Wiley Interdisciplinary Reviews: Water

WILEY

WIREs WATER

ADVANCED REVIEW OPEN ACCESS

Uncertainties as a Guide for Global Water Model Advancement

Robert Reinecke[1,2] | Lina Stein[2] | Sebastian Gnann[2,3] | Jafet C. M. Andersson[4] | Berit Arheimer[4] | Marc Bierkens[5,6] | Sara Bonetti[7] | Andreas Güntner[2,8] | Stefan Kollet[9] | Sulagna Mishra[10] | Nils Moosdorf[11,12] | Sara Nazari[11] | Yadu Pokhrel[13] | Christel Prudhomme[14,15] | Jacob Schewe[16] | Chaopeng Shen[17] | Thorsten Wagener[2]

[1]Institute of Geography, Johannes Gutenberg-University Mainz, Mainz, Germany | [2]Institute of Environmental Science and Geography, University of Potsdam, Germany | [3]Chair of Hydrology, Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany | [4]Hydrology Research Unit, Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden | [5]Department of Physical Geography, Utrecht University, the Netherlands | [6]Unit Subsurface and Groundwater Systems, Deltares, Utrecht, the Netherlands | [7]Laboratory of Catchment Hydrology and Geomorphology, École Polytechnique Fédérale de Lausanne, Sion, Switzerland | [8]GFZ Helmholtz Centre for Geosciences, Potsdam, Germany | [9]Centre for High-Performance Scientific Computing in Terrestrial System, Jülich, Germany | [10]World Meteorological Organization WMO, Geneva, Switzerland | [11]Department for Biogeochemistry and Geology, Leibniz Centre for Tropical Marine Research (ZMT), Bremen, Germany | [12]Kiel University, Kiel, Germany | [13]Department of Civil and Environmental Engineering, Michigan State University, East Lansing, Michigan, USA | [14]Forecast Department, European Centre for Medium-Range Weather Forecasts, Reading, UK | [15]Geography Department, University of Loughborough, Loughborough, UK | [16]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany | [17]Civil and Environmental Engineering, Pennsylvania State University, University Park, Pennsylvania, USA

Correspondence: Robert Reinecke (reinecke@uni-mainz.de)

Received: 4 November 2024 | **Revised:** 9 April 2025 | **Accepted:** 19 April 2025

Associate Editor: Qiuhong Tang | **Editor-in-Chief:** Wendy Jepson

Funding: This work was supported by NSF (1752729); ERC (101019185); Alexander von Humboldt-Stiftung.

Keywords: GHMs | global models | uncertainty | water cycle

ABSTRACT

Global water models allow us to explore the terrestrial water cycle in earth-sized digital laboratories to support science and guide policy. However, these models are still subject to considerable but also reducible uncertainties that can be attributed to mainly three sources: (1) imbalances in data quality and availability across geographical regions and between hydrologic variables, (2) poorly quantified human influence on the water cycle, and (3) difficulties in tailoring process representations to regionally diverse hydrologic systems. New, more accurate, and larger datasets, as well as better accumulated and even enhanced process knowledge, will help to reduce these uncertainties and thus improve model consistency with our perceptions and accuracy given existing observations. This review examines the sources of uncertainty crucial for global water models and proposes actions to mitigate them, thereby providing a roadmap for model advancement. Following this path will yield more consistent and accurate models that are urgently needed to tackle key scientific and societal challenges.

1 | Introduction

In 1972, the Blue Marble picture showed us, for the first time, a color image of Earth from space, laying bare its vulnerability and interconnectedness through the water cycle (Eagleson 1991). It suggested that we require an understanding of the past, present, and future of Earth's freshwater resources to safeguard the blue planet for future generations. Global

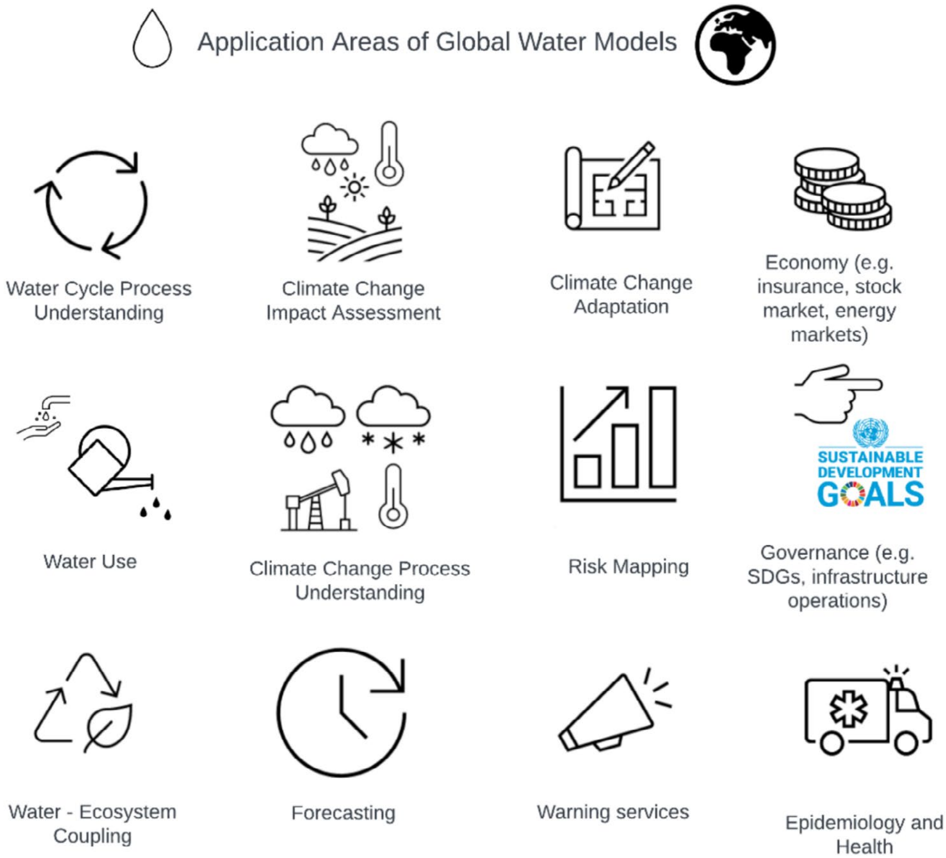

FIGURE 1 | Global water models inform a wide range of applications, but their potential has not been extensively explored in all possible areas. References to a subjective selection of examples and further explanations of potential applications can be found in Table S1.

water models (GWMs) are a central tool to foster this understanding by simulating the global terrestrial water cycle with the help of computer programs. Starting with the first coarse-resolution models using simple process representation in the 1990s, these models have evolved tremendously due to remote sensing (Chahine 1992; Wulder et al. 2022) and an increased process understanding from regional hydrologic research (Shuttleworth 1994). See Figure S1 for a history of GWM evolution. They can now be run at high temporal (hourly–daily) and spatial (up to 1 km) resolutions and include an increasing range of hydrologic processes and anthropogenic influences (Sellers et al. 1997; Pokhrel et al. 2016; Arheimer et al. 2020; Müller Schmied et al. 2021; Hoch et al. 2023). Until recently, these models have been solely based on (sometimes very simplified) representations of hydrological processes, but machine learning and hybrid approaches are starting to emerge (Feng et al. 2023). Here, we discuss multiple classes of models that we collectively categorize as GWMs: global hydrological models, land surface models, Earth system models (ESMs), and dynamic vegetation models (Bierkens 2015). While all model classes were initially built for different purposes, they all simulate the global terrestrial water cycle (though at various levels of detail) and are used to answer questions related to global hydrology (e.g., Haddeland et al. 2011; Gudmundsson, Wagener, et al. 2012; Schewe et al. 2019; Pokhrel et al. 2021; Seo et al. 2025). We chose to jointly discuss them as GWMs as all model classes benefit from a more accurate representation of the terrestrial water cycle, and they all suffer from uncertainties specific to modeling these processes on a global scale.

Because water connects all spheres of the Earth, GWMs have a wide range of applications and offer opportunities for future use in many fields (Figure 1). In contrast to models of specific places or regions (e.g., catchments or river basins), the capacity of GWMs to generate continuous and consistent global hydrological time series for variables such as streamflow, soil moisture, evapotranspiration, or snow water equivalent makes them a valuable resource. Their global coverage allows usage in regions with limited or no observations; they help to understand spatiotemporal patterns of hydrological extremes (Ward et al. 2014; Emerton et al. 2017; Arheimer et al. 2020) in support of global early warning systems and risk maps (Emerton et al. 2016; He et al. 2020), and they help to assess future risks such as water scarcity and explore possible adaptation measures (Veldkamp et al. 2016; van Vliet et al. 2021). Other research fields have used GWMs to assess issues related to the Water-Energy-Food nexus (Lodge et al. 2023) or the impacts of climate change on freshwater ecosystems (Döll and Zhang 2010; Bartosova et al. 2021). Furthermore, GWMs have improved the representation of the Earth system in climate and weather models because one of the primary outputs of water models, streamflow, naturally integrates various terrestrial hydrological processes and can thus be used to evaluate mass balances of other models (Zsoter et al. 2019; Boussetta et al. 2021).

While GWMs provide global coverage of multiple water cycle components, current uncertainties, and their poor quantification may limit their value (e.g., for global change impact analysis) as the reliability of model estimates remains unclear (Wagener et al. 2022). For example, the sixth IPCC assessment report concluded that our knowledge of climate change impacts on groundwater is still poor, partly because groundwater remains inadequately represented in many models (IPCC 2023). Even if implemented, the uncertainty of groundwater levels remains high (Reinecke et al. 2024). Complex process interrelationships such as atmospheric CO_2 fertilization and its long-term impact on global water availability are poorly known (Milly and Dunne 2016), while current representations of anthropogenic impacts such as irrigation, groundwater abstractions (Arheimer et al. 2020; Puy et al. 2021; McDermid et al. 2023), and river regulation (Arheimer et al. 2017) are still in their infancy. This has direct consequences for other impact assessments, such as assessing the economic impacts of floods (Willner et al. 2018) or vulnerability to food insecurity (Betts et al. 2018).

Much has been said about the topic of uncertainty in the context of local and regional models (Wagener and Montanari 2011; Beven 2016; Nearing et al. 2016), but what are the problems and solutions specific to global water models? Here, we focus on so-called epistemic uncertainty, defined by Walker et al. (2003) as "uncertainty due to the imperfection of our knowledge, which may be reduced by more research and empirical efforts." If we can identify those uncertainties that can be reduced, they can guide advancements of GWMs and our scientific hydrological understanding in general. By advancements, we mean both improved consistency (i.e., more realistic simulation and inclusion of dominant hydrological processes given our perception of real-world processes (Wagener et al. 2022); also referred to as fidelity (Clark, Fan, et al. 2015)) and improved accuracy (i.e., smaller differences between model outputs and observations; Clark, Fan, et al. 2015; Wagener et al. 2022). To a lesser extent, we also mean more flexible models (i.e., a model structure that can be tailored to different conditions) and sustainable and reproducible modeling software (Nyenah et al. 2024).

In this review, we outline that the advancement of GWMs can take different paths, by gathering new, more accurate, and larger datasets, improving process knowledge, and ultimately building more consistent and accurate models. Based on a critical review of the current literature, we identify three key sources of uncertainty specific to modeling the terrestrial water cycle on a global scale. We then outline how these uncertainties affect our understanding of past, present, and future water cycles, and suggest ways to reduce them. Reducing uncertainties that dominate model outputs should yield models that provide improved predictions and align better with our scientific understanding and, as a result, provide more reliable information to policymakers.

Our review is structured as follows: In Section 2, we review how uncertainties originate in the modeling process, namely in the building, forcing, and evaluation of GWMs. In Section 3, we describe the resulting uncertainties that arise in simulating the past, future, and near-future water cycle. In Section 4, we summarize our findings and outline how the identified uncertainties can guide the advancement of GWMs in the future. Finally, we conclude our review in Section 5 by discussing the future of GWMs.

2 | Uncertainties in Building, Forcing, and Evaluating Global Water Models

In this section, we discuss the main sources of epistemic uncertainty in GWMs and explore where they originate during the building, forcing, and evaluation stages of the modeling process. All models are subject to uncertainties, as they are, by definition, an approximation and, thus, an imperfect representation of reality. Knowing when, where, and why models are uncertain is a starting point for refinement and improved scientific understanding (Eyring et al. 2019; Gleeson et al. 2021). Importantly, uncertainties can affect the robustness of and confidence in impact assessments, policies, and decisions derived from model results (Haddeland et al. 2011; Puy et al. 2022). Insufficient and inaccurate quantification and communication of existing uncertainties may lead to overconfident decisions and potentially to a loss of trust in models (Beven 2018).

We focus our discussion on uncertainties that relate to developing and implementing GWMs, though additional uncertainties originate from the natural variability of human and environmental systems. Such aleatory uncertainties represent variability, imprecision, and randomness, or factors that can generally be modeled as probabilities in statistical frameworks (Beven et al. 2018). In addition, uncertainties might arise when testing specific intervention strategies to guide policymaking, such as land use change or water use scenarios, which cannot be assessed against observations. What is important for our discussion here is that many of the uncertainties currently impacting GWMs originate from a lack of knowledge, that is, they are epistemic and can be reduced (in principle) through new or better-utilized observations (e.g., through new algorithms or different models) or through new knowledge (Beven et al. 2018). They exist because we lack system understanding; we cannot measure certain variables in all places, at all times, at the right scale—or sometimes at all—and measurements themselves carry uncertainties (Sivapalan 2018; Condon et al. 2021). In addition, we are often interested in future system states that are possibly very different from the past and thus may lack historical analogs (e.g., due to climate or land use change).

We find that uncertainties can largely be attributed to three sources (Figure 2): (1) imbalances in data quality/availability across geographical regions and between hydrologic variables, (2) poorly quantified human alterations of the water cycle, and (3) difficulties in tailoring models to regionally diverse hydrologic systems. Imbalances in existing datasets create considerable problems in our ability to assess how consistent models and real-world dynamics are. For example, (in situ) data availability in temperate regions such as Europe and North America tends to be high, and we generally have more data in regions with higher population densities (Krabbenhoft et al. 2022; Figure 2a). If observations are available, they frequently suffer from inconsistencies due to differences in data collection between administrative boundaries (Figure 2b), which are

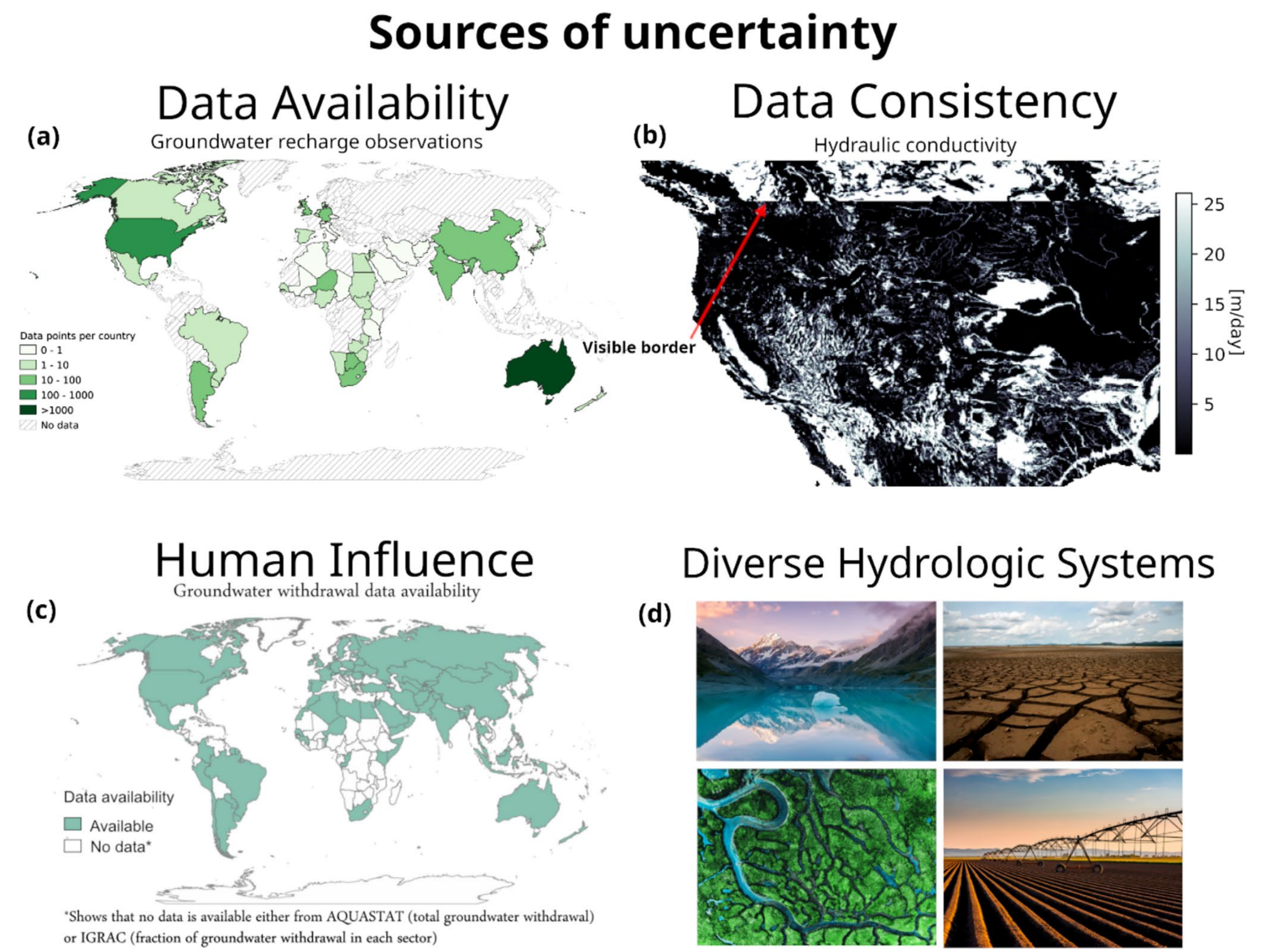

FIGURE 2 | Sources of uncertainty for global water models. Uncertainties mainly originate from three sources: (a) Imbalances in data availability, exemplified by showing the number of available data points in a global dataset of groundwater recharge (Moeck et al. 2020), and (b) consistency across geographical regions illustrated by a zoom-in showing that hydraulic conductivity in a widely used global dataset (Gleeson et al. 2014) changes abruptly at the border between Canada and the USA, (c) poorly quantified human influence on the water cycle demonstrated with available data on groundwater withdrawal on a country level in the AQUASTAT and IGRAC database, and (d) difficulties in tailoring process representations to regionally diverse hydrologic systems.

challenging to eradicate and can significantly impact model results. In addition, observations are often available in places where anthropogenic influences are considerable but unquantified (often because the data is not publicly available; Figure 2c). For example, while global datasets on reservoirs and dams are available (Mulligan et al. 2020), their operation schemes are largely unknown (Hanasaki et al. 2006). The uncertainty due to availability and consistency spans all variables, including hydrological, meteorological, geological, and hydrogeological variables.

Furthermore, it is difficult to tailor GWMs to reflect diverse regional hydrological systems (Figure 2d) due to incomplete process understanding, the coarse spatial model resolution of most current models, and the lack of data at the scale of model resolution or generally for specific regions. (1) Regional process knowledge is either not (easily) available or biased (i.e., our perceptual understanding is limited; Stein et al. 2024) or we lack the flexibility to implement that knowledge in the current generation of models. (2) Similarly, there are some processes that are highly heterogeneous and difficult to model accurately at a coarse spatial resolution of often 0.5°, even though models approach km-scale resolutions. Examples of these processes include snowmelt across altitudes and soil wetting across different soil types. (3) In addition, observations rarely represent the scale of model units (Weber et al. 2023). In situ, observations of individual variables (e.g., soil moisture) are often only representative of areas much smaller than the scale of the modeling unit (e.g., a raster cell; this is also true for regional models, but the scale difference is likely more severe for GWMs). In contrast, observations measured at larger spatial scales (i.e., satellite measurements) often integrate over multiple state variables and/or larger areas than the modeling unit. Streamflow observations are a special case as they integrate various processes within a catchment. However, (especially) for large catchments, the influence of diverse spatially distributed and heterogeneous runoff generation processes cannot be identified easily (if at all) from the signal that arrives at the catchment outlet (van Werkhoven et al. 2008).

2.1 | Building Global Water Models

Any process-based water model, independent of its application scale, is generally built through a modeling process that establishes a perceptual (or conceptual) model of the system under study and ends with a computational model that can be executed (Beven 2018). For regional models, much discussion has been placed on how these stages are implemented (Beven 2012). Current GWMs are generally built along the following steps: (1) Outlining the different hydrological landscapes that should be modeled explicitly, which provides the conceptual foundation for the model structure, model equations, as well as strategies to estimate or set model parameters. Given the size of the domain, one or a few representative perceptual models are typically used (e.g., separating landscapes into mountains and sedimentary basins; Hartmann et al. 2017; Wagener et al. 2021). (2) The perceptual model(s) are then translated into model structures applied at the modeling unit (typically a raster cell or a catchment), and routing functions are added to connect the individual modeling units. (3) Each model unit is then tailored to regional conditions (e.g., hydro-climatic) via the model parameters, mostly through a priori estimates directly derived from globally available soil, geology, vegetation, and other datasets. (4) Models may consider different aspects of human interventions, such as water abstractions or reservoirs. (5) The model is then driven by forcing inputs such as precipitation, temperature, and radiation, either using observational records, reanalysis products, or projections from Global Climate Models (GCMs). Sometimes, GWMs are directly coupled to a GCM together with other additional physical, chemical, and biological processes, then referred to as an ESM (Clark, Fan, et al. 2015), but definitions may vary—we here also include ESMs as part of the umbrella term GWMs when they simulate the terrestrial water cycle. (6) GWMs can be evaluated based on their main outputs, such as streamflow, soil moisture, or evapotranspiration, though whether and how this is done might vary by model and study focus. Calibration of global models to streamflow or other observations is not standard yet (Kupzig et al. 2023) but can be achieved in principle (Döll et al. 2003; Arheimer et al. 2020). All these modeling steps offer options and choices that introduce uncertainty, and they result in a diverse range of models. However, the current model diversity does not necessarily reflect the diversity of hydrologic processes found across the global land surface. It is rather a consequence of different modeling groups selecting a specific underlying model structure to build their GWM with (Addor and Melsen 2019; Melsen 2022), which often reflects the specific environments these modeling groups work in (e.g., Kiraz et al. 2023). For example, TOPMODEL (Beven et al. 2021) was developed to reflect the topographic control on soil moisture found in the UK, while HBV (Seibert and Bergström 2022) was developed to reflect the snow processes found in Sweden. It is the uncritical transfer to other regions that can be problematic. The model diversity, though not necessarily reflecting process uncertainty, can be captured by model ensembles, which often offer more robust predictions and help reveal knowledge gaps regarding the appropriate representation of the terrestrial water cycle (e.g., Reinecke et al. 2021; Gnann et al. 2023). We note, though, that the ensemble is usually an ensemble of opportunity and does not necessarily reflect model structural uncertainty in a coherent manner.

Data availability is a key problem for all steps of the model-building process for GWMs. It starts during perceptual model development, where the availability of observational data limits how detailed our system perception can be. Given that modelers are also limited in their knowledge of global hydrologic diversity, this limitation influences what structural representations we might even consider. While local experts may possess a thorough understanding of a specific hydrologic system, integrating and synthesizing that local knowledge into a comprehensive global database has yet to be achieved. Even though some first steps have been made to systemize perceptual understanding of different hydrological landscapes (Pechlivanidis and Arheimer 2015; Andersson et al. 2017; McMillan et al. 2023; McMillan et al. 2025), the transferability of system understanding, especially across scales, remains difficult (Wagener et al. 2010). Ultimately, the lack of trustworthy (not uncertainty-free) perceptual models limits our capability of tailoring global models to the diversity of hydrologic systems that we find on Earth.

Once a (or several competing) perceptual model(s) has (have) been chosen, it is translated into a model structure and tailored to different hydrologic systems, mainly through global datasets (Table S2). The datasets are used to estimate a priori model parameters, and in many cases, these data may have already influenced the equations that were used to build the model structure in the first place. For example, the Harmonized World Soil Database (HWSD; Nachtergaele et al. 2010) is utilized as a soil map in eight GWMs in the global water sector of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP; Frieler et al. 2017), and is playing a crucial role in estimating soil-related parameters. Without alternative datasets, quantifying the uncertainty that this choice introduces remains a challenge. Some processes, such as groundwater recharge, for which there is a lack of direct measurements of the relevant system properties are parameterized by combining geology, soil, topography, and permafrost datasets, as well as expert knowledge (Döll and Fiedler 2008). However, such a complex combination of different sources of information may lead to the inability to explain model differences (Reinecke et al. 2021).

Our ability to represent the processes we assume to be present is related to finding an adequate model structure, especially for scales far finer than the spatial or temporal resolution of GWMs. Adequate here means that the model can be used for its intended purpose (see also Figure 1). The representation of sub-grid scale processes and their variability is a challenge to GWMs, with long-lasting debates around the issues of model and parameter adequacy as well as data limitation and uncertainty (Beven and Cloke 2012; Clark et al. 2017). Reasons for this ongoing dispute are questions regarding the validity of theories when applied beyond their scale of derivation, the representation of interactions and feedback among processes, strategies to describe the effects of sub-grid scale heterogeneity, and the availability of data to parameterize and test model formulations. A prominent and illustrative example is the structural representation of the land surface, particularly the representation of soil processes (Fatichi et al. 2020; Or 2020; Weber et al. 2023), whose description is generally rooted in theories and limited observations (Vereecken et al. 2016). As mentioned, most models use maps of soil types [e.g., HWSD (Nachtergaele et al. 2010) or SoilGrids (Poggio et al. 2021)] and correlate them with the model parameters of interest via pedotransfer functions (PTFs). However, current parameterizations of soil hydraulic properties based on PTFs rely on geographically limited data, generally derived from small samples taken from agricultural fields, thus not accounting for soil structure effects and spatial heterogeneities (Or 2020; Gleeson et al. 2021). Such effects may significantly alter infiltration–runoff and other exchange processes at larger scales (Fatichi et al. 2020; Bonetti et al. 2021). Recent research showed that it is possible to incorporate soil structure corrections into pedotransfer functions based on remotely sensed vegetation metrics and local soil texture (Bonetti et al. 2021). The uncertainties in soil process representation influence multiple other processes within GWMs. Besides infiltration and runoff, different soil and land use representations can also influence model translation from radiation forcing into evapotranspiration, thus significantly altering the water balance representation of the model (Gnann et al. 2023).

Another critical aspect of process representation and parameter estimation is anthropogenic alterations of the terrestrial water cycle through land and water management. Humans have profoundly impacted freshwater systems by changing land use patterns, expanding irrigation, building dams, transporting water across catchment boundaries, and pumping groundwater (Abbott et al. 2019). It is challenging to represent these human water cycle alterations in GWMs, making it very difficult to distinguish natural and anthropogenic components in hydrological signals as a consequence (Salwey et al. 2023). Many GWMs now represent water management processes (e.g., irrigation, domestic water use, reservoir operation, and groundwater pumping), but with great difficulty, especially when trying to capture complex human decision-making processes and field-scale management practices (McDermid et al. 2023). The representation of water management is often challenging because of data paucity, especially at the global scale [e.g., water used for irrigation, manufacturing, domestic use, and cooling processes, as well as reservoir operation rules (Pokhrel et al. 2016; Wada et al. 2017)]. Future projections considering human activities are even more problematic because scenarios of future water use and management practices are almost nonexistent due to limited data from the past and a lack of approaches to model the future. This applies not only to irrigation water use but also to water use in growing urban areas and especially in megacities that rely on remote water transfers (He et al. 2021). On the positive side, there is a growing body of literature on attributing observed changes to natural versus human drivers (Felfelani et al. 2017), even though a comprehensive quantification is often challenging.

Despite all these challenges, there are multiple possibilities for improving GWM building. Regional information can be collected through community data portals (Crochemore et al. 2020; Zipper et al. 2023), increasingly high-resolution satellite products are available, and some studies have shown that structural improvements can be derived from more informed and diverse perceptual models, for example, the inclusion of preferential recharge in karst regions, an important yet often omitted process (Hartmann et al. 2017).

2.2 | Forcing Global Models

Once a model is established and a priori parameter values have been defined based on available data, GWMs are driven by time-varying inputs of meteorological variables such as precipitation, temperature, and radiation. These inputs may be based on historical observations, as in gauge-based, satellite-based, or reanalysis products that combine observations with simulations (Beck, Vergopolan, et al. 2017; Sun et al. 2018), or climate projections of future conditions, possibly with additional downscaling and bias correction steps (Maraun et al. 2017) (see also Section 3). A less common forcing for GWMs (so far) is a reconstruction of the deeper past with paleo-hydro-climatic conditions (Gladstone et al. 2007) to understand how the hydrologic cycle has evolved over long time scales.

Observation-driven products depend strongly on observational data, which means that any uncertainties will propagate into the final forcing product. This is due to uncertainties in the measurements themselves (e.g., satellite or station data) and in interpolation and modeling techniques to derive spatial data fields, for example, for precipitation (Viviroli et al. 2011). For example, precipitation stations cover only a small area of the world (likely $< 1\%$ of the Earth's surface is represented; Kidd et al. 2017), and generally, fewer and more uncertain observations are available in mountainous or economically poorer regions, leading to unbalanced datasets (Viviroli et al. 2011). Reanalysis products do not assimilate precipitation data (e.g., from precipitation gauges), so they are more uncertain and need biased adjustment against observations when used for historical simulations or forecasting by GWMs (Berg et al. 2021). The data availability issue also affects simulations of the future, for example, for climate impact studies because uncertainty in historical observations still matters as we require them as reference data (Tarek et al. 2021).

Uncertainty in GWMs forcing for projections of future climates is caused by three primary factors: GCM/ESM or structural uncertainty (e.g., different models giving different outcomes for the same data or initializations; Wu et al. 2024), scenario-related uncertainty (e.g., differences in outcomes due to varying scenarios/input specified, e.g., atmospheric composition), and uncertainty caused by internal variability (i.e., arising from natural processes such as multi-decadal oscillations; Deser et al. 2020; Lehner et al. 2020). The specific contribution of these three components to the total uncertainty generally depends on the time horizon considered (the more distant the time window is from the present states, the higher the uncertainty), the variable of interest (e.g., uncertainty in precipitation is generally higher than for temperature), the GCM/ESM used, and the geographic region (Schwarzwald and Lenssen 2022). While structural and scenario uncertainties are important, uncertainties from internal variability can account for over 50% of the total uncertainty in climate projections (Xie et al. 2015; Kumar and Ganguly 2018; Deser et al. 2020; Schwarzwald and Lenssen 2022). This implies that the uncertainty in climate forcing should be a key consideration in future projections by GWMs, for example, by utilizing model ensembles.

2.3 | Evaluation of Global Models

Evaluation is a key step to assess a model's ability to perform a specific task, or to adjust its parameters as part of an iterative calibration process (sometimes referred to as tuning). Ideally, model evaluation should be diagnostic (Gupta et al. 2008) and help to identify model deficiencies in, for example, capturing water fluxes and storage dynamics. Evaluating process realism (i.e., consistency or model fidelity) is an important step towards enhancing the credibility of GWMs' multiple uses (Figure 1), for example, for climate change impact assessments (Krysanova et al. 2020).

GWMs are mostly evaluated by comparing simulated and observed streamflow time series (Arheimer et al. 2020), given that streamflow is relatively widely available and provides information for model performance across a larger region—the upstream catchment. Other fluxes and state variables, such as evapotranspiration, snow, terrestrial water storage, and soil moisture, are less commonly used (Pimentel, Arheimer, et al. 2023), but are gaining more traction with new evaluation protocols (Collier et al. 2018; Tiwari et al. 2025). Globally, datasets of streamflow are biased towards large rivers (Downing 2012) with extensive anthropogenic influences (Wagener and Montanari 2011; Krabbenhoft et al. 2022), while other variables like groundwater recharge are often only available as long-term averages and certain climatic regions (dry regions; Moeck et al. 2020). Some variables are not measured at all (e.g., lateral groundwater fluxes, macropore flow), and others are measured at scales that are very different from the current model scales [e.g., evapotranspiration (Wartenburger et al. 2018), soil moisture (Crow et al. 2012), or groundwater table depth (Reinecke et al. 2020) and groundwater recharge (Moeck et al. 2020)]. Due to the insufficient length and homogeneity of observational records, long-term variations and trends in components of the water cycle can only be quantified with large uncertainty (Dorigo et al. 2021). The uncertainties of observational data themselves are also rarely accounted for (even if they can be substantial; Westerberg and McMillan 2015) during GWM evaluation (or calibration; Moges et al. 2021), and little is known about how this in turn affects the predictive uncertainty of GWMs.

Choices of error metrics in model evaluation can be a substantial source of uncertainty as well. Different metrics emphasize different aspects of model performance, for example, bias, variability, and timing (Gupta et al. 2009). While an array of metrics exists, most studies have focused on the Nash–Sutcliffe efficiency (NSE) and similar statistical metrics (Melsen et al. 2025) and thus on evaluating the long-term water balance rather than extremes (Zaherpour et al. 2018). This has been criticized because these metrics do not provide diagnostic insight into model behavior (Gupta et al. 2008), are not tailored to the specific research question (e.g., flood prediction), and because of limitations in interpreting and comparing the error metrics themselves. For instance, the choice of an acceptable threshold for goodness of fit metrics is subjective and differs between error metrics (Knoben et al. 2019) and between hydrological systems (Schaefli and Gupta 2007), calling for more nuanced approaches in the future. First, a standard set of performance metrics and reasonable benchmarks (Seibert et al. 2018; Knoben 2024) would facilitate an impartial comparison between models and allow tracing of model improvement over time. Second, a move away from statistical metrics towards hydrological signatures (Donnelly et al. 2016; McMillan 2020) that focus on hydrologically relevant aspects of model outputs (e.g., low flows or high flows) would provide an evaluation strategy that is more fit for purpose (e.g., if we are interested in extremes such as droughts or floods; Gupta et al. 2008). There is a wide range of signatures and extensive experience from catchment modeling that can be utilized to this end (Kuentz et al. 2017; McMillan 2020; McMillan 2021). Note, though, that GWMs are typically uncalibrated, have a less tailored model structure, and are forced with climate data that are often subject to high biases, so a lower performance is to be expected when comparing global models with local or regional models for a specific region.

Evaluating the accuracy of hydrological extremes (floods and droughts) is of critical importance since GWM outputs are often used to inform both early warnings in data-scarce regions (e.g., Harrigan et al. 2020) as well as long-term policy recommendations to adapt to these extremes (e.g., Dottori et al. 2018), which is heavily featured in the IPCC impact report. For example, the Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL) land surface model is used in combination with a routing and inundation model to create the input for the Global Flood Awareness System (GLOFAS; Harrigan et al. 2020), which provides flood warnings for the world, including data-scarce regions. For predictions of local inundation areas, an additional high-resolution routing and/or inundation model (e.g., LISFLOOD; Harrigan et al. 2020) is necessary. This adds further uncertainty in the modeling chain and the need for additional evaluation.

Direct comparisons with observations (e.g., with statistical metrics as described above) are difficult in data-sparse regions (an especially common problem for global models), in regions with substantial anthropogenic impacts (Döll et al. 2014), and generally not possible for the future. Thus, alternative evaluation strategies have been proposed. To provide information for evaluation, also in regions without measurements, we may use regionalized streamflow signatures (Troy et al. 2008) or functional relationships that capture the co-variability of forcing and response variables in space (e.g., Randerson et al. 2009; Koster and Mahanama 2012; Gnann et al. 2023). As a complementary method to point-by-point comparisons with historical data, evaluation focusing on input–output relationships can help to reveal additional insights into model functioning (e.g., Luo et al. 2012; Mahnken et al. 2022). This is particularly relevant for climate change impact studies, where response-based analysis methods can provide insight into whether a model is fit for purpose, for instance, by showing whether a model's sensitivity to changing forcing is as expected (Wagener et al. 2022; Zhang et al. 2023).

If the evaluation shows discrepancies between observations and model results, calibration is a potential strategy to reduce them. Few global models so far have been calibrated against observed variables to improve a priori parameter estimates (Kupzig et al. 2023). In light of the equifinality of parameter sets and the risk that the model is adjusted only for the variable that is used for calibration, but not for other flux or state variables, multi-criteria calibration of GWMs with different observables and signatures has been recommended (Döll et al. 2016;

Pimentel, Crochemore, et al. 2023). Model calibration beyond streamflow with multiple remote-sensing-based observations, such as evapotranspiration, snow cover, snow water equivalent, soil moisture, terrestrial water storage, water level, surface temperature, among others, has been realized for several river basin studies (Meyer Oliveira et al. 2021). However, neither single- nor multi-objective calibration has become the general practice for GWMs (Telteu et al. 2021). Apart from a lack of observational data for many regions, the observations themselves may carry large uncertainties (Pimentel, Crochemore, et al. 2023) and calibration procedures can be computationally expensive. With km-scale models and nested catchments, calibration can be very time- and resource-consuming. Ultimately, calibration on historical observations does not ensure that GWMs are also providing robust projections of future changes (Wagener et al. 2022), for example, a particular parameter might not be adequate to represent streamflow processes if conditions have changed substantially due to climate and land use change (Milly et al. 2008). Recent studies also highlight that calibration might not be necessary for warm and humid regions (Zhao et al. 2025), which is in line with findings from catchment studies in similar regions (van Werkhoven et al. 2009).

Evaluating or calibrating GWMs requires observations of (at least some) simulated variables. Given that in situ observations do not seamlessly cover all land areas, there is the prospect of using satellite-based products with global coverage. Terrestrial water storage (TWS) is a variable estimated using satellite gravimetry (GRACE, GRACE-FollowOn; Landerer et al. 2020; Rodell and Reager 2023) and has become an important observation for assessing GWM performance as TWS is an integrative hydrologic state variable (Lee et al. 2023). First efforts have been made to assimilate GRACE data in GWMs (Tian et al. 2019). However, TWS products for model evaluation also hold substantial uncertainties due to the coarse spatial and temporal resolution (Scanlon et al. 2018; Rodell and Reager 2023) and complex attribution to specific water storage components (Döll et al. 2014). Other satellite missions like Landsat (e.g., extent of surface water bodies, surface temperature, land use change), altimetry missions for water level of inland water bodies, and SMAP (Soil Moisture Active Passive) have also proven to be essential sources to evaluate and build GWMs (McCabe et al. 2017), although they may also include considerable uncertainties (Pimentel, Crochemore, et al. 2023). Among recent satellite missions, the Surface Water and Ocean Topography (SWOT) mission is expected to be an important step for assessing surface water dynamics and storage variations of inland water bodies (instead of water level only) with unprecedented spatial resolutions and global coverage, providing insights into small and otherwise ungauged water bodies at scales of about 100 m (Papa and Frappart 2021).

To facilitate a more structured comparison between models, model intercomparison projects have gained importance by providing modeling protocols that define standardized forcing data, scenarios, and other modeling choices. These have been carried out very successfully in the climate community with the Climate Model Intercomparison Project (CMIP; Eyring et al. 2019). The Earth System Modeling community has implemented the ILAMB (International Land Model Benchmarking) benchmark, which offers a structured comparison of models to observations in a standardized software package (Collier et al. 2018). The ISIMIP project, specifically the global water sector (Frieler et al. 2017), has also developed standardized protocols for joint simulations to evaluate models, which have already yielded multiple insights. Models consistently perform better concerning streamflow in wetter than in drier climates (Zaherpour et al. 2018; Heinicke et al. 2024) and show large uncertainties for specific variables, for example, groundwater recharge (Reinecke et al. 2021). Further, different models fail to consistently reproduce both fluxes and storages with the same efficacy, and model performance is relatively poor in regions with high human impacts (Tiwari et al. 2025). Intercomparison projects can further be used to explore the differences in model structure, for example, in regards to the representation of human water uses (Telteu et al. 2021) and explore model differences in representing process relationships (Gnann et al. 2023).

To summarize, building GWMs remains challenging regarding the identification of adequate model structures that reflect current perceptions of heterogeneous hydrologic systems, estimating parameters, simulating human activity, driving models with uncertain inputs, and evaluating them with limited observations. All these issues require further study to understand and quantify existing uncertainties and their origins and to understand their implications for GWM applications. On the positive side, new and growing datasets, alternative methods for model evaluation, and increasing computational resources have the potential to push forward the development of GWMs.

3 | Uncertainties in Simulating the Past, Future, and Near Future Water Cycle

In this section, we discuss how GWM uncertainties influence simulations of past, future, and near-future water cycles. We focus on six essential hydrological variables: streamflow, evapotranspiration, groundwater recharge, soil moisture, terrestrial water storage, and anthropogenic water use (Table S3).

3.1 | Simulating the (Recent) Past

Reconstructions of the terrestrial water cycle over the last 100 years include different sources of uncertainty, such as model conceptualization and parameterization, meteorological forcing, and anthropogenic influences that will impact simulated hydrological variables (Section 2). This period largely covers our observational records (and thus enables the use of reanalysis products) and includes the main upswing of global economic growth after the Second World War. Multiple model comparison studies and global water balance studies that include GWM outputs reveal substantial uncertainties, even for global average fluxes.

For streamflow, studies estimate ranges from less than 40,000 km^3/year to over 60,000 km^3/year globally [<300 to >450 mm/year] (Haddeland et al. 2011; Schellekens et al. 2017; Abbott et al. 2019; Rockström et al. 2023) (Figure 3a for differences in a GWM ensemble). Correspondingly, estimates of evapotranspiration range from approximately 60,000 to over 80,000 km^3/year [450–600 mm/year] (Haddeland et al. 2011; Schellekens et al. 2017; Abbott et al. 2019) (see also Figures 3b

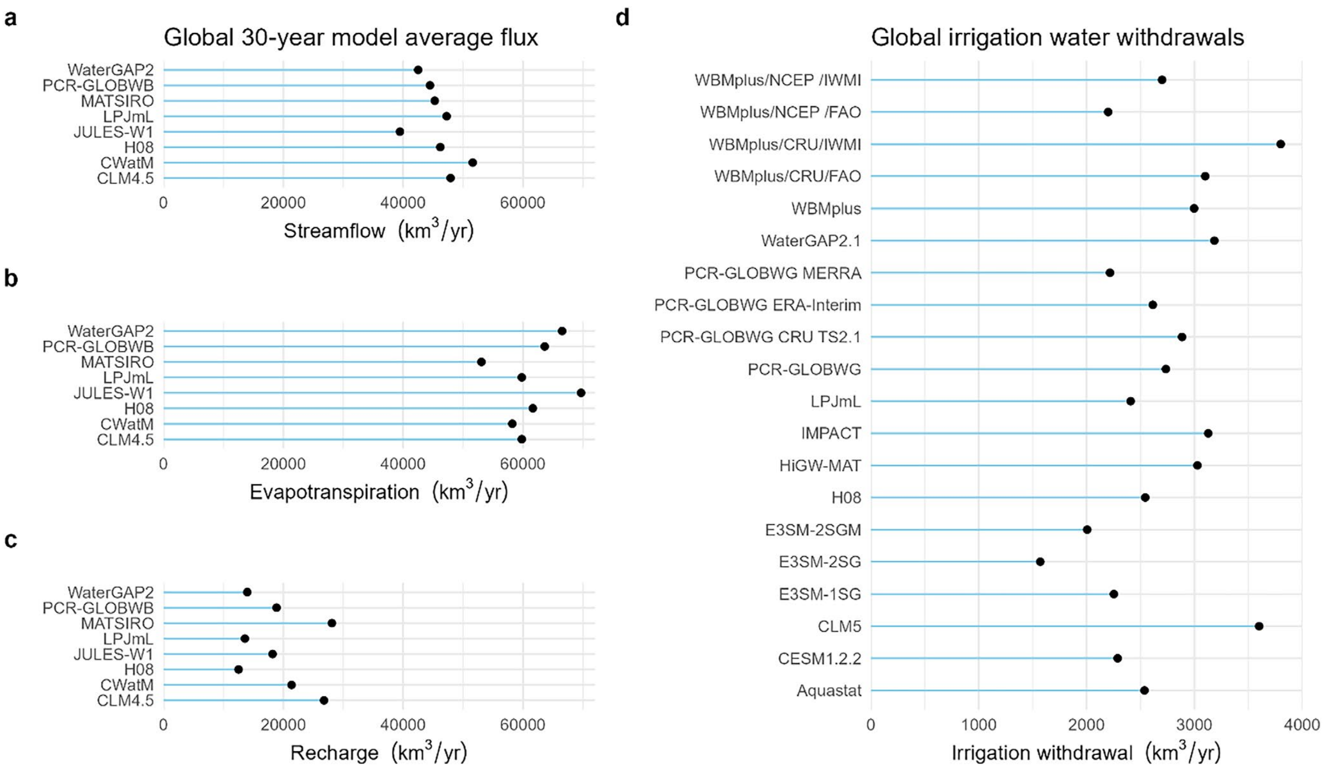

FIGURE 3 | Disagreements between key variables of global water models. (a–c) Differences between multiple global water models for streamflow, evapotranspiration, and groundwater recharge (replotted from Table S4 in Gnann et al. (2023) with a factor of 132 for conversion to km^3, which is an estimate of average model cell size). Note that while all models are forced with the same input data, streamflow, and evapotranspiration may not always add up to precipitation. (d) *Irrigation*: Estimated global irrigation water withdrawals show large disagreement between models, with global water models tending to show larger values than reported by the United Nations Food and Agriculture Organization in Aquastat (replotted from McDermid et al. 2023).

and 5d). The (relative) disagreement is similarly large for fluxes such as groundwater recharge, ranging from approximately 12,000 to 25,000 km^3/year [90–190 mm/year] (Abbott et al. 2019; Rockström et al. 2023) (see also Figure 3c), with a recent data-based study suggesting that models generally underestimate diffuse groundwater recharge (water that percolates down to groundwater rather evenly across the landscape, in contrast to focused recharge that enters groundwater at certain "focused" points, such as rivers and lakes) compared to observations [observation-based estimate: 218 mm/year] (Berghuijs et al. 2022).

For terrestrial water storage, we can only assess relative differences between models and observations, often done by examining anomalies and trends. For example, Scanlon et al. (2018) noted substantial uncertainties among different models, reporting that terrestrial water storage anomaly trends (summed over all investigated basins) are "positive for GRACE (~71–82 km^3/year) but negative for models (−450 to −12 km^3/year)." A recent study that used multiple GHMs participating in the third phase of ISIMIP demonstrated that while TWS simulations have been improved in newer model versions, different models do not consistently reproduce fluxes and storages with the same efficacy across varied geographic and climatic regions (Tiwari et al. 2025). The study also demonstrated that model performance for river discharge deteriorates largely with increasing human influence. Storages, such as groundwater, accumulate errors (in contrast to fluxes) and may exhibit long-term memory. This is often not captured by the current generation of models that usually contain very simplified representations of groundwater systems, for example, bucket models that may not be able to represent long-term storage depletion (Fowler et al. 2020).

Anthropogenic water use is particularly uncertain, especially for irrigation, which accounts for ~70% of global water withdrawals, with estimates collated by a recent study ranging from 1571 to 3800 km^3/year (McDermid et al. 2023; i.e., differences by a factor of 2; Figure 3d). Uncertainties of water use are, however, not necessarily a direct consequence of GWM uncertainty but a reporting issue since minor changes to definitions can change reporting of water consumption of, for example, the US energy system, substantially (Grubert et al. 2020). Another example is the area equipped for irrigation data, which is used in multiple global water models and datasets (Siebert et al. 2015). If a municipality or private entity reports an area as being equipped with irrigation equipment, it includes planned-but-not-implemented and existing-but-not-used equipment as well. While challenges especially remain in correctly allocating withdrawals to their sources (e.g., surface water or groundwater), determining the withdrawals in the first place is highly uncertain as, for example, uncertain information on where irrigation equipment exists is fed into an uncertain irrigation model to determine irrigation estimates (McDermid et al. 2023).

While global long-term averages already reveal large differences, these uncertainties are even larger in specific regions, during specific time periods, and for extremes. Streamflow is typically more uncertain, in relative terms, in dry regions (Zaherpour et al. 2018; Heinicke et al. 2024), and in lake-rich and snowy places (Giuntoli et al. 2015; Beck, van Dijk, et al. 2017; Gädeke et al. 2020). Possible reasons are that in these regions, precipitation is a weaker or less direct control on streamflow, because other processes (e.g., related to filling and spilling the storage) and their representation in GWMs are more important. For example, seasonal dynamics are often poorly captured and timing bias of the annual flow maximum is regularly more than 1 month, in part because of poor representation of snow, lake dynamics, and other storage processes (Gudmundsson, Wagener, et al. 2012; Zaherpour et al. 2018). Also, studies tend to find larger uncertainties for extremes such as high and low flows (Gudmundsson, Tallaksen, et al. 2012; Schewe et al. 2019), which are particularly important for flood and drought impact analyses. For instance, assessing modeled streamflow during the European heatwave in 2003, Schewe et al. (2019) found that many models underestimate low flows compared to observations, likely because these models poorly represent groundwater flow to streams that become particularly important during dry conditions. Devitt et al. (2021) tested four global models regarding their ability to simulate historical floods in the USA and found that two models underestimated floods by more than 25% in roughly two-thirds of all catchments, while the other two overestimated flows by the same amount in a similar fraction of catchments. In a global analysis, Heinicke et al. (2024) found GWMs tend to overestimate both high flows and low flows. For both extremes, predictions in arid regions are less accurate than in humid regions. Moreover, Kumar et al. (2022) showed that while GWMs were not able to capture observed hydrologic drought events (quantified by runoff-deficits above the Q80), they performed similarly to catchment scale models in estimating runoff-based standardized drought index (Kumar et al. 2022). Given that natural hazards such as floods and droughts pose a severe threat to society and are projected to increase in many regions around the globe (Tabari et al. 2021), it will be critical to ensure that GWMs realistically simulate these extremes.

3.2 | Simulating the Near Future—Seasonal Forecasts

Generation of near-future (seasonal to sub-seasonal) climate and hydrological forecasts is crucial for integrated water resources management as well as for generating early warnings for hazards, such as floods and droughts, which can have long transition periods. While seasonal climate predictions (see Figure S2 for mapping meteorological to hydrological forecasts based on timescales) are common practice, their utilization for global hydrological forecasts remains uncertain and is largely under development and not yet operational, except for flood forecasting (https://www.globalfloods.eu). The quality/accuracy of global seasonal hydrological forecasts is driven by three major factors: (a) initial hydrological conditions, which consider various observed variables (both hydrological and meteorological), (b) GWM uncertainties, and (c) seasonal climate forecasts accuracy, which depends on historical climate data and climate forecasts (temperature and precipitation being of utmost significance to hydrological models; Figure S3; WMO 2021). However, the contribution of (a)–(c) to the uncertainty of the seasonal hydrological forecast varies with hydroclimatic zones (Shukla et al. 2013), size of the catchment (Paiva et al. 2012), chosen reference year(s) (Shukla et al. 2013; Sinha and Sankarasubramanian 2013), and prediction method (Wood et al. 2016). An important component to counteract these uncertainties is data assimilation (see also Section 4.1.) which is pivotal for state updates when simulating the near future (Zhang et al. 2021).

3.3 | Simulating the Far Future—Climate Change and Water Use Projections

The further we move away from the instrumental record of observations, into the past or future, the more we expect uncertainties to grow due to the influence of scenarios and other choices (e.g., quantification of human influences, land-cover changes; Collins et al. 2012). Uncertainties related to the reconstruction of the recent (50–100 years) water cycle by GWMs are thus naturally smaller than the uncertainties related to future projections. This is because meteorological forcings are constrained by data assimilation (reanalysis data instead of climate model projections), actual land use data are obtained from remote sensing (at least in recent decades), and socio-economic forcings (GDP, population, water demand) are constrained by regular country reporting. In addition, GWMs may become more uncertain in future regimes for which they were not developed or calibrated, for example, because of changes in biophysical processes related to CO_2 fertilization.

A question often posed is whether input uncertainty or model structure uncertainty dominates the uncertainty in the model output. Is the uncertainty in climate forcing originating from GCMs or are the differences related to the GWMs? While this has been evaluated in multiple studies (Prudhomme et al. 2014; Schewe et al. 2014; Giuntoli et al. 2015; Wartenburger et al. 2018; Reinecke et al. 2021), the answer depends on which models, variables, time periods, and geographic regions are included in the analysis. Wartenburger et al. (2018) showed that evapotranspiration differences between different model choices largely explain overall variability but that the spatio-temporal differences can mainly be explained by forcing uncertainty. Schewe et al. (2014) (depicted in Figure 4) compared sources of uncertainty for streamflow and found high spatial variability in which sources dominated. In some warm arid regions, GWMs dominate uncertainty more than GCMs, while at least in some humid and some cold regions forcing is the major contributor to output uncertainty. This is comparable to the results by Giuntoli et al. (2015), who evaluated sources of uncertainty separately for low and high streamflow. In their analysis, GCMs generally dominate uncertainty for both flow regimes with exceptions in snow-dominated and arid regions. They conclude that GWMs dominate uncertainty where flow processes are more relevant than precipitation input. Somewhat similarly, if we investigate subsurface water fluxes like groundwater recharge, GWM uncertainty becomes more important because these hydrological processes are less directly controlled by climatic input. For instance, Reinecke et al. (2021) (depicted in Figure 4) found that in most regions the variability in process representation for groundwater

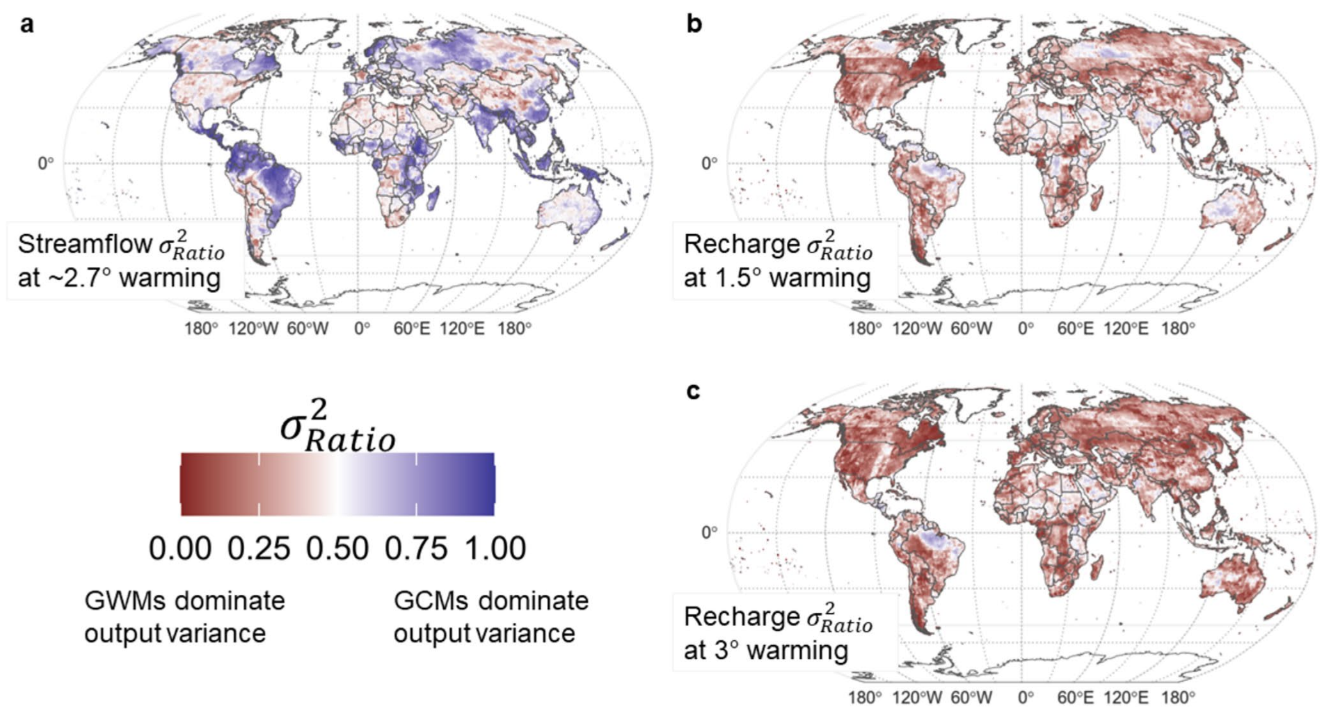

FIGURE 4 | Do Global Water Models or their forcing (Global Climate Models) dominate output uncertainty? Panels (a)–(c) show maps of GWM output variance of different GCM forcing divided by total output variance for different variables and warming scenarios (in comparison to pre-industrial temperatures). (a) Variance ratio for streamflow replotted from Schewe et al. (2014). (b) and (c) Variance ratio for groundwater recharge replotted from Reinecke et al. (2021).

recharge modeling has a larger impact on output differences than GCMs.

While these studies suggest that the ratio of GWM to GCM uncertainty depends on climate characteristics and hydrological processes, a comprehensive review of sources of uncertainty across different variables is currently lacking. As already shortly mentioned in Section 2, model ensembles are also often ensembles of opportunity rather than ensembles that were carefully selected to cover multiple uncertainty aspects (as, e.g., done by using multiple plausible conceptual models as basis for modeling a certain groundwater system; Enemark et al. 2019). It would be worthwhile to study systematically for which variables, which time periods, and which regions each of these uncertainties dominate globally to show where future model improvement would have the most leverage.

Uncertainty originating from the GWMs in the context of climate projections is at least partially related to the representations of various biophysical processes, such as those related to vegetation and soils (Wartenburger et al. 2018), which influence how future atmospheric moisture and energy translate into hydrologic impact variables (Samaniego et al. 2017). How, for example, the function of the vegetation-soil sub-system will evolve under different climate change trajectories is uncertain, thus adding uncertainty to estimates of evapotranspiration or percolation. One would expect that the overall uncertainty increases with longer projection horizons, yet current literature is inconclusive as to whether GCMs or GWMs dominate projection uncertainty. For example, Pokhrel et al. (2021) found that for terrestrial water storage, GCM uncertainty is significantly larger than GWM uncertainty for any given Representative Concentration Pathway (RCP) scenario, with variations between regions (e.g., Figure 5a). However, GWM uncertainty increases with time within a scenario, potentially surpassing GCM uncertainty for the distant future (e.g., a century into the future) (Pokhrel et al. 2021). While one might expect GCM uncertainty to become increasingly dominant in the future, this does not always seem to be the case. One reason might be that variables like TWS accumulate errors from different compartments in GWM, which can be linked to a not fully closed model water balance and thus become increasingly uncertain. Another reason might be that some water models reach thresholds at which certain processes or their numerical representations change unpredictably. Some models, for example, may encode a specific fixed process behavior or factor for dryer and wetter regions, respectively (e.g., Müller Schmied et al. 2021). If a region shifts from wet to dry in the future, this may lead to inconsistent model behaviors, thus amplifying GWM uncertainty. The uncertainty arising from socio-economic and water use scenarios is challenging to quantify owing to the lack of data and thus limited model accuracy (see also Section 2). The limited number of studies that have quantified this uncertainty indicates that future projections of water availability and use, especially for irrigation, are greatly influenced by the scenarios considered (Wada et al. 2013; Rosenzweig et al. 2014).

One aspect of projection uncertainty is that different processes may be active or inactive, or play a dominant or minor role during changing conditions. For example, CO_2 levels can increase leaf-level water use efficiency of plants, potentially offsetting reductions in water availability due to higher temperatures through reduced evapotranspiration (Rosenzweig et al. 2014; Berg et al. 2016; Lemordant et al. 2018; Hatfield

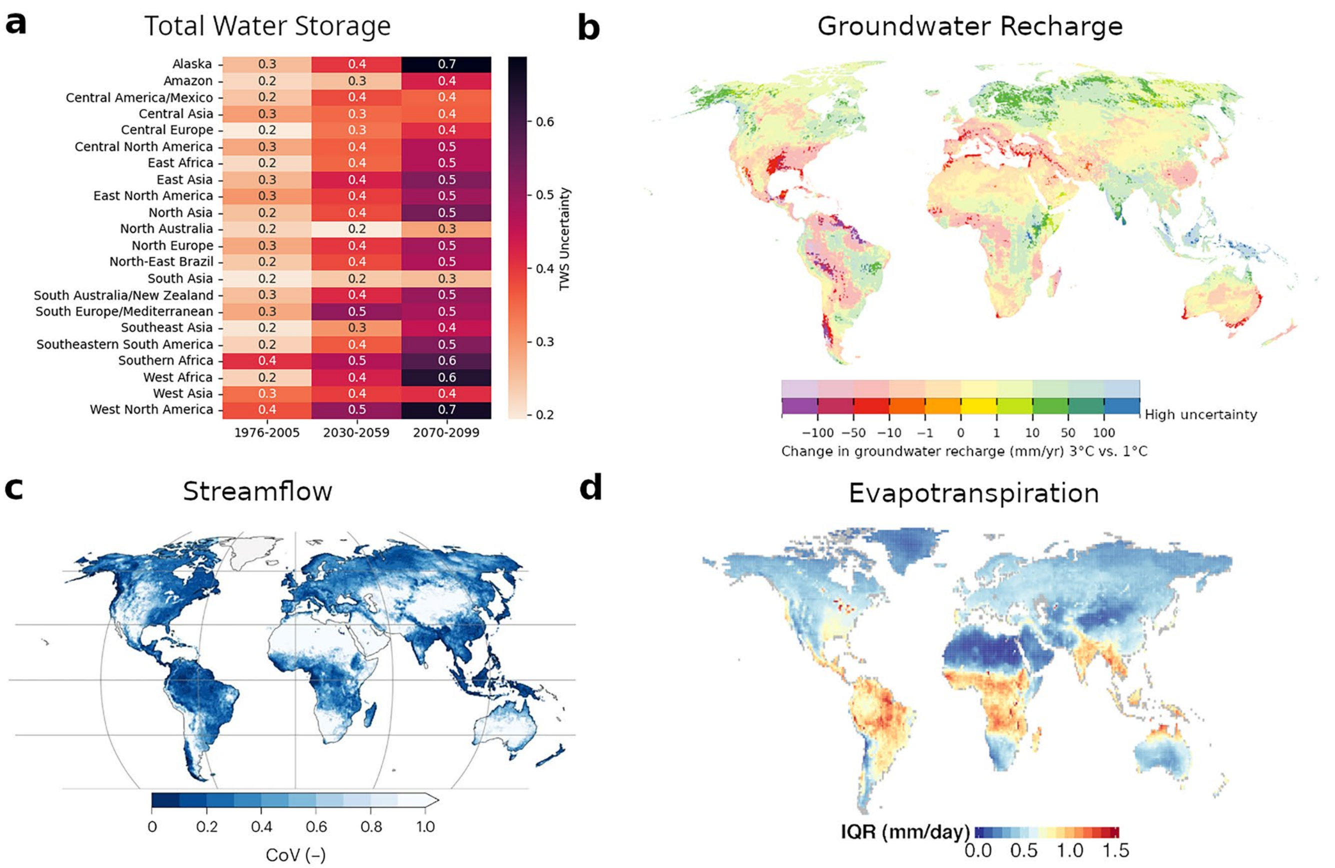

FIGURE 5 | Uncertainty in projected total water storage (a) and groundwater recharge (b), replotted from Pokhrel et al. (2021) and Reinecke et al. (2021) respectively. Uncertainty in streamflow (c) and uncertainty in evapotranspiration (d) were reproduced from Gnann et al. (2023) and Wartenburger et al. (2018), respectively. In (b), only regions where model agreement is significantly large are plotted in solid colors; all other regions are shown in an opaque color. In (c), the coefficient of variation (CoV) is calculated for an ensemble of eight GWMs over a 30-year period. In (d), the interquartile range (IQR) is shown for an ensemble of 11 GWMs. The ensembles of all four studies may not use the same models.

and Dold 2019). However, recent work also suggests that the change in water use efficiency is already exhausted due to an increased vapor pressure deficit (Li et al. 2023). Water use efficiency has been suggested to strongly influence total terrestrial runoff and evapotranspiration (Gedney et al. 2006; Piao et al. 2007), and is not represented in some GWMs despite predictions of CO_2 fertilization effects being credited as a source of uncertainties. This can lead to contradicting findings regarding the extent to which a climate-induced decline in water availability and improved plant water use efficiency counterbalance each other (Mankin et al. 2019; Adams et al. 2020; Singh et al. 2020). Uncertainty from CO_2 fertilization effects has also been linked to the uncertainty in future projections of crop productivity, irrigation water use, and groundwater recharge (Figure 5b) (Elliott et al. 2014; Rosenzweig et al. 2014; Reinecke et al. 2021). A second example is cold regions, many of which will experience considerable change as rising temperatures will affect frozen water storage in snow, glaciers, and permafrost. Given that some models are already associated with considerable uncertainty in cold and lake-rich places (Giuntoli et al. 2015; Beck, van Dijk, et al. 2017; Gädeke et al. 2020), it is unclear how robust future projections are and what role threshold behavior will play. For example, reduction in snow cover and/or greening due to vegetation growth can lead to albedo feedback, reducing streamflow due to increased net radiation (Milly and Dunne 2020). A better representation of glaciers increasing runoff in glacierized basins is already included in some models (Cáceres et al. 2020; Wiersma et al. 2022). Other aspects related to human decisions, such as land use change (Sterling et al. 2013) and water regulation (Arheimer et al. 2017), may even mask climatic change but are difficult to investigate due to a lack of historical data, as discussed earlier.

4 | Uncertainties as a Guide for the Advancement of Global Water Models

The use of global water models is unavoidably associated with uncertainties that arise during model-building and execution. These uncertainties influence simulations of past, near-future, and future water cycles. Key sources of uncertainties in GWMs are deficits and imbalances in data quality and availability across geographical regions and between hydrologic variables, poorly quantified human influences on the water cycle, and difficulties in tailoring process representations to regionally diverse hydrologic systems. Due to these GWM uncertainties, we have a limited understanding of when and where our models provide accurate results and behavior consistent with our system understanding. Specifically, it is unclear to what extent the models realistically reflect regional hydrological behavior and a distinction between natural variability and human impacts remains challenging. The scientific community can use uncertainties to guide future GWM development by gathering new, more accurate, and larger datasets, improving process knowledge, and ultimately building more consistent and accurate models. Here, we outline concrete ideas to achieve that by creating more consistent information (data and process knowledge) (Section 4.1), building more consistent and accurate models (Section 4.2), and using machine learning to gain new understanding and build different models (Section 4.3).

4.1 | Towards More Consistent Information

Data is used in all stages of the model-building and execution process. Hence, new and improved data will be critical for advancing GWMs, including data on local/regional hydrologic knowledge, newly measured or collated datasets, and more information on data uncertainties. So, where is new information likely to come from? Upcoming satellite missions will provide information on variations in inland water bodies at unprecedented spatial resolutions, providing new insights into the hydrology of ungauged regions (Papa and Frappart 2021). Some initiatives, like ESA-CCI (https://climate.esa.int), offer possibilities for better representation of, for example, land cover, which would lead to a better representation of surface heterogeneity and many other essential climate variables to be used for model forcing, parameterization, calibration, or evaluation (even if they may also carry substantial uncertainties; Pimentel, Crochemore, et al. 2023). Recent efforts have also shown that remote sensing data can be used to monitor human alterations of the water cycle, such as dam construction (Zhang and Gu 2023), and can be used to extend in situ observations of streamflow (Elmi et al. 2024).

Beyond satellites, there are ongoing efforts to collect high-resolution data that are not yet fully utilized, such as the upcoming updated catchment polygons of HydroBASINS (Lehner and Grill 2013). For example, increasing the spatial resolution of digital elevation data offers new possibilities for high-resolution river routing (Yamazaki et al. 2019). However, to profit from these advancements requires new methods to utilize such high-resolution data in the current comparably coarse-scale models. Advances in data availability also require advancements in how we merge data and models. Data assimilation methods established in other communities can be utilized to adjust GWMs and have already been adopted (Gerdener et al. 2023). In addition, subsurface data that cannot readily be acquired from satellite remote sensing, such as hydraulic conductivity and thickness of soils or aquifers, lag behind in their improvement. Importantly, satellite data, and in fact any observations, carry uncertainty that should be explicitly assessed in their effect on, for example, model calibration (Hasan et al. 2025) and propagated through GWM estimates. While propagating this uncertainty into model results is computationally expensive, advancements in computational resources and model code can contribute to feasible solutions.

The spatial resolution of GWMs is increasing, with the hope of improving model accuracies because parameter heterogeneities and spatial variability can be better resolved (Wood et al. 2011; Beven and Cloke 2012; Bierkens 2015). However, existing problems in process representation and data availability will not be eliminated through increased resolutions (Beven and Cloke 2012). Hoch et al. (2023) showed that improvements may be found for streamflow but not necessarily for other hydrological variables, due to a lack of accurate high-resolution forcing data, inaccurate scale transitions of model parameters, and challenges in realistically representing scale-dependent processes (Beven and Cloke 2012). However, increasing resolutions might enable more regional information to be used in GWMs, for example, by assimilating regional system conceptualizations. In addition, this could also lead to a certain convergence of global models and (local or regional) catchment models. While we would not expect global models to perform as well as tailored catchment models, the more detailed regional models or datasets can serve as benchmarks and potentially help in identifying the most important structural or parametric deficiencies in global models (Arheimer et al. 2012) or shortcomings of global datasets (Clerc-Schwarzenbach et al. 2024).

Besides collecting new data, the scientific community started investing resources in collecting local knowledge to improve model structures and tailor them to specific regions. For example, community portals (Crochemore et al. 2020; Zipper et al. 2023) allow uploading existing local models (including data, code, and documentation) or perceptual models (McMillan et al. 2023) that encode the hydrological knowledge and human influences of particular regions. Partnering with diverse disciplines could yield new data as well (Razavi et al. 2025). We can extract and synthesize information from such databases to tailor global models to specific hydro-climatic or socio-economic conditions around the world. New research may also extract knowledge from the enormous number of existing publications, which could be automatically built into new global datasets of regional knowledge (Stein et al. 2022; McMillan et al. 2025).

Existing data may also be found in nonscientific resources and can be combined into valuable products of human influences on the water cycle, such as freshwater demand for energy production (Gerbens-Leenes et al. 2024). Local knowledge can be combined with remote sensing into joint datasets. For example, the Global Gravity-based Groundwater Product (G3P; Güntner et al. 2024) combines different observational products into one global product of groundwater storage variations (other examples of combined datasets exist, such as Moeck et al. 2020). Available data can also be transferred to regions with a lack of data, for example, to synthesize new irrigation water use data (Kragh et al. 2025). Furthermore, datasets often lack uncertainty quantification, which would be valuable to rigorously test how uncertainties affect GWM output uncertainties (when used to force or calibrate models) and allow for a more improved model evaluation (Kiang et al. 2018; Beven et al. 2020; Pimentel, Crochemore, et al. 2023). Model evaluation should also include not just streamflow but other components of the water cycle, such as evaporation (Pimentel, Arheimer, et al. 2023), groundwater recharge (Wan et al. 2024), soil moisture (Crow et al. 2012), snow (Arheimer et al. 2017), or TWS (Döll et al. 2024). However, this again requires datasets that are less biased towards specific regions, quantified uncertainties of the measurements themselves, and methods to compare point measurements to coarse-scale model estimates.

4.2 | Towards More Consistent and Accurate Models

Models always simplify reality, and many choices can be made to build and evaluate them. A better representation of the diversity of hydrologic systems, new model evaluation strategies, more model intercomparison, and better modeling software can yield GWM more consistent with our current process knowledge and provide accurate results. Increasing awareness of the problem of model structural uncertainty has led to a range of modular modeling frameworks that maximize flexibility to allow users, in principle, to tailor models to the specific perceptual model of a particular system (Clark et al. 2008; Fenicia et al. 2011; McMillan et al. 2011; Dembélé et al. 2020). It has also led to increasing efforts to highlight and demonstrate the need to explicitly formulate perceptual models to distinguish uncertainty in system perception from uncertainty in model implementation (Wagener et al. 2021; McMillan et al. 2023). GWMs have not yet fully explored the possibility of modularity. This is due to the high diversity of the subsystems (Figure 2d) and missing approaches to disaggregate them sensibly and efficiently on the global scale. Global models already separate between, for example, humid and arid regions (Müller Schmied et al. 2021) or mountains and plains (Hoch et al. 2023), but do not account for the full hydrologic diversity that exists. To tailor GWM better to local hydrosystems, we need GWM software architectures that allow for flexible parameterization of different subsystems with diverse process equations. Secondly, we need strategies and initiatives to collect, validate, organize, store, and share the vast knowledge of diverse hydrologic systems in a structured way that allows a robust, transparent, and computationally efficient integration in GWMs. Finally, the diversity of global hydrologic systems also underlines the need for methods to analyze and evaluate the diverse outputs of GWMs. Likely, the lack of structural tailoring and, thus, limited flexibility of current global models entails significant structural uncertainty.

Collecting existing regional knowledge provides the opportunity to tailor GWMs to particular regions. However, current modeling frameworks (Clark et al. 2008; McMillan et al. 2011; Clark, Nijssen, et al. 2015) that allow for a more modular approach to hydrologic modeling require information about the model structure that best fits particular hydrologic settings, something we rarely have. Modular catchment-scale modeling mainly works through model structure comparison using performance metrics, which do not allow for upscaling to largely ungauged global settings and suffer from model structure equifinality (Knoben et al. 2020). New approaches that express global hydrologic diversity in a model without the computational and diagnostic drawbacks of existing frameworks (i.e., sampling a multitude of model structures) are necessary. One path might be better utilization of perceptual models to guide a priori model component selection based on our hydrologic understanding (Kiraz et al. 2023). During the model-building process, different assumptions may lead to different models. Making the perceptual (or conceptual) model explicit (e.g., through a graphical representation) can help to understand assumptions about the underlying system and resulting uncertainties (Wagener et al. 2021). Making the computer code of these models openly available is equally important as it enables the community to work towards open science goals and make internal assumptions such as hard-coded empirical factors more transparent (Hutton et al. 2016).

Tailoring GWMs can be based on utilizing existing knowledge and offers the opportunity for collaboration between local-scale modeling experts and the global modeling community. Hydrologic systems with dynamics that significantly diverge from the average require structural diversity in models, as their behavior is unlikely to be captured by merely changing parameter values. One such example is carbonate rock regions in which the combination of uniform and strongly preferential recharge fluxes creates dynamics that are difficult to capture with models that do not explicitly consider preferential flow (Rahman and Rosolem 2017). Different strategies to include such processes exist and have been tested in well-monitored catchments (Rahman and Rosolem 2017) or smaller sites (Hartmann et al. 2021). Hartmann et al. (2015) and Sarrazin et al. (2018) demonstrate how both conceptual and process-based (mechanistic) model structures that reflect large-scale recharge dynamics in karst regions can be built and tested in ways that would make them transferrable to large-scale modeling efforts. Both strategies rely heavily on the utilization of generalized perceptual models for karst regions and thus expected dominant process controls. However, transferring process details from local studies to global frameworks can still be limited by the information that can be found in global datasets. For example, West et al. (2022) demonstrated that not all information used in local perceptual models of recharge processes is available in global datasets.

Another pathway to GWM improvement lies in evaluation strategies that use the global process variability represented already in these models as an advantage. Compared to single catchment models, GWMs simulate a large diversity of regions simultaneously. This enables us to search for similar patterns of hydrologic behavior (Falkenmark and Chapman 1989; Kuentz et al. 2017), for instance, functional relationships between forcing and response variables, which can be used for model diagnosis and improvement (Gnann et al. 2023). Also, the wider application of diagnostic signatures (Gnann et al. 2021; McMillan 2021) provides a pathway to improve GWM evaluation. Methods for uncertainty attribution can guide the necessary reduction in data and model uncertainty (by focusing on main sources of uncertainty) and help evaluate the models' sensitivity to future changes (Wagener et al. 2022). Connected to those ideas is the development of calibration in general and new calibration schemes adapted for GWMs as mentioned in Section 2. Currently, most of the existing methods have been developed in a catchment or regional modeling context and lack the flexibility to account for the hydrologic diversity of GWMs (Kupzig et al. 2023) and would benefit from tailoring the model structure to different regions (Beck, van Dijk, et al. 2017; Santos et al. 2022).

Model intercomparison projects have become widely used and provide another avenue for improving global models. While intercomparison projects such as CMIP, ISIMIP, and ILAMB have resulted in important insights, a challenge is the development of standardized evaluation metrics and thresholds embedded in automated comparison frameworks to gain more diagnostic insights. The first steps have been made, for example, through automated evaluation tools in CMIP (Eyring et al. 2020) and ISIMIP (https://github.com/ISI-MIP/isimip-qc). These could be extended with frameworks for hydrologic signatures (Gnann et al. 2021), unified application of performance metrics (e.g., as in Blöschl et al. 2013), and new DataMIPs that specifically investigate uncertainties in forcing and attribute data.

The GWM community increasingly couples their models with those developed by other communities (climate models, flood models, crop models, water quality models, groundwater models, socio-economic models, and many more) to obtain a more integrated Earth system view. The hope is that these models will be able to represent feedback loops that could otherwise not be simulated, such as land-atmosphere feedback (Koster et al. 2004; Zipper et al. 2019) or groundwater–surface water interactions and the supply of groundwater to the atmosphere through capillary rise (Reinecke et al. 2019). These coupled systems can then be used to, for example, investigate the nitrogen cycle (Vilmin et al. 2020), the carbon cycle (Zhang et al. 2020) or in the future possibly to better represent human interaction, for example, by coupling socio-economic models with GWMs. However, it is still unclear how model coupling affects model uncertainty. Coupling requires additional assumptions and thus likely increases uncertainty, yet examples of coupling land surface processes with atmospheric processes have shown that coupling might help to constrain model dynamics and possibly reduce uncertainty (Lewis and Dadson 2021). While some efforts have already achieved online coupling (allowing feedback to flow into both model domains; Furusho-Percot et al. 2019), others are limited by the computational burden of one model or the resulting coupling feedbacks (e.g., flood modeling; Hoch and Trigg 2019). In addition, the challenge of multi-parameter calibration (e.g., for energy and water fluxes) escalates for these coupled systems (Sellar et al. 2019).

Finally, an aspect that is often overlooked is the modeling software itself. Several GWMs have been developed for almost 30 years (e.g., WaterGAP since 1996 and VIC since 1994; Liang et al. 1994) by many students and researchers with diverse programming experience (Reinecke et al. 2022). While some have been published as open-source, most models remain closed-source projects primarily used by one research group (Melsen 2022). Openly available models rarely contain comprehensible documentation of the code itself, that is, internal documentation, and of how to use and modify it, that is, external documentation, and require experts to execute them. The unavailability and complexity of the models and their code mean that they usually do not comply with FAIR principles (Findability, Accessibility, Interoperability, Reusability; Barker et al. 2022; Nyenah et al. 2024), affecting the reproducibility of research relying on GWMs. It is currently unclear to what extent the code complexity and lack of application of established software engineering best practices affect model uncertainty (Nyenah et al. 2024). Poorly documented code may lead to unintentional wrong use, missing rigorous automated tests may lead to mistakes, and hidden physical constants and assumptions contribute to uncertainty (Mendoza et al. 2015; Cuntz et al. 2016; Hutton et al. 2016). Still, efforts have been made to make underlying assumptions and equations more transparent (Telteu et al. 2021). A likely source for this missing compliance with FAIR principles is that current funding and hiring practices undervalue model development efforts (Reinecke et al. 2022; Nyenah et al. 2024). GWMs should strive to make their code openly available, including comprehensive internal and external documentation. Open code will lead to fewer hidden and implicit assumptions, reducing model uncertainty and leading to faster progress. Modular software, for example, offers the possibility to transfer implementations (e.g., human water use, routing) between models. In addition, a more flexible code would allow for a more flexible implementation of different model structures, which in turn would be an essential step towards achieving regional tailoring. Reproducible experiments will provide a pathway to more in-depth knowledge of model differences, and more modular and modern software code will lead to experiments that can pinpoint uncertainties in process understanding more accurately.

4.3 | Machine Learning as a Complementary Modeling Approach

Machine learning (ML) methods are rapidly entering global hydrology and will likely help gather better information, fill knowledge gaps regarding hydrologic processes, human dynamics, and so forth, and build better models (Tsai et al. 2021). Purely data-driven ML has already demonstrated high performance in predicting hydrologic variables across the water cycle (Shen et al. 2021), including streamflow (Feng et al. 2020), soil moisture (Fang et al. 2017; O and Orth 2021), snow water equivalent (Meyal et al. 2020), and groundwater levels (Wunsch et al. 2022). One benefit of ML is that it absorbs information directly from data. This produces models that are inherently consistent with (observational) data, but which also inherit the imbalances and uncertainties of these data. In addition, purely data-driven models may be hard to interpret (Zaherpour et al. 2019) or helpful in making a specific scientific inquiry, as they do not directly encode physical concepts. As an important mission of GWMs is to create long-term projections under future climate, it remains unclear (and difficult to evaluate) if pure ML models are suitable for such long-term tasks. At the same time, ML models have already been shown to provide more accurate real-time flood forecasts than state-of-the-art global modeling systems, indicating their potential (Nearing et al. 2024).

Other approaches have shown the capability of ML approaches to emulate complex physical models, which could foster future uncertainty quantification in continental to global scales water models (Bennett et al. 2024). ML models can also be used to estimate human water uses and thus reduce uncertainty in GWMs (Shrestha et al. 2024). Hybrid models of classical GWMs and ML (Kraft et al. 2022; Slater et al. 2023), such as differentiable models (Shen et al. 2023), are developed to reap the benefits of both worlds while circumventing their respective limitations. Differentiable models mix process-based equations with neural networks and offer the ability to learn unknown physical relationships. By finding new relationships that govern large-scale processes of the water cycle, they could help to reduce GWM uncertainty. While ML models potentially carry less uncertainty as they can skip steps in the classical modeling chain (Nearing et al. 2021), they will ultimately also suffer from limited global hydrological information and thus will equally benefit from new, more accurate, and larger datasets (Beven 2020; Nearing et al. 2024). In the future, machine learning and process-based models are likely to be particularly powerful when used as complementary approaches that can both efficiently learn from data and be enriched with hydrological and other process knowledge (Reichstein et al. 2019).

New ways to use machine learning in modeling will come from using large language models (think ChatGPT) to support and inform decision-making along the entire modeling process (Eythorsson and Clark 2025). Large language models can help to generate local perceptual models, choose appropriate model structures, create reproducible code and documentation, and manage the data. On the one hand, such a system might be the way forward to a GWM that can quickly adapt to local information. On the other hand, there is the risk that it will only reproduce the same biases and uncertainties GWM already faces today. Given the explosion in published studies, we will not be able to avoid utilizing such help to synthesize what we know about places and processes, though its success might still depend on how well such information can be extracted from existing and future publications (Stein et al. 2022).

5 | The Future of Global Water Models

The water cycle is a central element in the Earth system, transporting and storing water, energy, nutrients, sediments, pollutants, and pathogens. Thus, the water cycle influences our climate, societal development, and the evolution of ecosystems. Global water models have evolved into widely used tools that help us understand and predict the terrestrial water cycle under past, current, and potential future conditions. Key water fluxes such as streamflow, evapotranspiration, groundwater recharge, and water storage in its various forms can now be simulated across the whole global land surface. Outputs of these models support critical global policy discussions and scientific analyses around a central resource needed by all life on Earth and a dominant source of disaster risk for society.

However, considerable uncertainties remain despite significant advances in GWMs in recent years. Key reasons for these uncertainties are that data quality and availability are highly imbalanced, that human alterations in the water cycle are poorly quantified, and that it is difficult to represent the diversity of hydrological systems around the world. Yet, these uncertainties are mostly epistemic in nature and can, in principle, be reduced if knowledge gaps are closed. Understanding the relative importance of different sources of uncertainty thus serves as a guide for prioritizing future research efforts to reduce these gaps. This will likely happen through new observations, the synthesis of existing regional hydrologic knowledge, diagnostic model evaluation strategies tailored to the specific needs and characteristics of GWMs, and through the development of improved and new modeling approaches, including machine learning and hybrid strategies. It is likely that, especially the synthesis of regional perceptual models, will lead to improved tailoring of GWM so that dominant process controls are more realistic compared with globally consistent model structures that are only tailored through adjustment of free parameters (Wagener et al. 2021; McMillan et al. 2023). This modeling step offers a tremendous opportunity for collaboration between catchment-scale and global-scale modelers.

Multiple developments will enable the advancement of global water models:

- In the future, new data sources, including those from upcoming satellite missions and open collections of existing observations, will provide considerably more information that can be assimilated into GWMs to improve model structures, parameters, and model evaluation.
- Increasing the space–time resolution of existing models further means that model parameters and state variables move closer in scale to many of the variables we can observe. Whether this will lead to improved model performances, as seen in meteorology (Bauer et al. 2015), is questionable (Beven and Cloke 2012). Examples show that increasing hydrologic model resolution does not necessarily lead to better results (Hoch et al. 2023). Also, blind spots in our observations of properties and dynamics of hydrologic systems remain despite advancements in observations (Tarasova et al. 2024).
- Significant opportunities for advancement remain in the context of GWM evaluation. Few strategies have so far been developed that benefit from the nature of GWMs, for example, by looking across large gradients in the model domain or by establishing contrasting expectations of model form (structure and parameters) and behavior (e.g., Gnann et al. 2023).
- Finally, new approaches, including those from machine learning, enable us to interrogate large and diverse datasets to estimate model parameters and structures (with more or less stringent physical constraints).

Regardless of any uncertainty reduction discussed here, it remains important to quantify and acknowledge dominant uncertainties to provide as much information as possible to support decision-making or scientific inquiry. In practice, this is often done through the use of ensembles of model simulations coming from combinations of model forcing/structure/parameters, which can support robust decision-making under uncertainty (e.g., Smith et al. 2018).

Keeping track of and utilizing advancements in machine learning, observational capabilities, computer science, and other relevant fields will be increasingly difficult for individuals or even for specific research groups. Integrating diverse knowledge and skill sets will be critical to developing and maintaining a highly dynamic research field. In addition, various other fields within Earth Sciences also attempt to establish global modeling capabilities, for example, vegetation modeling to understand carbon fluxes. Cross-communication and exchange will likely be highly beneficial for all areas where similar problems of building, testing, and utilizing global models exist. Therefore, the reasoning presented in this review extends beyond the topic of GWMs, providing a broader blueprint for how understanding epistemic uncertainties can serve as a guiding light for knowledge discovery and accumulation in global model improvement.

Author Contributions

Robert Reinecke: conceptualization (lead), visualization (lead), writing – original draft (lead), writing – review and editing (lead). **Lina Stein:** conceptualization (supporting), writing – original draft (supporting), writing – review and editing (supporting). **Sebastian Gnann:** conceptualization (supporting), writing – original draft (supporting), writing – review and editing (supporting). **Jafet C. M. Andersson:** writing – original draft (supporting), writing – review and editing (supporting). **Berit Arheimer:** writing – original draft (supporting), writing – review and editing (supporting). **Marc Bierkens:** writing – original draft (supporting), writing – review and editing (supporting). **Sara Bonetti:** writing – original draft (supporting), writing – review and editing (supporting). **Andreas Güntner:** writing – original draft (supporting), writing – review and editing (supporting). **Stefan Kollet:** writing – original draft (supporting), writing – review and editing (supporting). **Sulagna Mishra:** writing – original draft (supporting), writing – review and editing (supporting). **Nils Moosdorf:** writing – original draft (supporting), writing – review and editing (supporting). **Sara Nazari:** writing – original draft (supporting), writing – review and editing (supporting). **Yadu Pokhrel:** writing – original draft (supporting), writing – review and editing (supporting). **Christel Prudhomme:** writing – original draft (supporting), writing – review and editing (supporting). **Jacob Schewe:** writing – review and editing (supporting). **Chaopeng Shen:** writing – original draft (supporting), writing – review and editing (supporting). **Thorsten Wagener:** conceptualization (supporting), supervision (lead), writing – original draft (supporting), writing – review and editing (supporting).

Acknowledgments

R.R., L.S., S.G., and T.W. acknowledge support from the Alexander von Humboldt Foundation in the framework of the Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research (BMBF). Y.P. acknowledges the support from the National Science Foundation (Grant #: 1752729). M.B. acknowledges support from the ERC Advanced Grant scheme (Grant no. 101019185, GEOWAT). Open Access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Related WIREs Articles

A short history of philosophies of hydrological model evaluation and hypothesis testing

References

Abbott, B. W., K. Bishop, J. P. Zarnetske, et al. 2019. "Human Domination of the Global Water Cycle Absent From Depictions and Perceptions." *Nature Geoscience* 12, no. 7: 533–540. https://doi.org/10.1038/s41561-019-0374-y.

Adams, M. A., T. N. Buckley, and T. L. Turnbull. 2020. "Diminishing CO_2-Driven Gains in Water-Use Efficiency of Global Forests." *Nature Climate Change* 10, no. 5: 466–471. https://doi.org/10.1038/s41558-020-0747-7.

Addor, N., and L. A. Melsen. 2019. "Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models." *Water Resources Research* 55, no. 1: 378–390. https://doi.org/10.1029/2018WR022958.

Andersson, J. C., B. Arheimer, F. Traoré, D. Gustafsson, and A. Ali. 2017. "Process Refinements Improve a Hydrological Model Concept Applied to the Niger River Basin." *Hydrological Processes* 31, no. 25: 4540–4554. https://doi.org/10.1002/hyp.11376.

Arheimer, B., J. Dahné, C. Donnelly, G. Lindström, and J. Strömqvist. 2012. "Water and Nutrient Simulations Using the HYPE Model for Sweden vs. the Baltic Sea Basin—Influence of Input-Data Quality and Scale." *Hydrology Research* 43, no. 4: 315–329. https://doi.org/10.2166/nh.2012.010.

Arheimer, B., C. Donnelly, and G. Lindström. 2017. "Regulation of Snow-Fed Rivers Affects Flow Regimes More Than Climate Change." *Nature Communications* 8, no. 1: 62. https://doi.org/10.1038/s41467-017-00092-8.

Arheimer, B., R. Pimentel, K. Isberg, et al. 2020. "Global Catchment Modelling Using World-Wide HYPE (WWH), Open Data, and Stepwise Parameter Estimation." *Hydrology and Earth System Sciences* 24, no. 2: 535–559. https://doi.org/10.5194/hess-24-535-2020.

Barker, M., N. P. Chue Hong, D. S. Katz, et al. 2022. "Introducing the FAIR Principles for Research Software." *Scientific Data* 9, no. 1: 622. https://doi.org/10.1038/s41597-022-01710-x.

Bartosova, A., B. Arheimer, A. de Lavenne, R. Capell, and J. Strömqvist. 2021. "Large-Scale Hydrological and Sediment Modeling in Nested Domains Under Current and Changing Climate." *Journal of Hydrologic Engineering* 26, no. 5: 05021009. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002078.

Bauer, P., A. Thorpe, and G. Brunet. 2015. "The Quiet Revolution of Numerical Weather Prediction." *Nature* 525, no. 7567: 47–55. https://doi.org/10.1038/nature14956.

Beck, H. E., A. I. J. M. van Dijk, A. de Roo, et al. 2017. "Global Evaluation of Runoff From 10 State-Of-The-Art Hydrological Models." *Hydrology and Earth System Sciences* 21, no. 6: 2881–2903. https://doi.org/10.5194/hess-21-2881-2017.

Beck, H. E., N. Vergopolan, M. Pan, et al. 2017. "Global-Scale Evaluation of 22 Precipitation Datasets Using Gauge Observations and Hydrological Modeling." *Hydrology and Earth System Sciences* 21, no. 12: 6201–6217. https://doi.org/10.5194/hess-21-6201-2017.

Bennett, A., H. Tran, L. De la Fuente, et al. 2024. "Spatio-Temporal Machine Learning for Regional to Continental Scale Terrestrial Hydrology." *Journal of Advances in Modeling Earth Systems* 16: e2023MS004095. https://doi.org/10.1029/2023MS004095.

Berg, A., K. Findell, B. Lintner, et al. 2016. "Land–Atmosphere Feedbacks Amplify Aridity Increase Over Land Under Global Warming." *Nature Climate Change* 6, no. 9: 869–874. https://doi.org/10.1038/nclimate3029.

Berg, P., F. Almén, and D. Bozhinova. 2021. "HydroGFD3.0 (Hydrological Global Forcing Data): A 25 Km Global Precipitation and Temperature Data Set Updated in Near-Real Time." *Earth System Science Data* 13, no. 4: 1531–1545. https://doi.org/10.5194/essd-13-1531-2021.

Berghuijs, W. R., E. Luijendijk, C. Moeck, Y. van der Velde, and S. T. Allen. 2022. "Global Recharge Data Set Indicates Strengthened Groundwater Connection to Surface Fluxes." *Geophysical Research Letters* 49, no. 23: e2022GL099010. https://doi.org/10.1029/2022GL099010.

Betts, R. A., L. Alfieri, C. Bradshaw, et al. 2018. "Changes in Climate Extremes, Fresh Water Availability and Vulnerability to Food Insecurity Projected at 1.5°C and 2°C Global Warming With a Higher-Resolution Global Climate Model." *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 376, no. 2119: 20160452. https://doi.org/10.1098/rsta.2016.0452.

Beven, K. 2012. *Rainfall-Runoff Modelling.* 2nd ed. John Wiley & Sons Inc.

Beven, K. 2016. "Facets of Uncertainty: Epistemic Uncertainty, Non-Stationarity, Likelihood, Hypothesis Testing, and Communication." *Hydrological Sciences Journal* 61, no. 9: 1652–1665. https://doi.org/10.1080/02626667.2015.1031761.

Beven, K. 2018. *Environmental Modelling.* CRC Press.

Beven, K. 2020. "Deep Learning, Hydrological Processes and the Uniqueness of Place." *Hydrological Processes* 34, no. 16: 3608–3613. https://doi.org/10.1002/hyp.13805.

Beven, K., A. Asadullah, P. Bates, et al. 2020. "Developing Observational Methods to Drive Future Hydrological Science: Can We Make a Start as a Community?" *Hydrological Processes* 34, no. 3: 868–873. https://doi.org/10.1002/hyp.13622.

Beven, K. J., S. Almeida, W. P. Aspinall, et al. 2018. "Epistemic Uncertainties and Natural Hazard Risk Assessment—Part 1: A Review of Different Natural Hazard Areas." *Natural Hazards and Earth System Sciences* 18, no. 10: 2741–2768. https://doi.org/10.5194/nhess-18-2741-2018.

Beven, K. J., and H. L. Cloke. 2012. "Comment on 'Hyperresolution Global Land Surface Modeling: Meeting a Grand Challenge for Monitoring Earth's Terrestrial Water' by Eric F. Wood Et al." *Water Resources Research* 48, no. 1: 2011WR010982. https://doi.org/10.1029/2011WR010982.

Beven, K. J., M. J. Kirkby, J. E. Freer, and R. Lamb. 2021. "A History of TOPMODEL." *Hydrology and Earth System Sciences* 25, no. 2: 527–549. https://doi.org/10.5194/hess-25-527-2021.

Bierkens, M. F. P. 2015. "Global Hydrology 2015: State, Trends, and Directions." *Water Resources Research* 51, no. 7: 4923–4947. https://doi.org/10.1002/2015WR017173.

Blöschl, G., M. Sivapalan, T. Wagener, A. Viglione, and H. Savenije, eds. 2013. *Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales.* Cambridge University Press.

Bonetti, S., Z. Wei, and D. Or. 2021. "A Framework for Quantifying Hydrologic Effects of Soil Structure Across Scales." *Communications Earth & Environment* 2, no. 1: 107. https://doi.org/10.1038/s43247-021-00180-0.

Boussetta, S., G. Balsamo, G. Arduini, et al. 2021. "ECLand: The ECMWF Land Surface Modelling System." *Atmosphere* 12, no. 6: 723. https://doi.org/10.3390/atmos12060723.

Cáceres, D., B. Marzeion, J. H. Malles, B. D. Gutknecht, H. Müller Schmied, and P. Döll. 2020. "Assessing Global Water Mass Transfers From Continents to Oceans Over the Period 1948–2016." *Hydrology and Earth System Sciences* 24, no. 10: 4831–4851. https://doi.org/10.5194/hess-24-4831-2020.

Chahine, M. T. 1992. "GEWEX: The Global Energy and Water Cycle Experiment." *EOS, Transactions American Geophysical Union* 73, no. 2: 9–14. https://doi.org/10.1029/91EO00007.

Clark, M. P., M. F. P. Bierkens, L. Samaniego, et al. 2017. "The Evolution of Process-Based Hydrologic Models: Historical Challenges and the Collective Quest for Physical Realism." *Hydrology and Earth System Sciences* 21, no. 7: 3427–3440. https://doi.org/10.5194/hess-21-3427-2017.

Clark, M. P., Y. Fan, D. M. Lawrence, et al. 2015. "Improving the Representation of Hydrologic Processes in Earth System Models." *Water Resources Research* 51, no. 8: 5929–5956. https://doi.org/10.1002/2015WR017096.

Clark, M. P., B. Nijssen, J. D. Lundquist, et al. 2015. "A Unified Approach for Process-Based Hydrologic Modeling: 2. Model Implementation and Case Studies." *Water Resources Research* 51, no. 4: 2515–2542. https://doi.org/10.1002/2015WR017200.

Clark, M. P., A. G. Slater, D. E. Rupp, et al. 2008. "Framework for Understanding Structural Errors (FUSE): A Modular Framework to Diagnose Differences Between Hydrological Models." *Water Resources Research* 44, no. 12: 2007WR006735. https://doi.org/10.1029/2007WR006735.

Clerc-Schwarzenbach, F., G. Selleri, M. Neri, E. Toth, I. van Meerveld, and J. Seibert. 2024. "Large-Sample Hydrology—A Few Camels or a Whole Caravan?" *Hydrology and Earth System Sciences* 28, no. 17: 4219–4237. https://doi.org/10.5194/hess-28-4219-2024.

Collier, N., F. M. Hoffman, D. M. Lawrence, et al. 2018. "The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation." *Journal of Advances in Modeling Earth Systems* 10, no. 11: 2731–2754. https://doi.org/10.1029/2018MS001354.

Collins, M., R. E. Chandler, P. M. Cox, J. M. Huthnance, J. Rougier, and D. B. Stephenson. 2012. "Quantifying Future Climate Change." *Nature Climate Change* 2, no. 6: 403–409. https://doi.org/10.1038/nclimate1414.

Condon, L. E., S. Kollet, M. F. P. Bierkens, et al. 2021. "Global Groundwater Modeling and Monitoring: Opportunities and Challenges." *Water Resources Research* 57, no. 12: e2020WR029500. https://doi.org/10.1029/2020WR029500.

Crochemore, L., K. Isberg, R. Pimentel, L. Pineda, A. Hasan, and B. Arheimer. 2020. "Lessons Learnt From Checking the Quality of Openly Accessible River Flow Data Worldwide." *Hydrological Sciences Journal* 65, no. 5: 699–711. https://doi.org/10.1080/02626667.2019.1659509.

Crow, W. T., A. A. Berg, M. H. Cosh, et al. 2012. "Upscaling Sparse Ground-Based Soil Moisture Observations for the Validation of Coarse-Resolution Satellite Soil Moisture Products." *Reviews of Geophysics* 50, no. 2: RG2002. https://doi.org/10.1029/2011RG000372.

Cuntz, M., J. Mai, L. Samaniego, et al. 2016. "The Impact of Standard and Hard-Coded Parameters on the Hydrologic Fluxes in the Noah-MP Land Surface Model." *Journal of Geophysical Research: Atmospheres* 121, no. 18: 676–700. https://doi.org/10.1002/2016JD025097.

Dembélé, M., M. Hrachowitz, H. H. G. Savenije, G. Mariéthoz, and B. Schaefli. 2020. "Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets." *Water Resources Research* 56, no. 1: e2019WR026085. https://doi.org/10.1029/2019WR026085.

Deser, C., F. Lehner, K. B. Rodgers, et al. 2020. "Insights From Earth System Model Initial-Condition Large Ensembles and Future Prospects." *Nature Climate Change* 10, no. 4: 277–286. https://doi.org/10.1038/s41558-020-0731-2.

Devitt, L., J. Neal, T. Wagener, and G. Coxon. 2021. "Uncertainty in the Extreme Flood Magnitude Estimates of Large-Scale Flood Hazard Models." *Environmental Research Letters* 16, no. 6: 64013. https://doi.org/10.1088/1748-9326/abfac4.

Döll, P., H. Douville, A. Güntner, H. Müller Schmied, and Y. Wada. 2016. "Modelling Freshwater Resources at the Global Scale: Challenges and Prospects." *Surveys in Geophysics* 37, no. 2: 195–221. https://doi.org/10.1007/s10712-015-9343-1.

Döll, P., and K. Fiedler. 2008. "Global-Scale Modeling of Groundwater Recharge." *Hydrology and Earth System Sciences* 12, no. 3: 863–885. https://doi.org/10.5194/hess-12-863-2008.

Döll, P., H. M. M. Hasan, K. Schulze, et al. 2024. "Leveraging Multi-Variable Observations to Reduce and Quantify the Output Uncertainty of a Global Hydrological Model: Evaluation of Three Ensemble-Based Approaches for the Mississippi River Basin." *Hydrology and Earth System Sciences* 28, no. 10: 2259–2295. https://doi.org/10.5194/hess-28-2259-2024.

Döll, P., F. Kaspar, and B. Lehner. 2003. "A Global Hydrological Model for Deriving Water Availability Indicators: Model Tuning and Validation." *Journal of Hydrology* 270, no. 1–2: 105–134. https://doi.org/10.1016/S0022-1694(02)00283-4.

Döll, P., H. Müller Schmied, C. Schuh, F. T. Portmann, and A. Eicker. 2014. "Global-Scale Assessment of Groundwater Depletion and Related

Groundwater Abstractions: Combining Hydrological Modeling With Information From Well Observations and GRACE Satellites." *Water Resources Research* 50, no. 7: 5698–5720. https://doi.org/10.1002/2014WR015595.

Döll, P., and J. Zhang. 2010. "Impact of Climate Change on Freshwater Ecosystems: A Global-Scale Analysis of Ecologically Relevant River Flow Alterations." *Hydrology and Earth System Sciences* 14, no. 5: 783–799. https://doi.org/10.5194/hess-14-783-2010.

Donnelly, C., J. C. Andersson, and B. Arheimer. 2016. "Using Flow Signatures and Catchment Similarities to Evaluate the E-HYPE Multi-Basin Model Across Europe." *Hydrological Sciences Journal* 61, no. 2: 255–273. https://doi.org/10.1080/02626667.2015.1027710.

Dorigo, W., S. Dietrich, F. Aires, et al. 2021. "Closing the Water Cycle From Observations Across Scales: Where Do we Stand?" *Bulletin of the American Meteorological Society* 102, no. 10: E1897–E1935. https://doi.org/10.1175/BAMS-D-19-0316.1.

Dottori, F., W. Szewczyk, J.-C. Ciscar, et al. 2018. "Increased Human and Economic Losses From River Flooding With Anthropogenic Warming." *Nature Climate Change* 8, no. 9: 781–786. https://doi.org/10.1038/s41558-018-0257-z.

Downing, J. 2012. "Global Abundance and Size Distribution of Streams and Rivers." *Inland Waters* 2, no. 4: 229–236. https://doi.org/10.5268/IW-2.4.502.

Eagleson, P. S. 1991. "Hydrologic Science: A Distinct Geoscience." *Reviews of Geophysics* 29, no. 2: 237–248. https://doi.org/10.1029/90RG02615.

Elliott, J., D. Deryng, C. Müller, et al. 2014. "Constraints and Potentials of Future Irrigation Water Availability on Agricultural Production Under Climate Change." *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 9: 3239–3244. https://doi.org/10.1073/pnas.1222474110.

Elmi, O., M. J. Tourian, P. Saemian, and N. Sneeuw. 2024. "Remote Sensing-Based Extension of GRDC Discharge Time Series—A Monthly Product With Uncertainty Estimates." *Scientific Data* 11, no. 1: 240. https://doi.org/10.1038/s41597-024-03078-6.

Emerton, R., H. L. Cloke, E. M. Stephens, E. Zsoter, S. J. Woolnough, and F. Pappenberger. 2017. "Complex Picture for Likelihood of ENSO-Driven Flood Hazard." *Nature Communications* 8: 14796. https://doi.org/10.1038/ncomms14796.

Emerton, R. E., E. M. Stephens, F. Pappenberger, et al. 2016. "Continental and Global Scale Flood Forecasting Systems." *WIREs Water* 3, no. 3: 391–418. https://doi.org/10.1002/wat2.1137.

Enemark, T., L. J. Peeters, D. Mallants, and O. Batelaan. 2019. "Hydrogeological Conceptual Model Building and Testing: A Review." *Journal of Hydrology* 569: 310–329. https://doi.org/10.1016/j.jhydrol.2018.12.007.

Eyring, V., L. Bock, A. Lauer, et al. 2020. "Earth System Model Evaluation Tool (ESMValTool) v2.0 – An Extended Set of Large-Scale Diagnostics for Quasi-Operational and Comprehensive Evaluation of Earth System Models in CMIP." *Geoscientific Model Development* 13, no. 7: 3383–3438. https://doi.org/10.5194/gmd-13-3383-2020.

Eyring, V., P. M. Cox, G. M. Flato, et al. 2019. "Taking Climate Model Evaluation to the Next Level." *Nature Climate Change* 9, no. 2: 102–110. https://doi.org/10.1038/s41558-018-0355-y.

Eythorsson, D., and M. Clark. 2025. "Toward Automated Scientific Discovery in Hydrology: The Opportunities and Dangers of AI Augmented Research Frameworks." *Hydrological Processes* 39, no. 1: e70065. https://doi.org/10.1002/hyp.70065.

Falkenmark, M., and T. G. Chapman. 1989. *Comparative Hydrology: An Ecological Approach to Land and Water Resources.* Unesco.

Fang, K., C. Shen, D. Kifer, and X. Yang. 2017. "Prolongation of SMAP to Spatiotemporally Seamless Coverage of Continental U.S. Using a Deep Learning Neural Network." *Geophysical Research Letters* 44: 11,030–11,039. https://doi.org/10.1002/2017GL07561.

Fatichi, S., D. Or, R. Walko, et al. 2020. "Soil Structure Is an Important Omission in Earth System Models." *Nature Communications* 11, no. 1: 522. https://doi.org/10.1038/s41467-020-14411-z.

Felfelani, F., Y. Wada, L. Longuevergne, and Y. N. Pokhrel. 2017. "Natural and Human-Induced Terrestrial Water Storage Change: A Global Analysis Using Hydrological Models and GRACE." *Journal of Hydrology* 553: 105–118. https://doi.org/10.1016/j.jhydrol.2017.07.048.

Feng, D., H. Beck, K. Lawson, and C. Shen. 2023. "The Suitability of Differentiable, Physics-Informed Machine Learning Hydrologic Models for Ungauged Regions and Climate Change Impact Assessment." *Hydrology and Earth System Sciences* 27, no. 12: 2357–2373. https://doi.org/10.5194/hess-27-2357-2023.

Feng, D., K. Fang, and C. Shen. 2020. "Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales." *Water Resources Research* 56: e2019WR026793. https://doi.org/10.1029/2019WR026793.

Fenicia, F., D. Kavetski, and H. H. G. Savenije. 2011. "Elements of a Flexible Approach for Conceptual Hydrological Modeling: 1. Motivation and Theoretical Development." *Water Resources Research* 47: W11510. https://doi.org/10.1029/2010WR010174.

Fowler, K., W. Knoben, M. Peel, et al. 2020. "Many Commonly Used Rainfall-Runoff Models Lack Long, Slow Dynamics: Implications for Runoff Projections." *Water Resources Research* 56, no. 5: e2019WR025286. https://doi.org/10.1029/2019WR025286.

Frieler, K., S. Lange, F. Piontek, et al. 2017. "Assessing the Impacts of 1.5°C Global Warming—Simulation Protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b)." *Geoscientific Model Development* 10, no. 12: 4321–4345. https://doi.org/10.5194/gmd-10-4321-2017.

Furusho-Percot, C., K. Goergen, C. Hartick, K. Kulkarni, J. Keune, and S. Kollet. 2019. "Pan-European Groundwater to Atmosphere Terrestrial Systems Climatology From a Physically Consistent Simulation." *Scientific Data* 6, no. 1: 320. https://doi.org/10.1038/s41597-019-0328-7.

Gädeke, A., V. Krysanova, A. Aryal, et al. 2020. "Performance Evaluation of Global Hydrological Models in Six Large Pan-Arctic Watersheds." *Climatic Change* 163, no. 3: 1329–1351. https://doi.org/10.1007/s10584-020-02892-2.

Gedney, N., P. M. Cox, R. A. Betts, O. Boucher, C. Huntingford, and P. A. Stott. 2006. "Detection of a Direct Carbon Dioxide Effect in Continental River Runoff Records." *Nature* 439, no. 7078: 835–838. https://doi.org/10.1038/nature04504.

Gerbens-Leenes, P. W., S. D. Vaca-Jiménez, B. Holmatov, and D. Vanham. 2024. "Spatially Distributed Freshwater Demand for Electricity in Africa." *Environmental Science: Water Research & Technology* 10, no. 8: 1795–1808. https://doi.org/10.1039/D4EW00246F.

Gerdener, H., J. Kusche, K. Schulze, P. Döll, and A. Klos. 2023. "The Global Land Water Storage Data Set Release 2 (GLWS2.0) Derived via Assimilating GRACE and GRACE-FO Data Into a Global Hydrological Model." *Journal of Geodesy* 97, no. 7: 73. https://doi.org/10.1007/s00190-023-01763-9.

Giuntoli, I., J.-P. Vidal, C. Prudhomme, and D. M. Hannah. 2015. "Future Hydrological Extremes: The Uncertainty From Multiple Global Climate and Global Hydrological Models." *Earth System Dynamics* 6, no. 1: 267–285. https://doi.org/10.5194/esd-6-267-2015.

Gladstone, R., R. Flecker, P. Valdes, D. Lunt, and P. Markwick. 2007. "The Mediterranean Hydrologic Budget From a Late Miocene Global Climate Simulation." *Palaeogeography, Palaeoclimatology, Palaeoecology* 251, no. 2: 254–267. https://doi.org/10.1016/j.palaeo.2007.03.050.

Gleeson, T., N. Moosdorf, J. Hartmann, and L. P. H. van Beek. 2014. "A Glimpse Beneath Earth's Surface: GLobal HYdrogeology MaPS

(GLHYMPS) of Permeability and Porosity." *Geophysical Research Letters* 41, no. 11: 3891–3898. https://doi.org/10.1002/2014GL059856.

Gleeson, T., T. Wagener, P. Döll, et al. 2021. "GMD Perspective: The Quest to Improve the Evaluation of Groundwater Representation in Continental- to Global-Scale Models." *Geoscientific Model Development* 14, no. 12: 7545–7571. https://doi.org/10.5194/gmd-14-7545-2021.

Gnann, S., R. Reinecke, L. Stein, et al. 2023. "Functional Relationships Reveal Differences in the Water Cycle Representation of Global Water Models." *Nature Water* 1: 1079–1090. https://doi.org/10.31223/X50S9R.

Gnann, S. J., G. Coxon, R. A. Woods, N. J. Howden, and H. K. McMillan. 2021. "TOSSH: A Toolbox for Streamflow Signatures in Hydrology." *Environmental Modelling & Software* 138: 104983. https://doi.org/10.1016/j.envsoft.2021.104983.

Grubert, E., E. Rogers, and K. T. Sanders. 2020. "Consistent Terminology and Reporting Are Needed to Describe Water Quantity Use." *Journal of Water Resources Planning and Management* 146, no. 8: 04020064. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001241.

Gudmundsson, L., L. M. Tallaksen, K. Stahl, et al. 2012. "Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe." *Journal of Hydrometeorology* 13, no. 2: 604–620. https://doi.org/10.1175/JHM-D-11-083.1.

Gudmundsson, L., T. Wagener, L. M. Tallaksen, and K. Engeland. 2012. "Evaluation of Nine Large-Scale Hydrological Models With Respect to the Seasonal Runoff Climatology in Europe." *Water Resources Research* 48, no. 11: W11504. https://doi.org/10.1029/2011WR010911.

Güntner, A., E. Sharifi, J. Haas, et al. 2024. Global Gravity-Based Groundwater Product (G3P). https://doi.org/10.5880/G3P.2024.001.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez. 2009. "Decomposition of the Mean Squared Error and NSE Performance Criteria: Implications for Improving Hydrological Modelling." *Journal of Hydrology* 377, no. 1–2: 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Gupta, H. V., T. Wagener, and Y. Liu. 2008. "Reconciling Theory With Observations: Elements of a Diagnostic Approach to Model Evaluation." *Hydrological Processes* 22, no. 18: 3802–3813. https://doi.org/10.1002/hyp.6989.

Haddeland, I., D. B. Clark, W. Franssen, et al. 2011. "Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results." *Journal of Hydrometeorology* 12, no. 5: 869–884. https://doi.org/10.1175/2011JHM1324.1.

Hanasaki, N., S. Kanae, and T. Oki. 2006. "A Reservoir Operation Scheme for Global River Routing Models." *Journal of Hydrology* 327, no. 1–2: 22–41. https://doi.org/10.1016/j.jhydrol.2005.11.011.

Harrigan, S., E. Zsoter, L. Alfieri, et al. 2020. "GloFAS-ERA5 Operational Global River Discharge Reanalysis 1979–Present." *Earth System Science Data* 12, no. 3: 2043–2060. https://doi.org/10.5194/essd-12-2043-2020.

Hartmann, A., T. Gleeson, R. Rosolem, F. Pianosi, Y. Wada, and T. Wagener. 2015. "A Large-Scale Simulation Model to Assess Karstic Groundwater Recharge Over Europe and the Mediterranean." *Geoscientific Model Development* 8, no. 6: 1729–1746. https://doi.org/10.5194/gmd-8-1729-2015.

Hartmann, A., T. Gleeson, Y. Wada, and T. Wagener. 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, no. 11: 2842–2847. https://doi.org/10.1073/pnas.1614941114.

Hartmann, A., Y. Liu, T. Olarinoye, R. Berthelin, and V. Marx. 2021. "Integrating Field Work and Large-Scale Modeling to Improve Assessment of Karst Water Resources." *Hydrogeology Journal* 29, no. 1: 315–329. https://doi.org/10.1007/s10040-020-02258-z.

Hasan, H. M. M., P. Döll, S.-M. Hosseini-Moghari, F. Papa, and A. Güntner. 2025. "The Benefits and Trade-Offs of Multi-Variable Calibration of the WaterGAP Global Hydrological Model (WGHM) in the Ganges and Brahmaputra Basins." *Hydrology and Earth System Sciences* 29, no. 2: 567–596. https://doi.org/10.5194/hess-29-567-2025.

Hatfield, J. L., and C. Dold. 2019. "Water-Use Efficiency: Advances and Challenges in a Changing Climate." *Frontiers in Plant Science* 10: 103. https://doi.org/10.3389/fpls.2019.00103.

He, C., Z. Liu, J. Wu, et al. 2021. "Future Global Urban Water Scarcity and Potential Solutions." *Nature Communications* 12, no. 1: 4667. https://doi.org/10.1038/s41467-021-25026-3.

He, X., M. Pan, Z. Wei, E. F. Wood, and J. Sheffield. 2020. "A Global Drought and Flood Catalogue From 1950 to 2016." *Bulletin of the American Meteorological Society* 101, no. 5: E508–E535. https://doi.org/10.1175/BAMS-D-18-0269.1.

Heinicke, S., J. Volkholz, J. Schewe, et al. 2024. "Global Hydrological Models Continue to Overestimate River Discharge." *Environmental Research Letters* 19, no. 7: 74005. https://doi.org/10.1088/1748-9326/ad52b0.

Hoch, J. M., E. H. Sutanudjaja, N. Wanders, R. L. P. H. van Beek, and M. F. P. Bierkens. 2023. "Hyper-Resolution PCR-GLOBWB: Opportunities and Challenges From Refining Model Spatial Resolution to 1 Km Over the European Continent." *Hydrology and Earth System Sciences* 27, no. 6: 1383–1401. https://doi.org/10.5194/hess-27-1383-2023.

Hoch, J. M., and M. A. Trigg. 2019. "Advancing Global Flood Hazard Simulations by Improving Comparability, Benchmarking, and Integration of Global Flood Models." *Environmental Research Letters* 14, no. 3: 34001. https://doi.org/10.1088/1748-9326/aaf3d3.

Hutton, C., T. Wagener, J. Freer, D. Han, C. Duffy, and B. Arheimer. 2016. "Most Computational Hydrology Is Not Reproducible, So Is It Really Science?" *Water Resources Research* 52, no. 10: 7548–7555. https://doi.org/10.1002/2016WR019285.

IPCC. 2023. *Climate Change 2022—Impacts, Adaptation and Vulnerability*. Cambridge University Press.

Kiang, J. E., C. Gazoorian, H. McMillan, et al. 2018. "A Comparison of Methods for Streamflow Uncertainty Estimation." *Water Resources Research* 54, no. 10: 7149–7176. https://doi.org/10.1029/2018WR022708.

Kidd, C., A. Becker, G. J. Huffman, et al. 2017. "So, How Much of the Earth's Surface Is Covered by Rain Gauges?" *Bulletin of the American Meteorological Society* 98, no. 1: 69–78. https://doi.org/10.1175/BAMS-D-14-00283.1.

Kiraz, M., G. Coxon, and T. Wagener. 2023. "A Priori Selection of Hydrological Model Structures in Modular Modelling Frameworks: Application to Great Britain." *Hydrological Sciences Journal* 68, no. 14: 2042–2056. https://doi.org/10.1080/02626667.2023.2251968.

Knoben, W. J. M. 2024. "Setting Expectations for Hydrologic Model Performance With an Ensemble of Simple Benchmarks." *Hydrological Processes* 38, no. 10: e15288. https://doi.org/10.1002/hyp.15288.

Knoben, W. J. M., J. E. Freer, M. C. Peel, K. J. A. Fowler, and R. A. Woods. 2020a. "A Brief Analysis of Conceptual Model Structure Uncertainty Using 36 Models and 559 Catchments." *Water Resources Research* 56: e2019WR025975. https://doi.org/10.1029/2019WR025975.

Knoben, W. J. M., J. E. Freer, and R. A. Woods. 2019. "Technical Note: Inherent Benchmark or Not? Comparing Nash–Sutcliffe and Kling–Gupta Efficiency Scores." *Hydrology and Earth System Sciences* 23, no. 10: 4323–4331. https://doi.org/10.5194/hess-23-4323-2019.

Koster, R. D., P. A. Dirmeyer, Z. Guo, et al. 2004. "Regions of Strong Coupling Between Soil Moisture and Precipitation." *Science (New York, N.Y.)* 305, no. 5687: 1138–1140. https://doi.org/10.1126/science.1100217.

Koster, R. D., and S. P. P. Mahanama. 2012. "Land Surface Controls on Hydroclimatic Means and Variability." *Journal of Hydrometeorology* 13, no. 5: 1604–1620. https://doi.org/10.1175/JHM-D-12-050.1.

Krabbenhoft, C. A., G. H. Allen, P. Lin, et al. 2022. "Assessing Placement Bias of the Global River Gauge Network." *Nature Sustainability* 5: 586–592. https://doi.org/10.1038/s41893-022-00873-0.

Kraft, B., M. Jung, M. Körner, S. Koirala, and M. Reichstein. 2022. "Towards Hybrid Modeling of the Global Hydrological Cycle." *Hydrology and Earth System Sciences* 26, no. 6: 1579–1614. https://doi.org/10.5194/hess-26-1579-2022.

Kragh, S. J., R. Schneider, R. Fensholt, S. Stisen, and J. Koch. 2025. "Synthesizing Regional Irrigation Data Using Machine Learning—Towards Global Upscaling via Metamodeling." *Agricultural Water Management* 311: 109404. https://doi.org/10.1016/j.agwat.2025.109404.

Krysanova, V., J. Zaherpour, I. Didovets, et al. 2020. "How Evaluation of Global Hydrological Models Can Help to Improve Credibility of River Discharge Projections Under Climate Change." *Climatic Change* 163, no. 3: 1353–1377. https://doi.org/10.1007/s10584-020-02840-0.

Kuentz, A., B. Arheimer, Y. Hundecha, and T. Wagener. 2017. "Understanding Hydrologic Variability Across Europe Through Catchment Classification." *Hydrology and Earth System Sciences* 21, no. 6: 2863–2879. https://doi.org/10.5194/hess-21-2863-2017.

Kumar, A., S. N. Gosling, M. F. Johnson, et al. 2022. "Multi-Model Evaluation of Catchment- and Global-Scale Hydrological Model Simulations of Drought Characteristics Across Eight Large River Catchments." *Advances in Water Resources* 165: 104212. https://doi.org/10.1016/j.advwatres.2022.104212.

Kumar, D., and A. R. Ganguly. 2018. "Intercomparison of Model Response and Internal Variability Across Climate Model Ensembles." *Climate Dynamics* 51, no. 1–2: 207–219. https://doi.org/10.1007/s00382-017-3914-4.

Kupzig, J., R. Reinecke, F. Pianosi, M. Flörke, and T. Wagener. 2023. "Towards Parameter Estimation in Global Hydrological Models." *Environmental Research Letters* 18, no. 7: 74023. https://doi.org/10.1088/1748-9326/acdae8.

Landerer, F. W., F. M. Flechtner, H. Save, et al. 2020. "Extending the Global Mass Change Data Record: GRACE Follow-On Instrument and Science Data Performance." *Geophysical Research Letters* 47: e2020GL088306. https://doi.org/10.1029/2020GL088306.

Lee, H., M. Jung, N. Carvalhais, et al. 2023. "Diagnosing Modeling Errors in Global Terrestrial Water Storage Interannual Variability." *Hydrology and Earth System Sciences* 27, no. 7: 1531–1563. https://doi.org/10.5194/hess-27-1531-2023.

Lehner, B., and G. Grill. 2013. "Global River Hydrography and Network Routing: Baseline Data and New Approaches to Study the World's Large River Systems." *Hydrological Processes* 27, no. 15: 2171–2186. https://doi.org/10.1002/hyp.9740.

Lehner, F., C. Deser, N. Maher, et al. 2020. "Partitioning Climate Projection Uncertainty With Multiple Large Ensembles and CMIP5/6." *Earth System Dynamics* 11, no. 2: 491–508. https://doi.org/10.5194/esd-11-491-2020.

Lemordant, L., P. Gentine, A. S. Swann, B. I. Cook, and J. Scheff. 2018. "Critical Impact of Vegetation Physiology on the Continental Hydrologic Cycle in Response to Increasing CO2." *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 16: 4093–4098. https://doi.org/10.1073/pnas.1720712115.

Lewis, H. W., and S. J. Dadson. 2021. "A Regional Coupled Approach to Water Cycle Prediction During Winter 2013/14 in the United Kingdom." *Hydrological Processes* 35, no. 12: e14438. https://doi.org/10.1002/hyp.14438.

Li, F., J. Xiao, J. Chen, et al. 2023. "Global Water Use Efficiency Saturation due to Increased Vapor Pressure Deficit." *Science (New York, N.Y.)* 381, no. 6658: 672–677. https://doi.org/10.1126/science.adf5041.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges. 1994. "A Simple Hydrologically Based Model of Land Surface Water and Energy Fluxes for General Circulation Models." *Journal of Geophysical Research: Atmospheres* 99, no. D7: 14415–14428. https://doi.org/10.1029/94JD00483.

Lodge, J. W., A. P. Dansie, and F. Johnson. 2023. "A Review of Globally Available Data Sources for Modelling the Water-Energy-Food Nexus." *Earth-Science Reviews* 243: 104485. https://doi.org/10.1016/j.earscirev.2023.104485.

Luo, Y. Q., J. T. Randerson, G. Abramowitz, et al. 2012. "A Framework for Benchmarking Land Models." *Biogeosciences* 9, no. 10: 3857–3874. https://doi.org/10.5194/bg-9-3857-2012.

Mahnken, M., M. Cailleret, A. Collalti, et al. 2022. "Accuracy, Realism and General Applicability of European Forest Models." *Global Change Biology* 28, no. 23: 6921–6943. https://doi.org/10.1111/gcb.16384.

Mankin, J. S., R. Seager, J. E. Smerdon, B. I. Cook, and A. P. Williams. 2019. "Mid-Latitude Freshwater Availability Reduced by Projected Vegetation Responses to Climate Change." *Nature Geoscience* 12, no. 12: 983–988. https://doi.org/10.1038/s41561-019-0480-x.

Maraun, D., T. G. Shepherd, M. Widmann, et al. 2017. "Towards Process-Informed Bias Correction of Climate Change Simulations." *Nature Climate Change* 7, no. 11: 764–773. https://doi.org/10.1038/nclimate3418.

McCabe, M. F., M. Rodell, D. E. Alsdorf, et al. 2017. "The Future of Earth Observation in Hydrology." *Hydrology and Earth System Sciences* 21, no. 7: 3879–3914. https://doi.org/10.5194/hess-21-3879-2017.

McDermid, S., M. Nocco, P. Lawston-Parker, et al. 2023. "Irrigation in the Earth System." *Nature Reviews Earth and Environment* 4, no. 7: 435–453. https://doi.org/10.1038/s43017-023-00438-5.

McMillan, H. 2020. "Linking Hydrologic Signatures to Hydrologic Processes: A Review." *Hydrological Processes* 34, no. 6: 1393–1409. https://doi.org/10.1002/hyp.13632.

McMillan, H., R. Araki, L. Bolotin, et al. 2025. "Global Patterns in Observed Hydrologic Processes." *Nature Water* 3, no. 4: 497–506. https://doi.org/10.1038/s44221-025-00407-w.

McMillan, H., R. Araki, S. Gnann, R. Woods, and T. Wagener. 2023. "How Do Hydrologists Perceive Watersheds? A Survey and Analysis of Perceptual Model Figures for Experimental Watersheds." *Hydrological Processes* 37, no. 3: e14845. https://doi.org/10.1002/hyp.14845.

McMillan, H. K. 2021. "A Review of Hydrologic Signatures and Their Applications." *WIREs Water* 8, no. 1: e1499. https://doi.org/10.1002/wat2.1499.

McMillan, H. K., M. P. Clark, W. B. Bowden, M. Duncan, and R. A. Woods. 2011. "Hydrological Field Data From a Modeller's Perspective: Part 1. Diagnostic Tests for Model Structure." *Hydrological Processes* 25, no. 4: 511–522. https://doi.org/10.1002/hyp.7841.

Melsen, L. A. 2022. "It Takes a Village to Run a Model—The Social Practices of Hydrological Modeling." *Water Resources Research* 58, no. 2: e2021WR030600. https://doi.org/10.1029/2021WR030600.

Melsen, L. A., A. Puy, P. J. J. F. Torfs, and A. Saltelli. 2025. "The Rise of the Nash-Sutcliffe Efficiency in Hydrology." *Hydrological Sciences Journal* 1–12. https://doi.org/10.1080/02626667.2025.2475105.

Mendoza, P. A., M. P. Clark, M. Barlage, et al. 2015. "Are We Unnecessarily Constraining the Agility of Complex Process-Based Models?" *Water Resources Research* 51, no. 1: 716–728. https://doi.org/10.1002/2014WR015820.

Meyal, A. Y., R. Versteeg, E. Alper, et al. 2020. "Automated Cloud Based Long Short-Term Memory Neural Network Based SWE Prediction." *Frontiers in Water* 2: 574917. https://doi.org/10.3389/frwa.2020.574917.

Meyer Oliveira, A., A. S. Fleischmann, and R. Paiva. 2021. "On the Contribution of Remote Sensing-Based Calibration to Model Hydrological and Hydraulic Processes in Tropical Regions." *Journal of Hydrology* 597: 126184. https://doi.org/10.1016/j.jhydrol.2021.126184.

Milly, P. C. D., J. Betancourt, M. Falkenmark, et al. 2008. "Climate Change. Stationarity Is Dead: Whither Water Management?" *Science (New York, N.Y.)* 319, no. 5863: 573–574. https://doi.org/10.1126/science.1151915.

Milly, P. C. D., and K. A. Dunne. 2016. "Potential Evapotranspiration and Continental Drying." *Nature Climate Change* 6, no. 10: 946–949. https://doi.org/10.1038/nclimate3046.

Milly, P. C. D., and K. A. Dunne. 2020. "Colorado River Flow Dwindles as Warming-Driven Loss of Reflective Snow Energizes Evaporation." *Science (New York, N.Y.)* 367, no. 6483: 1252–1255. https://doi.org/10.1126/science.aay9187.

Moeck, C., N. Grech-Cumbo, J. Podgorski, et al. 2020. "A Global-Scale Dataset of Direct Natural Groundwater Recharge Rates: A Review of Variables, Processes and Relationships." *Science of the Total Environment* 717: 137042. https://doi.org/10.1016/j.scitotenv.2020.137042.

Moges, E., Y. Demissie, L. Larsen, and F. Yassin. 2021. "Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis." *Water* 13, no. 1: 28. https://doi.org/10.3390/w13010028.

Müller Schmied, H., D. Cáceres, S. Eisner, et al. 2021. "The Global Water Resources and Use Model WaterGAP v2.2d: Model Description and Evaluation." *Geoscientific Model Development* 14, no. 2: 1037–1079. https://doi.org/10.5194/gmd-14-1037-2021.

Mulligan, M., A. van Soesbergen, and L. Sáenz. 2020. "GOODD, a Global Dataset of More Than 38,000 Georeferenced Dams." *Scientific Data* 7, no. 1: 31. https://doi.org/10.1038/s41597-020-0362-5.

Nachtergaele, F., H. van Velthuizen, L. Verelst, et al. 2010. "The Harmonized World Soil Database." In *Proceedings of the 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia, 1–6 August 2010*, edited by R. J. Gilkes and N. Prakongkep, 34–37. International Union of Soil Sciences.

Nearing, G., D. Cohen, V. Dube, et al. 2024. "Global Prediction of Extreme Floods in Ungauged Watersheds." *Nature* 627, no. 8004: 559–563. https://doi.org/10.1038/s41586-024-07145-1.

Nearing, G. S., F. Kratzert, A. K. Sampson, et al. 2021. "What Role Does Hydrological Science Play in the Age of Machine Learning?" *Water Resources Research* 57: e2020WR028091. https://doi.org/10.1029/2020WR028091.

Nearing, G. S., Y. Tian, H. V. Gupta, M. P. Clark, K. W. Harrison, and S. V. Weijs. 2016. "A Philosophical Basis for Hydrological Uncertainty." *Hydrological Sciences Journal* 61, no. 9: 1666–1678. https://doi.org/10.1080/02626667.2016.1183009.

Nyenah, E., P. Döll, D. S. Katz, and R. Reinecke. 2024. Software Sustainability of Global Impact Models.

O, S., and R. Orth. 2021. "Global Soil Moisture Data Derived Through Machine Learning Trained With In-Situ Measurements." *Scientific Data* 8, no. 1: 170. https://doi.org/10.1038/s41597-021-00964-1.

Or, D. 2020. "The Tyranny of Small Scales—On Representing Soil Processes in Global Land Surface Models." *Water Resources Research* 56, no. 6: 0043–1397. https://doi.org/10.1029/2019WR024846.

Paiva, R. C. D., W. Collischonn, M. P. Bonnet, and L. G. G. de Gonçalves. 2012. "On the Sources of Hydrological Prediction Uncertainty in the Amazon." *Hydrology and Earth System Sciences* 16, no. 9: 3127–3137. https://doi.org/10.5194/hess-16-3127-2012.

Papa, F., and F. Frappart. 2021. "Surface Water Storage in Rivers and Wetlands Derived From Satellite Observations: A Review of Current Advances and Future Opportunities for Hydrological Sciences." *Remote Sensing* 13, no. 20: 4162. https://doi.org/10.3390/rs13204162.

Pechlivanidis, I. G., and B. Arheimer. 2015. "Large-Scale Hydrological Modelling by Using Modified PUB Recommendations: The India-HYPE Case." *Hydrology and Earth System Sciences* 19, no. 11: 4559–4579. https://doi.org/10.5194/hess-19-4559-2015.

Piao, S., P. Friedlingstein, P. Ciais, N. de Noblet-Ducoudré, D. Labat, and S. Zaehle. 2007. "Changes in Climate and Land Use Have a Larger Direct Impact Than Rising CO2 on Global River Runoff Trends." *Proceedings of the National Academy of Sciences of the United States of America* 104, no. 39: 15242–15247. https://doi.org/10.1073/pnas.0707213104.

Pimentel, R., B. Arheimer, L. Crochemore, J. C. M. Andersson, I. G. Pechlivanidis, and D. Gustafsson. 2023. "Which Potential Evapotranspiration Formula to Use in Hydrological Modeling World-Wide?" *Water Resources Research* 59, no. 5: e2022WR033447. https://doi.org/10.1029/2022WR033447.

Pimentel, R., L. Crochemore, J. C. M. Andersson, and B. Arheimer. 2023. "Assessing Robustness in Global Hydrological Predictions by Comparing Modelling and Earth Observations." *Hydrological Sciences Journal* 68, no. 16: 2357–2372. https://doi.org/10.1080/02626667.2023.2267544.

Poggio, L., L. M. de Sousa, N. H. Batjes, et al. 2021. "SoilGrids 2.0: Producing Soil Information for the Globe With Quantified Spatial Uncertainty." *Soil* 7, no. 1: 217–240. https://doi.org/10.5194/soil-7-217-2021.

Pokhrel, Y., F. Felfelani, Y. Satoh, et al. 2021. "Global Terrestrial Water Storage and Drought Severity Under Climate Change." *Nature Climate Change* 11, no. 3: 226–233. https://doi.org/10.1038/s41558-020-00972-w.

Pokhrel, Y. N., N. Hanasaki, Y. Wada, and H. Kim. 2016. "Recent Progresses in Incorporating Human Land–Water Management Into Global Land Surface Models Toward Their Integration Into Earth System Models." *WIREs Water* 3, no. 4: 548–574. https://doi.org/10.1002/wat2.1150.

Prudhomme, C., I. Giuntoli, E. L. Robinson, et al. 2014. "Hydrological Droughts in the 21st Century, Hotspots and Uncertainties From a Global Multimodel Ensemble Experiment." *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 9: 3262–3267. https://doi.org/10.1073/pnas.1222473110.

Puy, A., E. Borgonovo, S. Lo Piano, S. A. Levin, and A. Saltelli. 2021. "Irrigated Areas Drive Irrigation Water Withdrawals." *Nature Communications* 12, no. 1: 4525. https://doi.org/10.1038/s41467-021-24508-8.

Puy, A., R. Sheikholeslami, H. V. Gupta, et al. 2022. "The Delusive Accuracy of Global Irrigation Water Withdrawal Estimates." *Nature Communications* 13, no. 1: 3183. https://doi.org/10.1038/s41467-022-30731-8.

Rahman, M., and R. Rosolem. 2017. "Towards a Simple Representation of Chalk Hydrology in Land Surface Modelling." *Hydrology and Earth System Sciences* 21, no. 1: 459–471. https://doi.org/10.5194/hess-21-459-2017.

Randerson, J. T., F. M. Hoffman, P. E. Thornton, et al. 2009. "Systematic Assessment of Terrestrial Biogeochemistry in Coupled Climate–Carbon Models." *Global Change Biology* 15, no. 10: 2462–2484. https://doi.org/10.1111/j.1365-2486.2009.01912.x.

Razavi, S., A. Duffy, L. Eamen, et al. 2025. "Convergent and Transdisciplinary Integration: On the Future of Integrated Modeling of Human-Water Systems." *Water Resources Research* 61: e2024WR038088. https://doi.org/10.1029/2024WR038088.

Reichstein, M., G. Camps-Valls, B. Stevens, et al. 2019. "Deep Learning and Process Understanding for Data-Driven Earth System Science." *Nature* 566, no. 7743: 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Reinecke, R., L. Foglia, S. Mehl, T. Trautmann, D. Cáceres, and P. Döll. 2019. "Challenges in Developing a Global Gradient-Based Groundwater Model (G3M v1.0) for the Integration Into a Global Hydrological Model." *Geoscientific Model Development* 12, no. 6: 2401–2418. https://doi.org/10.5194/gmd-12-2401-2019.

Reinecke, R., S. Gnann, L. Stein, et al. 2024. "Uncertainty in Model Estimates of Global Groundwater Depth." *Environmental Research Letters* 19, no. 11: 114066. https://doi.org/10.1088/1748-9326/ad8587.

Reinecke, R., H. Müller Schmied, T. Trautmann, et al. 2021. "Uncertainty of Simulated Groundwater Recharge at Different Global Warming Levels: A Global-Scale Multi-Model Ensemble Study." *Hydrology and Earth System Sciences* 25, no. 2: 787–810. https://doi.org/10.5194/hess-25-787-2021.

Reinecke, R., T. Trautmann, T. Wagener, and K. Schüler. 2022. "The Critical Need to Foster Computational Reproducibility." *Environmental Research Letters* 17, no. 4: 41005. https://doi.org/10.1088/1748-9326/ac5cf8.

Reinecke, R., A. Wachholz, S. Mehl, L. Foglia, C. Niemann, and P. Döll. 2020. "Importance of Spatial Resolution in Global Groundwater Modeling." *Ground Water* 58, no. 3: 363–376. https://doi.org/10.1111/gwat.12996.

Rockström, J., J. Gupta, D. Qin, et al. 2023. "Safe and Just Earth System Boundaries." *Nature* 619, no. 7968: 102–111. https://doi.org/10.1038/s41586-023-06083-8.

Rodell, M., and J. T. Reager. 2023. "Water Cycle Science Enabled by the GRACE and GRACE-FO Satellite Missions." *Nature Water* 1, no. 1: 47–59. https://doi.org/10.1038/s44221-022-00005-0.

Rosenzweig, C., J. Elliott, D. Deryng, et al. 2014. "Assessing Agricultural Risks of Climate Change in the 21st Century in a Global Gridded Crop Model Intercomparison." *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 9: 3268–3273. https://doi.org/10.1073/pnas.1222463110.

Salwey, S., G. Coxon, F. Pianosi, M. B. Singer, and C. Hutton. 2023. "National-Scale Detection of Reservoir Impacts Through Hydrological Signatures." *Water Resources Research* 59, no. 5: e2022WR033893. https://doi.org/10.1029/2022WR033893.

Samaniego, L., R. Kumar, L. Breuer, et al. 2017. "Propagation of Forcing and Model Uncertainties on to Hydrological Drought Characteristics in a Multi-Model Century-Long Experiment in Large River Basins." *Climatic Change* 141, no. 3: 435–449. https://doi.org/10.1007/s10584-016-1778-y.

Santos, L., J. C. M. Andersson, and B. Arheimer. 2022. "Evaluation of Parameter Sensitivity of a Rainfall-Runoff Model Over a Global Catchment Set." *Hydrological Sciences Journal* 67, no. 3: 342–357. https://doi.org/10.1080/02626667.2022.2035388.

Sarrazin, F., A. Hartmann, F. Pianosi, R. Rosolem, and T. Wagener. 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, no. 12: 4933–4964. https://doi.org/10.5194/gmd-11-4933-2018.

Scanlon, B. R., Z. Zhang, H. Save, et al. 2018. "Global Models Underestimate Large Decadal Declining and Rising Water Storage Trends Relative to GRACE Satellite Data." *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 6: E1080–E1089. https://doi.org/10.1073/pnas.1704665115.

Schaefli, B., and H. V. Gupta. 2007. "Do Nash Values Have Value?" *Hydrological Processes* 21, no. 15: 2075–2080. https://doi.org/10.1002/hyp.6825.

Schellekens, J., E. Dutra, A. La Martínez-de Torre, et al. 2017. "A Global Water Resources Ensemble of Hydrological Models: The eartH2Observe Tier-1 Dataset." *Earth System Science Data* 9, no. 2: 389–413. https://doi.org/10.5194/essd-9-389-2017.

Schewe, J., S. N. Gosling, C. Reyer, et al. 2019. "State-Of-The-Art Global Models Underestimate Impacts From Climate Extremes." *Nature Communications* 10, no. 1: 1005. https://doi.org/10.1038/s41467-019-08745-6.

Schewe, J., J. Heinke, D. Gerten, et al. 2014. "Multimodel Assessment of Water Scarcity Under Climate Change." *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 9: 3245–3250. https://doi.org/10.1073/pnas.1222460110.

Schwarzwald, K., and N. Lenssen. 2022. "The Importance of Internal Climate Variability in Climate Impact Projections." *Proceedings of the National Academy of Sciences of the United States of America* 119, no. 42: e2208095119. https://doi.org/10.1073/pnas.2208095119.

Seibert, J., and S. Bergström. 2022. "A Retrospective on Hydrological Catchment Modelling Based on Half a Century With the HBV Model." *Hydrology and Earth System Sciences* 26, no. 5: 1371–1388. https://doi.org/10.5194/hess-26-1371-2022.

Seibert, J., M. J. P. Vis, E. Lewis, and H. J. van Meerveld. 2018. "Upper and Lower Benchmarks in Hydrological Modelling." *Hydrological Processes* 32, no. 8: 1120–1125. https://doi.org/10.1002/hyp.11476.

Sellar, A. A., C. G. Jones, J. P. Mulcahy, et al. 2019. "UKESM1: Description and Evaluation of the U.K. Earth System Model." *Journal of Advances in Modeling Earth Systems* 11, no. 12: 4513–4558. https://doi.org/10.1029/2019MS001739.

Sellers, P. J., R. E. Dickinson, D. A. Randall, et al. 1997. "Modeling the Exchanges of Energy, Water, and Carbon Between Continents and the Atmosphere." *Science (New York, N.Y.)* 275, no. 5299: 502–509. https://doi.org/10.1126/science.275.5299.502.

Seo, K.-W., D. Ryu, T. Jeon, et al. 2025. "Abrupt Sea Level Rise and Earth's Gradual Pole Shift Reveal Permanent Hydrological Regime Changes in the 21st Century." *Science (New York, N.Y.)* 387, no. 6741: 1408–1413. https://doi.org/10.1126/science.adq6529.

Shen, C., A. P. Appling, P. Gentine, et al. 2023. "Differentiable Modelling to Unify Machine Learning and Physical Models for Geosciences." *Nature Reviews Earth and Environment* 4, no. 8: 552–567. https://doi.org/10.1038/s43017-023-00450-9.

Shen, C., X. Chen, and E. Laloy. 2021. "Editorial: Broadening the Use of Machine Learning in Hydrology." *Frontiers in Water* 3. https://doi.org/10.3389/frwa.2021.681023.

Shrestha, P. K., L. Samaniego, O. Rakovec, et al. 2024a. "Toward Improved Simulations of Disruptive Reservoirs in Global Hydrological Modeling." *Water Resources Research* 60: e2023WR035433. https://doi.org/10.1029/2023WR035433.

Shukla, S., J. Sheffield, E. F. Wood, and D. P. Lettenmaier. 2013. "On the Sources of Global Land Surface Hydrologic Predictability." *Hydrology and Earth System Sciences* 17, no. 7: 2781–2796. https://doi.org/10.5194/hess-17-2781-2013.

Shuttleworth, W. J. 1994. "Large-Scale Experimental and Modelling Studies of Hydrological Processes." *Ambio* 23, no. 1: 82–86.

Siebert, S., M. Kummu, M. Porkka, P. Döll, N. Ramankutty, and B. R. Scanlon. 2015. "A Global Data Set of the Extent of Irrigated Land From 1900 to 2005." *Hydrology and Earth System Sciences* 19, no. 3: 1521–1545. https://doi.org/10.5194/hess-19-1521-2015.

Singh, A., S. Kumar, S. Akula, D. M. Lawrence, and D. L. Lombardozzi. 2020. "Plant Growth Nullifies the Effect of Increased Water-Use Efficiency on Streamflow Under Elevated CO_2 in the Southeastern United States." *Geophysical Research Letters* 47: e2019GL086940. https://doi.org/10.1029/2019GL086940.

Sinha, T., and A. Sankarasubramanian. 2013. "Role of Climate Forecasts and Initial Conditions in Developing Streamflow and Soil Moisture Forecasts in a Rainfall–Runoff Regime." *Hydrology and Earth System Sciences* 17, no. 2: 721–733. https://doi.org/10.5194/hess-17-721-2013.

Sivapalan, M. 2018. "From Engineering Hydrology to Earth System Science: Milestones in the Transformation of Hydrologic Science." *Hydrology and Earth System Sciences* 22, no. 3: 1665–1693. https://doi.org/10.5194/hess-22-1665-2018.

Slater, L. J., L. Arnal, M.-A. Boucher, et al. 2023. "Hybrid Forecasting: Blending Climate Predictions With AI Models." *Hydrology and*

Earth System Sciences 27, no. 9: 1865–1889. https://doi.org/10.5194/hess-27-1865-2023.

Smith, K. A., R. L. Wilby, C. Broderick, et al. 2018. "Navigating Cascades of Uncertainty—As Easy as ABC? Not Quite…." *Journal of Extreme Events* 5, no. 1: 1850007. https://doi.org/10.1142/S2345737618500070.

Stein, L., S. K. Mukkavilli, B. M. Pfitzmann, et al. 2024. "Wealth Over Woe: Global Biases in Hydro-Hazard Research." *Earth's Future* 12: e2024EF004590. https://doi.org/10.1029/2024EF004590.

Stein, L., S. K. Mukkavilli, and T. Wagener. 2022. "Lifelines for a Drowning Science - Improving Findability and Synthesis of Hydrologic Publications." *Hydrological Processes* 36, no. 11: e14742. https://doi.org/10.1002/hyp.14742.

Sterling, S. M., A. Ducharne, and J. Polcher. 2013. "The Impact of Global Land-Cover Change on the Terrestrial Water Cycle." *Nature Climate Change* 3, no. 4: 385–390. https://doi.org/10.1038/nclimate1690.

Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K.-L. Hsu. 2018. "A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons." *Reviews of Geophysics* 56, no. 1: 79–107. https://doi.org/10.1002/2017RG000574.

Tabari, H., P. Hosseinzadehtalaei, W. Thiery, and P. Willems. 2021. "Amplified Drought and Flood Risk Under Future Socioeconomic and Climatic Change." *Earth's Future* 9: e2021EF002295. https://doi.org/10.1029/2021EF002295.

Tarasova, L., S. Gnann, S. Yang, A. Hartmann, and T. Wagener. 2024. "Catchment Characterization: Current Descriptors, Knowledge Gaps and Future Opportunities." *Earth-Science Reviews* 252: 104739. https://doi.org/10.1016/j.earscirev.2024.104739.

Tarek, M., F. Brissette, and R. Arsenault. 2021. "Uncertainty of Gridded Precipitation and Temperature Reference Datasets in Climate Change Impact Studies." *Hydrology and Earth System Sciences* 25, no. 6: 3331–3350. https://doi.org/10.5194/hess-25-3331-2021.

Telteu, C.-E., H. Müller Schmied, W. Thiery, et al. 2021. "Understanding Each Other's Models: An Introduction and a Standard Representation of 16 Global Water Models to Support Intercomparison, Improvement, and Communication." *Geoscientific Model Development* 14, no. 6: 3843–3878. https://doi.org/10.5194/gmd-14-3843-2021.

Tian, S., L. J. Renzullo, A. I. J. M. van Dijk, P. Tregoning, and J. P. Walker. 2019. "Global Joint Assimilation of GRACE and SMOS for Improved Estimation of Root-Zone Soil Moisture and Vegetation Response." *Hydrology and Earth System Sciences* 23, no. 2: 1067–1081. https://doi.org/10.5194/hess-23-1067-2019.

Tiwari, A. D., Y. Pokhrel, J. Boulange, P. Burek, L. Guillaumot, and S. N. Gosling. 2025. Similarities and Divergent Patterns in Hydrologic Fluxes and Storages Simulated by Global Water Models.

Troy, T. J., E. F. Wood, and J. Sheffield. 2008. "An Efficient Calibration Method for Continental-Scale Land Surface Modeling." *Water Resources Research* 44: W09411. https://doi.org/10.1029/2007WR006513.

Tsai, W.-P., D. Feng, M. Pan, et al. 2021. "From Calibration to Parameter Learning: Harnessing the Scaling Effects of Big Data in Geoscientific Modeling." *Nature Communications* 12, no. 1: 5988. https://doi.org/10.1038/s41467-021-26107-z.

van Vliet, M. T. H., E. R. Jones, M. Flörke, et al. 2021. "Global Water Scarcity Including Surface Water Quality and Expansions of Clean Water Technologies." *Environmental Research Letters* 16, no. 2: 24020. https://doi.org/10.1088/1748-9326/abbfc3.

van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang. 2008. "Rainfall Characteristics Define the Value of Streamflow Observations for Distributed Watershed Model Identification." *Geophysical Research Letters* 35: L11403. https://doi.org/10.1029/2008GL034162.

van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang. 2009. "Sensitivity-Guided Reduction of Parametric Dimensionality for Multi-Objective Calibration of Watershed Models." *Advances in Water Resources* 32, no. 8: 1154–1169. https://doi.org/10.1016/j.advwatres.2009.03.002.

Veldkamp, T. I. E., Y. Wada, J. C. J. H. Aerts, and P. J. Ward. 2016. "Towards a Global Water Scarcity Risk Assessment Framework: Incorporation of Probability Distributions and Hydro-Climatic Variability." *Environmental Research Letters* 11, no. 2: 24006. https://doi.org/10.1088/1748-9326/11/2/024006.

Vereecken, H., A. Schnepf, J. W. Hopmans, et al. 2016. "Modeling Soil Processes: Review, Key Challenges, and New Perspectives." *Vadose Zone Journal* 15: 1–57. https://doi.org/10.2136/vzj2015.09.0131.

Vilmin, L., J. M. Mogollón, A. H. W. Beusen, et al. 2020. "Modeling Process-Based Biogeochemical Dynamics in Surface Fresh Waters of Large Watersheds With the IMAGE-DGNM Framework." *Journal of Advances in Modeling Earth Systems* 12: e2019MS001796. https://doi.org/10.1029/2019MS001796.

Viviroli, D., D. R. Archer, W. Buytaert, et al. 2011. "Climate Change and Mountain Water Resources: Overview and Recommendations for Research, Management and Policy." *Hydrology and Earth System Sciences* 15, no. 2: 471–504. https://doi.org/10.5194/hess-15-471-2011.

Wada, Y., M. F. P. Bierkens, A. de Roo, et al. 2017. "Human–Water Interface in Hydrological Modelling: Current Status and Future Directions." *Hydrology and Earth System Sciences* 21, no. 8: 4169–4193. https://doi.org/10.5194/hess-21-4169-2017.

Wada, Y., D. Wisser, S. Eisner, et al. 2013. "Multimodel Projections and Uncertainties of Irrigation Water Demand Under Climate Change." *Geophysical Research Letters* 40, no. 17: 4626–4632. https://doi.org/10.1002/grl.50686.

Wagener, T., T. Gleeson, G. Coxon, et al. 2021. "On Doing Hydrology With Dragons: Realizing the Value of Perceptual Models and Knowledge Accumulation." *Wiley Interdisciplinary Reviews: Water* 8, no. 6: e1550. https://doi.org/10.1002/wat2.1550.

Wagener, T., and A. Montanari. 2011. "Convergence of Approaches Toward Reducing Uncertainty in Predictions in Ungauged Basins." *Water Resources Research* 47: W06301. https://doi.org/10.1029/2010WR009469.

Wagener, T., R. Reinecke, and F. Pianosi. 2022. "On the Evaluation of Climate Change Impact Models." *WIREs Climate Change* 13, no. 3: e772. https://doi.org/10.1002/wcc.772.

Wagener, T., M. Sivapalan, P. A. Troch, et al. 2010. "The Future of Hydrology: An Evolving Science for a Changing World." *Water Resources Research* 46: W05301. https://doi.org/10.1029/2009WR008906.

Walker, W. E., P. Harremoës, J. Rotmans, et al. 2003. "Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support." *Integrated Assessment* 4, no. 1: 5–17. https://doi.org/10.1076/iaij.4.1.5.16466.

Wan, W., P. Döll, and H. Müller Schmied. 2024. "Global-Scale Groundwater Recharge Modeling is Improved by Tuning Against Ground-Based Estimates for Karst and Non-Karst Areas." *Water Resources Research* 60: e2023WR036182. https://doi.org/10.1029/2023WR036182.

Ward, P. J., S. Eisner, M. Flörke, M. D. Dettinger, and M. Kummu. 2014. "Annual Flood Sensitivities to El Niño–Southern Oscillation at the Global Scale." *Hydrology and Earth System Sciences* 18, no. 1: 47–66. https://doi.org/10.5194/hess-18-47-2014.

Wartenburger, R., S. I. Seneviratne, M. Hirschi, et al. 2018. "Evapotranspiration Simulations in ISIMIP2a—Evaluation of Spatio-Temporal Characteristics With a Comprehensive Ensemble of Independent Datasets." *Environmental Research Letters* 13, no. 7: 75001. https://doi.org/10.1088/1748-9326/aac4bb.

Weber, T. K. D., L. Weihermüller, A. Nemes, M. Bechtold, A. Degré, and E. Diamantopoulos. 2023. Hydro-Pedotransfer Functions: A Roadmap for Future Development.

West, C., R. Rosolem, A. M. MacDonald, M. O. Cuthbert, and T. Wagener. 2022. "Understanding Process Controls on Groundwater Recharge Variability Across Africa Through Recharge Landscapes." *Journal of Hydrology* 612: 127967. https://doi.org/10.1016/j.jhydrol.2022.127967.

Westerberg, I. K., and H. K. McMillan. 2015. "Uncertainty in Hydrological Signatures." *Hydrology and Earth System Sciences* 19, no. 9: 3951–3968. https://doi.org/10.5194/hess-19-3951-2015.

Wiersma, P., J. Aerts, H. Zekollari, et al. 2022. "Coupling a Global Glacier Model to a Global Hydrological Model Prevents Underestimation of Glacier Runoff." *Hydrology and Earth System Sciences* 26, no. 23: 5971–5986. https://doi.org/10.5194/hess-26-5971-2022.

Willner, S. N., C. Otto, and A. Levermann. 2018. "Global Economic Response to River Floods." *Nature Climate Change* 8, no. 7: 594–598. https://doi.org/10.1038/s41558-018-0173-2.

WMO. 2021. *Guidelines on Seasonal Hydrological Prediction*. WMO.

Wood, A. W., T. Hopson, A. Newman, L. Brekke, J. Arnold, and M. Clark. 2016. "Quantifying Streamflow Forecast Skill Elasticity to Initial Condition and Climate Prediction Skill." *Journal of Hydrometeorology* 17, no. 2: 651–668. https://doi.org/10.1175/JHM-D-14-0213.1.

Wood, E. F., J. K. Roundy, T. J. Troy, et al. 2011. "Hyperresolution Global Land Surface Modeling: Meeting a Grand Challenge for Monitoring Earth's Terrestrial Water." *Water Resources Research* 47, no. 5: 2010WR010090. https://doi.org/10.1029/2010WR010090.

Wu, Y., C. Miao, L. Slater, X. Fan, Y. Chai, and S. Sorooshian. 2024. "Hydrological Projections Under CMIP5 and CMIP6: Sources and Magnitudes of Uncertainty." *Bulletin of the American Meteorological Society* 105, no. 1: E59–E74. https://doi.org/10.1175/BAMS-D-23-0104.1.

Wulder, M. A., D. P. Roy, V. C. Radeloff, et al. 2022. "Fifty Years of Landsat Science and Impacts." *Remote Sensing of Environment* 280: 113195. https://doi.org/10.1016/j.rse.2022.113195.

Wunsch, A., T. Liesch, and S. Broda. 2022. "Deep Learning Shows Declining Groundwater Levels in Germany Until 2100 due to Climate Change." *Nature Communications* 13, no. 1: 1221. https://doi.org/10.1038/s41467-022-28770-2.

Xie, S.-P., C. Deser, G. A. Vecchi, et al. 2015. "Towards Predictive Understanding of Regional Climate Change." *Nature Climate Change* 5, no. 10: 921–930. https://doi.org/10.1038/nclimate2689.

Yamazaki, D., D. Ikeshima, J. Sosa, P. D. Bates, G. H. Allen, and T. M. Pavelsky. 2019. "MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset." *Water Resources Research* 55, no. 6: 5053–5073. https://doi.org/10.1029/2019WR024873.

Zaherpour, J., S. N. Gosling, N. Mount, et al. 2018. "Worldwide Evaluation of Mean and Extreme Runoff From Six Global-Scale Hydrological Models That Account for Human Impacts." *Environmental Research Letters* 13, no. 6: 65015. https://doi.org/10.1088/1748-9326/aac547.

Zaherpour, J., N. Mount, S. N. Gosling, et al. 2019. "Exploring the Value of Machine Learning for Weighted Multi-Model Combination of an Ensemble of Global Hydrological Models." *Environmental Modelling & Software* 114: 112–128. https://doi.org/10.1016/j.envsoft.2019.01.003.

Zhang, A. T., and V. X. Gu. 2023. "Global Dam Tracker: A Database of More Than 35,000 Dams With Location, Catchment, and Attribute Information." *Scientific Data* 10, no. 1: 111. https://doi.org/10.1038/s41597-023-02008-2.

Zhang, C., P. Abbaszadeh, L. Xu, H. Moradkhani, Q. Duan, and W. Gong. 2021. "A combined Optimization-Assimilation Framework to Enhance the Predictive Skill of Community Land Model." *Water Resources Research* 57: e2021WR029879. https://doi.org/10.1029/2021WR029879.

Zhang, H., R. Lauerwald, P. Regnier, et al. 2020. "Simulating Erosion-Induced Soil and Carbon Delivery From Uplands to Rivers in a Global Land Surface Model." *Journal of Advances in Modeling Earth Systems* 12, no. 11: e2020MS002121. https://doi.org/10.1029/2020MS002121.

Zhang, Y., H. Zheng, X. Zhang, et al. 2023. "Future Global Streamflow Declines Are Probably More Severe Than Previously Estimated." *Nature Water* 1, no. 3: 261–271. https://doi.org/10.1038/s44221-023-00030-7.

Zhao, F., N. Nie, Y. Liu, et al. 2025. "Benefits of Calibrating a Global Hydrological Model for Regional Analyses of Flood and Drought Projections: A Case Study of the Yangtze River Basin." *Water Resources Research* 61: e2024WR037153. https://doi.org/10.1029/2024WR037153.

Zipper, S., K. M. Befus, R. Reinecke, et al. 2023. "GroMoPo: A Groundwater Model Portal for Findable, Accessible, Interoperable, and Reusable (FAIR) Modeling." *Ground Water* 61, no. 6: 764–767. https://doi.org/10.1111/gwat.13343.

Zipper, S. C., J. Keune, and S. J. Kollet. 2019. "Land Use Change Impacts on European Heat and Drought: Remote Land-Atmosphere Feedbacks Mitigated Locally by Shallow Groundwater." *Environmental Research Letters* 14, no. 4: 44012. https://doi.org/10.1088/1748-9326/ab0db3.

Zsoter, E., H. Cloke, E. Stephens, et al. 2019. "How Well Do Operational Numerical Weather Prediction Configurations Represent Hydrology?" *Journal of Hydrometeorology* 20, no. 8: 1533–1552. https://doi.org/10.1175/JHM-D-18-0086.1.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.