized (e.g., similar to those during PUB; Wood et al., 2005).

A key question is of course how we would motivate and incentivize the community to do so. Adding additional metadata to future papers would simply be a requirement by the journal. To add these to historical articles might require a paid activity maybe even done by nonhydrologists like in Mechanical Turk (https://www.mturk.com), which is a crowdsourcing marketplace where an outsourced virtual workforce can perform tasks. The focus lies on outsourcing tasks that humans can still perform better than computers, such as doing research (see example application by Bonnefon et al., 2016). Another interesting model for motivation might be to gamify this activity as well as the subsequent analysis. The idea is to use techniques borrowed from game-development to motivate consistent participation and long-term engagement. A nice example of gamification is the Moral Machine (moramachine.net). Users are asked to make decisions about moral dilemmas. For example, a self-driving car is approaching a pedestrian crossing and a break failure occurs. The car can either stay in the lane and likely kill a child, or it can swerve and likely kill an adult. What should the self-driving car do? The Moral Machine online experiment collected 40 million decisions from people in 233 countries and territories (Awad et al., 2018). Results showed, for example, that there are strong preferences for sparing humans over animals, for sparing more lives over fewer and for sparing young people over older people.

4 | CONCLUSIONS

The outputs of regional- to global-scale inquiries in hydrology unavoidably contain hydrologic dragons, that is, regions where the uncertainty in expected hydrologic behavior or relevant system properties is very high due to a lack of local knowledge. Our observations are often too sparse, our data sets are not equally valid everywhere due to the empirical post-processing models they are based on, and we cannot "see" key processes due a lack of measurement capability. All of these points are simply statements of the present state of our science, and we are not the first to point them out. The wider problem is that few studies highlight such knowledge gaps where they exist, how they propagate into model outputs, and what their consequences are for our conclusions. We drew a comparison with cartography in the 15th/16th century where cartographers shifted from filling all parts of the map to being content with leaving large areas blank. Blank spaces represented significant knowledge gaps that invited exploration. Highlighting knowledge gaps became a key outcome, rather than something to hide. How many large-scale maps of hydrologic model predictions or hydrologic data products have been published with blank areas highlighting knowledge gaps, for example, regions where the model does not reach specific predictive benchmarks—using local data or regionalized information (Seibert et al., 2018;

Wagener & Montanari, 2011)? Furthermore, while this would be a good start, how do we tackle these hydrologic dragons?

4.1 | First, open and shared collective perceptual models

Hydrology as a science is strongly dependent on experience. This experience is difficult to share and pass on fully. One strategy to improve this sharing, we believe, lies in the development of collective and open perceptual models that evolve if new insight becomes available. Such perceptual models would have to be developed with a granularity that is sufficient to derive testable hypotheses (Beven & Chappell, 2021), but not too fine, because this would distract the focus from dominant processes which should be captured. Simple perceptual models that capture our expectation of how a system will behave already exist—for example, within the comparative hydrology framework by Falkenmark and Chapman (1989). More complex and spatially distributed versions transferred to larger scales or across larger domains are required to facilitate where our understanding converges or diverges when applied outside of experimental catchments. Visualizing such changing knowledge landscape is an interesting challenge for computer science (Gil et al., 2019). Even weak constraints on hydrologic dynamics derived from such perceptual models might help to constrain acceptable model behavior as has been shown at the catchment scale and beyond (Hartmann et al., 2015; Seibert & McDonnell, 2002; Wagener & Montanari, 2011). We finally should stress that shared, open and evolving perceptual models would make exceptionally good tools for hydrology teaching across all levels of students.

4.2 | Second, improved knowledge accumulation

We argue that knowledge accumulation is poor in the field of hydrology and needs to become a much stronger focus. While the sharing of insights through collectively developed and shared perceptual models would be a first step, much of our knowledge is captured in journal articles, and not easily found or extracted. Here, semantic data mining algorithms might offer the chance to harvest the existing knowledge in an effective manner. In the future, we need to improve the efficiency of extracting and synthesizing knowledge from published work. To do so would require a metadata tagging of journal papers with Essential Hydrological Descriptors such as geolocation and time period studied. A separate open database—a Web of Hydrology—could become a community virtual laboratory by linking these essential metadata to evolving descriptors of climatic, topographic, or other properties.

Some 350 years after John Cabot had set sail to the West from Bristol harbor, Alexander von Humboldt published, for the time, an incredibly comprehensive portrait of nature in the first volume of his work *Cosmos: A Sketch of a Physical Description of the Universe* in 1845. Humboldt's *Cosmos* was largely the results of multiple expeditions in the Americas to explore some of the blank areas shown in the Salviati Planisphere (Figure 1b). His aim was "… to grasp Nature's essence under the cover of outer appearances" by studying the "perceptible world," an objective akin to Dooge's search for hydrologic laws (Dooge, 1986). Humboldt took an incredible 25 years to write his five-volume *Cosmos* while corresponding by letter with scientists across the globe on topics including botany, geology, geography, and volcanology. We have since moved on to communicate via (increasingly open) journal articles, and via exchanges at conferences and online meetings, but knowledge accumulation remains cumbersome and time-consuming. We need to urgently rethink how we share, debate, and ultimately accumulate hydrologic knowledge given the opportunities provided by web-based tools and machine learning—this would help us to tackle some of our dragons. It might even help us to realize that some systems are not hydrological monsters with unexpected behavior (a term coined by Kuczera et al., 2010), if we might find out that their behavior is not as unexpected as we think it is once we gain a better overview.

AUTHOR CONTRIBUTIONS

Thorsten Wagener: Conceptualization; writing - original draft; writing-review & editing. **Tom Gleeson:** Conceptualization; writing-review & editing. **Gemma Coxon:** Writing-review & editing. **Andreas Hartmann:** Writing-review & editing. **Nicholas Howden:** Writing-review & editing. **Francesca Pianosi:** Writing-review & editing. **Mostaquimur Rahman:** Writing-review & editing. **Rafael Rosolem:** Writing-review & editing. **Lina Stein:** Writing-review & editing. **Ross Woods:** Writing-review & editing.

DATA AVAILABILITY STATEMENT

The statistics in Figure 4 are included in the caption of the figure in the manuscript. Further data sharing is not applicable to this article as no further new data were created or analyzed in this study.

ORCID

Thorsten Wagener https://orcid.org/0000-0003-3881-5849
Tom Gleeson https://orcid.org/0000-0001-9493-7707
Gemma Coxon https://orcid.org/0000-0002-8837-460X
Andreas Hartmann https://orcid.org/0000-0003-0407-742X
Nicholas Howden https://orcid.org/0000-0002-0422-0524
Francesca Pianosi https://orcid.org/0000-0002-1516-2163
Mostaquimur Rahman https://orcid.org/0000-0003-0950-9009
Rafael Rosolem https://orcid.org/0000-0002-4914-692X
Lina Stein https://orcid.org/0000-0002-9539-9549
Ross Woods https://orcid.org/0000-0002-5732-5979

RELATED WIREs ARTICLES

Putting the cat in the box: why our models should consider subsurface heterogeneity at all scales
On hypothesis testing in hydrology: Why falsification of models is still a really good idea
Perceptual perplexity and parameter parsimony

Further Reading

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., ... Cudennec, C. (2013). A decade of Predictions in Ungauged Basins (PUB)—A review. *Hydrological Sciences Journal*, *58*(6), 1198–1255.
Or, D. (2020). The tyranny of small scales—On representing soil processes in global land surface models. *Water Resources Research*, *56*, 1–9. https://doi.org/10.1029/2019WR024846

REFERENCES

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). A ranking of hydrological signatures based on their predictability in space. *Water Resources Research*, *54*(11), 8792–8812. https://doi.org/10.1029/2018WR022606
Agostinho, D., D'Ignazio, C., Ring, A., Thylstrup, N. B., & Veel, K. (2019). Uncertain archives: Approaching the unknowns, errors, and vulnerabilities of Big Data through cultural theories of the archive. *Surveillance & Society*, *17*(3/4), 422–441.
Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W. H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S., & Over, T. (2015). Accelerating advances in continental domain hydrologic modelling. *Water Resources Research*, *51*, 10078–10091. https://doi.org/10.1002/2015WR017498
Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J. F., & Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*(7729), 59–64.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & de Roo, A. (2019). MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, *21*, 589–615. https://doi.org/10.5194/hess-21-589-2017

Beck, H. E., Wood, E. F., McVicar, T. R., Zambrano-Bigiarini, M., Alvarez-Garreton, C., Baez-Villanueva, O. M., Sheffield, J., & Karger, D. N. (2019). Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments. *Journal of Climate*, *33*, 1299–1315. https://doi.org/10.1175/JCLI-D-19-0332.1

Beven, K. J. (2000). Uniqueness of place and process representations in hydrological modelling. *Hydrology and Earth Systems Sciences*, *4*, 203–213. https://doi.org/10.5194/hess-4-203-2000

Beven, K. J. (2001). *Rainfall-runoff modelling: The primer*. Wiley.

Beven, K. J. (2018). On hypothesis testing in hydrology: Why falsification of models is still a really good idea. *WIREs Water*, *5*, 1–8. https://doi.org/10.1002/wat2.1278

Beven, K. J., Asadullah, A., Bates, P., Blyth, E., Chappell, N., Child, S., Cloke, H., Dadson, S., Everard, N., Fowler, H. J., Freer, J., Hannah, D. M., Heppell, K., Holden, J., Lamb, R., Lewis, H., Morgan, G., Parry, L., & Wagener, T. (2020). Developing observational methods to drive future hydrological science: Can we make a start as a community? *Hydrological Processes*, *34*(3), 868–873. https://doi.org/10.1002/hyp.13622

Beven, K. J., & Chappell, N. (2021). Perceptual perplexity and parameter parsimony in Hydrology. *WIREs Water*, *8*, 1–17. https://doi.org/10.1002/wat2.1530

Beven, K. J., Cloke, H., Pappenberger, F., Lamb, R., & Hunter, N. (2015). Hyperresolution information and hyperresolution ignorance in modelling the hydrology of the land surface. *Science China Earth Sciences*, *58*(1), 22–35.

Beven, K. J., & Cloke, H. L. (2012). Comment on "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water" by Wood et al (2011). *Water Resources Research*, *48*, W01801. https://doi.org/10.1029/2011WR010982

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., ... Wood, E. F. (2015). Hyper-resolution global hydrological modelling—What is next? "Everywhere and locally relevant". *Hydrological Processes*, *29*(2), 310–320.

Bird, A. (2007). What is scientific progress? *Noûs*, *41*, 64–89.

Bird, A. (2008). Scientific progress as accumulation of knowledge: A reply to Rowbottom. *Studies In History and Philosophy of Science Part A*, *39*(2), 279–281.

Blair, G. S., Beven, K., Lamb, R., Bassett, R., Cauwenberghs, K., Hankin, B., Dean, G., Hunter, N., Edwards, L., Nundloll, V., Samreen, F., Simm, W., & Towe, R. (2019). Models of everywhere revisited: A technological perspective. *Environmental Modelling and Software*, *122*, 104521.

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H. G., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., ... Zhang, Y. (2019). Twenty-three unsolved problems in hydrology (UPH)—a community perspective. *Hydrological Sciences Journal*, *64*(10), 1141–1158.

Blöschl, G., Sivapalan, M., Savenije, H., Wagener, T., & Viglione, A. (Eds.) (2013). Runoff prediction in ungauged basins: Synthesis across processes, places and scales. Cambridge University Press.

Boers, N., Goswami, B., Rheinwalt, A., Bookhagen, B., Hoskins, B., & Kurths, J. (2019). Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature*, *566*(7744), 373–377.

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573–1576.

Boorman, D. B., Hollis, J. M., & Lilly, A. (1995). *Hydrology of soil types: A hydrologically-based classification of the soils in the United Kingdom*. Institute of Hydrology Report No. 126.

Bosmans, J. H. C., van Beek, L. P. H., Sutanudjaja, E. H., & Bierkens, M. F. P. (2017). Hydrological impacts of global land cover change and human water use. *Hydrology and Earth System Sciences*, *21*, 5603–5626.

Brassington, F. C., & Younger, P. L. (2010). A proposed framework for hydrogeological conceptual modelling. *Water and Environment Journal*, *24*(4), 261–273. https://doi.org/10.1111/j.1747-6593.2009.00173.x

Buttle, J. (2006). Mapping first-order controls on streamflow from drainage basins: the T-3 template. *Hydrological Processes*, *20*, 3415–3422. https://doi.org/10.1002/Hyp.6519

Cherkauer, K. A., Bowling, L. C., & Lettenmaier, D. P. (2003). Variable infiltration capacity cold land process model updates. *Global and Planetary Change*, *38*, 151–159.

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijenhoet, R., Bennet, K. E., Pauwels, V. R. N., Cai, X., Wood, A. W., & Peters-Lidard, C. D. E. (2017). The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. *Hydrology and Earth System Sciences*, *21*, 3427–3440. https://doi.org/10.5194/hess-2016-693

De Graaf, I. E. M., Gleeson, T., van Beek, L. P. H., Sutanudjaja, E. H., & Bierkens, M. F. P. (2019). Environmental flow limits to global groundwater pumping. *Nature*, *574*, 90–94. https://doi.org/10.1038/s41586-019-1594-4

Döll, P., Trautmann, T., Gerten, D., Schmied, H. M., Ostberg, S., Saaed, F., & Schleussner, C. F. (2018). Risks for the global freshwater system at 1.5 C and 2 C global warming. *Environmental Research Letters*, *13*(4), 044038.

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

Enemark, T., Peeters, L. J. M., Mallants, D., & Batelaan, O. (2019). Hydrogeological model building and testing: A review. *Journal of Hydrology*, *569*, 310–329. https://doi.org/10.1016/j.jhydrol.2018.12.007

Evaristo, J., & McDonnell, J. J. (2017). A role for meta-analysis in hydrology. *Hydrological Processes*, *31*(20), 3588–3591. https://doi.org/10.1002/hyp.11253

Falkenmark, M., & Chapman, T. G. (1989). *Comparative hydrology: An ecological approach to land and water resources*. UNESCO.

Famiglietti, J. S., Cazenave, A., Eicker, A., Reager, J. T., Rodell, M., & Velicogna, I. (2015). Satellites provide the big picture. *Science (New York, NY)*, *349*(6249), 684–685.

Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., Brooks, P. D., Dietrich, W. E., Flores, A., Grant, G., Kirchner, J. W., Mackay, D. S., McDonnell, J. J., Milly, P. C. D., Sullivan, P. L., Tague, C., Ajami, H., Chaney, N., Hartmann, A., … Yamazaki, D. (2019). Hillslope hydrology in global change research and earth system modeling. *Water Resources Research*, *55*, 1–36. https://doi.org/10.1029/2018WR023903

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M. H., Ghezzehei, T. A., Hengl, T., Kollet, S., Agam, N., & Avissar, R. (2020). Soil structure is an important omission in Earth System Models. *Nature Communications*, *11*, 522. https://doi.org/10.1038/s41467-020-14411-z

Ghiggi, G., Humphrey, V., Seneviratne, S. I., & Gudmundsson, L. (2019). A global observations-driven runoff reconstruction for 1901–2014. *Earth System Sciences Data*, *11*, 1655–1674. https://doi.org/10.5194/essd-11-1655-2019

Gil, Y., Pierce, S. A., Babaie, H., Banerjee, A., Borne, K., Bust, G., Cheatham, M., Ebert-Uphoff, I., Gomes, C., Hill, M., Horel, J., Hsu, L., Kinter, J., Knoblock, C., Krum, D., Kumar, V., Lermusiaux, P., Liu, Y., North, C., … Zhang, J. (2018). Intelligent systems for geosciences: An essential research agenda. *Communications of the ACM*, *62*(1), 76–84. https://doi.org/10.1145/3192335

Gleeson, T., Smith, L., Moosdorf, N., Hartmann, J., Dürr, H. H., Manning, A. H., van Beek, L. P. H., & Jellinek, A. M. (2011). Mapping permeability over the surface of the Earth. *Geophysical Research Letters*, *38*, L02401. https://doi.org/10.1029/2010GL045565

Grabs, W., De Couet, T., & Pauler, J. (1996). *Freshwater fluxes from the continents into the world oceans based on data of the global runoff data base*. Tech. Rep. 10. Global Runoff Data Centre.

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), 1–16.

Gupta, H. V., & Nearing, G. S. (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*(6), 5351–5359.

Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: Towards a diagnostic approach to model evaluation. *Hydrological Processes*, *22*, 3802–3813. https://doi.org/10.1002/hyp.6989

Gutmann, E. D., & Small, E. E. (2007). A comparison of land surface model soil hydraulic properties estimated by inverse modeling and pedotransfer functions. *Water Resources Research*, *43*, W05418. https://doi.org/10.1029/2006WR005135

Harari, Y. N. (2015). *Sapiens: A Brief History of Humankind* (464 pp. Harper.

Hartmann, A., Gleeson, T., Rosolem, R., Pianosi, F., Wada, Y., & Wagener, T. (2015). A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean. *Geoscientific Model Development*, *8*, 1729–1746. https://doi.org/10.5194/gmd-8-1729-2015

Hartmann, A., Gleeson, T., Wada, Y., & Wagener, T. (2017). Enhanced groundwater recharge rates and altered recharge sensitivity to climate variability through subsurface heterogeneity. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(11), 2842–2847. https://doi.org/10.1073/pnas.1614941114

Hrachowitz, M., Savenije, H. H. S., Bloschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., Fenicia, F., Freer, J., Gelfan, A., Gupta, H., Hughes, D., Hut, R., Montanari, A., Pande, S., Tetzlaff, D., … Cudennec, C. (2013). A decade of Predictions in Ungauged Basins (PUB)—A review. *Hydrological Science Journal*, *58*, 1–58. https://doi.org/10.1080/02626667.2013.803183

Hradec, J., Ostlaender, N., Macmillan, C., Acs, S., Listorti, G., Tomas, R., & Arnes Novau, X. (2019). *Semantic text analysis tool: SeTA*. EUR 29708 EN. Publications Office of the European Union. https://doi.org/10.2760/577814 JRC116152.

Huscroft, J., Gleeson, T., Hartmann, J., & Börker, J. (2018). Compiling and mapping global permeability of the unconsolidated and consolidated earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). *Geophysical Research Letters*, *45*, 1897–1904. https://doi.org/10.1002/2017GL075860

Jencso, K. G., & McGlynn, B. L. (2012). Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation. *Water Resources Research*, *47*, W11527. https://doi.org/10.1029/2011WR010666

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., … Williams, C. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat-derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, *116*, G00J07. https://doi.org/10.1029/2010JG001566

Kauffeldt, A., Halldin, S., Rodhe, A., Xu, C.-Y., & Westerberg, I. K. (2013). Disinformative data in large-scale hydrological modelling. *Hydrology and Earth System Sciences*, *17*, 2845–2857. https://doi.org/10.5194/hess-17-2845-2013

Kelleher, C., McGlynn, B., & Wagener, T. (2013). Characterizing and reducing equifinality by constraining a distributed catchment model with regional signatures, local observations, and process understanding. *Hydrology and Earth System Sciences*, *21*(7), 3325–3352.

Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., Rouf, N., & Mohi Ud Din, M. (2020). Machine learning based approaches for detecting COVID-19 using clinical text data. *Int j inf tecnol*, *12*, 731–739. https://doi.org/10.1007/s41870-020-00495-9

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, *42*, W03S04. https://doi.org/10.1029/2005WR004362

Knoben, W. J. M., Woods, R. A., & Freer, J. E. (2018). A quantitative hydrological climate classification evaluated with independent streamflow data. *Water Resources Research*, *54*(7), 5088–5109.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, *55*(12), 11344–11354.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences.*, *23*, 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Kuczera, G., Renard, B., Thyer, M., & Kavetski, D. (2010). There are no hydrological monsters, just models and observations with large uncertainties! *Hydrological Sciences Journal*, *55*(6), 980–991.

Kuentz, A., Arheimer, B., Hundecha, Y., & Wagener, T. (2017). Understanding hydrologic variability across Europe through catchment classification. *Hydrology and Earth System Sciences*, *21*, 2863–2879. https://doi.org/10.5194/hess-21-2863-2017

Kumar, A., Dabas, V., & Hooda, P. (2018). Text classification algorithms for mining unstructured data: A SWOT analysis. *International Journal of Information Technology*, *12*, 1159–1169. https://doi.org/10.1007/s41870-017-0072-1

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., & Wisser, D. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*(9), 494–502.

Leonard, L., & Duffy, C. J. (2013). Essential terrestrial variable data workflows for distributed water resources modeling. *Environmental Modelling & Software*, *50*, 85–96.

Leonard, L., & Duffy, C. J. (2014). Automating data-model workflows at a level 12 HUC scale: Watershed modeling in a distributed computing environment. *Environmental Modelling & Software*, *61*, 174–190. https://doi.org/10.1016/j.envsoft.2014.07.015

Leonard, L., & Duffy, C. J. (2016). Visualization workflows for level-12 HUC scales: Towards an expert system for watershed analysis in a distributed computing environment. *Environmental Modelling & Software*, *78*, 163–178. https://doi.org/10.1016/j.envsoft.2016.01.001

Lindersson, S., Brandimarte, L., Mård, J., & Di Baldassarre, G. (2020). A review of freely accessible global datasets for the study of floods, droughts and their interactions with human societies. *WiresWater*, 1–20. https://doi.org/10.1002/wat2.1424

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., & Thieme, M. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific Data*, *6*, 283. https://doi.org/10.1038/s41597-019-0300-6

Liu, Y., Wagener, T., Beck, H. E., & Hartmann, A. (2020). What is the hydrologically effective size of a catchment? *Environmental Research Letters*, 1–10. https://doi.org/10.31223/osf.io/k2z5h

Loritz, R., Hassler, S. K., Jackisch, C., Allroggen, N., van Schaik, L., Wienhöfer, J., & Zehe, E. (2017). Picturing and modeling catchments by representative hillslopes. *Hydrology and Earth System Sciences*, *21*, 1225–1249. https://doi.org/10.5194/hess-21-1225-2017

Manabe, S., & Holloway, J. L., Jr. (1975). The seasonal variation of the hydrologic cycle as simulated by a global model of the atmosphere. *Journal of Geophysical Research*, *80*(12), 1617–1694.

Maxwell, R. M., & Condon, L. E. (2016). Connections between groundwater flow and transpiration partitioning. *Science*, *353*(6297), 377–380. https://doi.org/10.1126/science.aaf7891

McDonnell, J. J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M. L., Selker, J., & Weiler, M. (2007). Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research*, *43*, W07301. https://doi.org/10.1029/2006WR005467

McDonnell, J. J., & Woods, R. (2004). On the need for catchment classification. *Journal of Hydrology*, *299*, 2–3.

Merz, B., Vorogushyn, S., Lall, U., Viglione, A., & Blöschl, G. (2015). Charting unknown waters—On the role of surprise in flood risk assessment and management. *Water Resources Research*, *51*(8), 6399–6416. https://doi.org/10.1002/2015WR017464

Merz, R., Tarasova, L., & Basso, S. (2020). Parameter's controls of distributed catchment models—How much information is in conventional catchment descriptors? *Water Resources Research*, *56*, e2019WR026008. https://doi.org/10.1029/2019WR026008

Miralles, D. G., De Jeu, R. A., Gash, J. H., Holmes, T. R., & Dolman, A. J. (2011). Magnitude and variability of land evaporation and its components at the global scale. *Hydrology and Earth System Sciences*, *15*, 967–981.

Mizrahi, M. (2013). What is scientific progress? Lessons from scientific practice. *Journal for General Philosophy of Science*, *44*(2), 375–390.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and MetaAnalyses: The PRISMA Statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Mulligan, M., van Soesbergen, A., & Saenz, L. (2020). GOODD, a global dataset of more than 38,000 georeferenced dams. *Scientific Data*, *7* (31), 1–8.

Nardi, F., Annis, A., Di Baldassarre, G., Vivoni, E. R., & Grimaldi, S. (2019). GFPLAIN250m, a global high-resolution dataset of Earth's floodplains. *Scientific Data*, *6*, 180309. https://doi.org/10.1038/sdata.2018.309

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research*, in Press, *57*(e2020WR028091), 1–15. https://doi.org/10.1029/2020WR028091.

Olarinoye, T., Gleeson, T., Marx, V., Seeger, S., Adinehvand, R., Allocca, V., Andreo, B., Apaéstegui, J., Apolit, C., Arfib, B., Auler, A., Bailly-Comte, V., Barberá, J. A., Batiot-Guilhe, C., Bechtel, T., Binet, S., Bittner, D., Blatnik, M., Bolger, T., ... Hartmann, A. (2020). Global karst springs hydrograph dataset for research and management of the world's fastest-flowing groundwater. *Scientific Data*, *7*, 59. https://doi.org/10.1038/s41597-019-0346-5

Pfister, L., & Kirchner, J. W. (2017). Debates—Hypothesis testing in hydrology: Theory and practice. *Water Resources Research*, *53*, 1792–1798.

Popper, K. (1959). *The logic of scientific discovery*. 2002 Routledge Edition. Routledge.

Price, K. (2011). Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review. *Progress in Physical Geography*, *35*(4), 465–492.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*, 195–204.

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., & Lo, M.-H. (2018). Emerging trends in global freshwater availability. *Nature*, *557*(7707), 651–659.

Rogger, M., Pirkl, H., Viglione, A., Komma, J., Kohl, B., Kirnbauer, R., Merz, R., & Blöschl, G. (2012). Step changes in the flood frequency curve: Process controls. *Water Resources Research*, *48*, W05544. https://doi.org/10.1029/2011WR011187

Rosero, E., Yang, Z. L., Wagener, T., Gulden, L. E., Yatheendradas, S., & Niu, G. Y. (2010). Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season. *Journal of Geophysical Research: Atmospheres*, *115*(D3).

Sarrazin, F., Hartmann, A., Pianosi, F., Rosolem, R., & Wagener, T. (2018). V2Karst v1.1: A parsimonious large-scale integrated vegetation-recharge model to simulate the impact of climate and land cover change in karst regions. *Geoscientific Model Development*, *11*(12), 4933–4964. https://doi.org/10.5194/gmd-11-4933-2018

Savenije, H. H. G. (2010). HESS Opinions "Topography driven conceptual modelling (FLEX-Topo)". *Hydrology and Earth System Sciences*, *14*, 2681–2692. https://doi.org/10.5194/hess-14-2681-2010

Seibert, J., & McDonnell, J. J. (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multi-criteria model calibration. *Water Resources Research*, *38*(11), 23-1–23-14.

Seibert, J., & McGlynn, B. L. (2007). A new triangular multiple flow-direction algorithm for computing upslope areas from gridded digital elevation models. *Water Resources Research*, *43*, W04501. https://doi.org/10.1029/2006WR005128

Seibert, J., Vis, M. J. P., Lewis, E., & van Meerveld, H. J. (2018). Upper and lower benchmarks in hydrological modelling. *Hydrological Processes*, 1120–1125. https://doi.org/10.1002/hyp.11476

Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F. J., Ganguly, S., Hsu, K. L., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X., & Tsai, W. P. (2018). HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community. *Hydrology and Earth System Sciences*, *22*, 5639–5656. https://doi.org/10.5194/hess-22-5639-2018

Sivapalan, M. (2009). The secret to "doing better hydrological science": Change the question! *Hydrological Processes*, *23*, 1391–1396. https://doi.org/10.1002/hyp.7242

Sivapalan, M., Blöschl, G., Zhang, L., & Vertessy, R. (2003). Downward approach to hydrological prediction. *Hydrological Processes*, *17*(11), 2101–2111.

Stolbova, V., Surovyatkina, E., Bookhagen, B., & Kurths, J. (2016). Tipping elements of the Indian monsoon: Prediction of onset and withdrawal. *Geophysical Research Letters*, *43*, 3982–3990. https://doi.org/10.1002/2016GL068392

Straatsma, M. W., Vermeulen, P., Kuijper, M., Bonte, M., Niele, F., & Bierkens, M. F. P. (2016). Rapid screening of operational freshwater availability using global models. *Water Resources Management*, *30*(9), 3013–3026.

Syvitski, J. P. M., Vörösmarty, C. J., Kettner, A. J., & Green, P. (2005). Impact of humans on the flux of terrestrial sediment to the global coastal ocean. *Science*, *308*(5720), 376–380.

Tetzlaff, D., McDonnell, J. J., Uhlenbrook, S., McGuire, K. J., Bogaart, P. W., Naef, F., Baird, A. J., Dunn, S. M., & Soulsby, C. (2008). Conceptualizing catchment processes: simply too complex? *Hydrological Processes*, *22*, 1727–1730. https://doi.org/10.1002/hyp.7069

Troy, T. J., Wood, E. F., & Sheffield, J. (2008). An efficient calibration method for continental-scale land surface modeling. *Water Resources Research*, *44*(9), 1–13.

Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K. A., Ceder, G., & Jain, A. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature*, *571*, 95–98.

Turuncoglu, U. U., Dalfes, N., Murphy, S., & DeLuca, C. (2013). Toward self-describing and workflow integrated Earth system models: A coupled atmosphere-ocean modeling system application. *Environmental Modelling & Software*, *39*, 247–262.

Van Werkhoven, K., Wagener, T., Tang, Y., & Reed, P. (2009). Complexity reduction in multiobjective watershed model calibration. *Advances in Water Resources*, *32*(8), 1154–1169.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., & Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*, *467*(7315), 555–561.

Wada, Y., L. P. H. v. B., C. M. v. K., Reckman, J. W. T. M., Vasak, S., & Bierkens, M. F. P. (2010). Global depletion of groundwater resources. *Geophysical Research Letters*, *37*(20), 1–5.

Wagener, T., Weiler, M., McGlynn, B., Gooseff, M., Meixner, T., Marshall, L., McGuire, K., & McHale, M. (2007). Taking the pulse of hydrology education. *Hydrological Processes*, *21*, 1789–1792.

Wagener, T., Dadson, S., Hannah, D. M., Coxon, G., Beven, K., Bloomfield, J. P., Buytaert, W., Cloke, H., Bates, P., Holden, J., Parry, L., Lamb, R., Fry, M., & Old, G. (2021). Knowledge gaps in our perceptual model of Great Britain's hydrology. *Hydrological Processes*, *35*, 1–13. https://doi.org/10.1002/hyp.14288

Wagener, T., & Montanari, A. (2011). Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. *Water Resources Research*, *47*, W06301. https://doi.org/10.1029/2010WR009469

Winsemius, H. C., Aerts, J. C. J. H., van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., Kwadijk, J. C. J., Ligtvoet, W., Lucas, P. L., van Vuuren, D. P., & Ward, P. J. (2016). Global drivers of future river flood risk. *Nature Climate Change*, *6*(4), 381–385.

Winter, T. C. (2001). The concept of hydrologic landscapes. *Journal of the American Water Resources Association*, *37*, 335–349.

Wood, E. F., Boll, J., Bogaart, P., & Troch, P. (2005). The need for a virtual hydrologic laboratory for PUB, Ch 16. In S. Franks, M. Sivapalan, K. Takeuchi, & Y. Tachikawa (Eds.), IAHS Pub 301 *Predictions in ungauged basins: International perspectives on the state of the art and pathways forward* (pp. 189–203). Wallingford.

Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., … Whitehead, P. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, *47*(5), 1–10. https://doi.org/10.1029/2010WR010090

Yan, D., Wang, K., Qin, T., Weng, B., Wang, H., Bi, W., Li, X., Li, M., Lv, Z., Liu, F., He, S., Ma, J., Shen, Z., Wang, J., Bai, H., Man, Z., Sun, C., Liu, M., Shi, X., … Abiyu, A. (2019). A data set of global river networks and corresponding water resources zones divisions. *Sci Data*, *6*, 219. https://doi.org/10.1038/s41597-019-0243-y

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., & Dickinson, R. (2013). The role of satellite remote sensing in climate change studies. *Nature Climate Change*, *3*, 875–883. https://doi.org/10.1038/NCLIMATE1908

Yang, Y., Pan, M., Beck, H. E., Fisher, C. K., Beighley, R. E., Kao, S. C., Hong, Y., & Wood, E. F. (2019). In quest of calibration density and consistency in hydrologic modelling: Distributed parameter calibration against streamflow characteristics. *Water Resources Research*, *55*, 7784–7803. https://doi.org/10.1029/2018WR024178

Yatheendradas, S., Wagener, T., Gupta, H., Unkrich, C., Goodrich, D., Schaffner, M., & Stewart, A. (2008). Understanding uncertainty in distributed flash flood forecasting for semiarid regions. *Water Resources Research*, *44*, W05S19. https://doi.org/10.1029/2007WR005940

Young, P. (2003). Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale. *Hydrological Processes*, 2195–2217. https://doi.org/10.1002/hyp.1328