AMERICAN WATER RESOURCES ASSOCIATION

RESEARCH ARTICLE OPEN ACCESS

The NextGen Water Resources Modeling Framework: Community Innovation at the Intersection of Hydrologic, Data and Computer Sciences

Fred L. Ogden[1] | Keith Jennings[2] | Edward P. Clark[1] | Ethan Coon[3] | Brian Cosgrove[1] | Luciana Kindl da Cunha[4] | Matthew W. Farthing[5] | Trey Flowers[1] | Jonathan M. Frame[6,7] | Nels J. Frazier[6] | Jessica L. Garrett[6] | Thomas M. Graziano[1] | Joseph D. Hughes[8] | J. Michael Johnson[6] | Rachel McDaniel[1] | J. David Moulton[9] | Scott D. Peckham[10] | Fernando R. Salas[1] | Gaurav Savant[5] | Roland Viger[11] | Andy Wood[12,13]

[1]U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Office of Water Prediction, Silver Spring, Maryland, USA | [2]University of Vermont Water Resources Institute (Formerly Lynker Corp), Burlington, Vermont, USA | [3]U.S. Department of Energy, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA | [4]West Consultants Inc, Folsom, California, USA | [5]U.S. Department of Defense, U.S. Army Corps of Engineers, Engineer Research and Development Center, Vicksburg, Mississippi, USA | [6]Lynker, Leesburg, Virginia, USA | [7]Department of Geological Sciences, University of Alabama, Tuscaloosa, Alabama, USA | [8]INTERA Inc (Formerly U.S. Department of the Interior, Geological Survey, Water Mission Area), Austin, Texas, USA | [9]U.S. Department of Energy, Los Alamos National Laboratory, Los Alamos, New Mexico, USA | [10]Institute for Arctic and Alpine Research, University of Colorado, Boulder, Colorado, USA | [11]Retired, U.S. Department of the Interior, Geological Survey, Water Mission Area, Reston, Virginia, USA | [12]Colorado School of Mines, Golden, Colorado, USA | [13]US National Science Foundation, National Center for Atmospheric Research, Boulder, Colorado, USA

Correspondence: Fred L. Ogden (fred.ogden@noaa.gov)

Received: 6 February 2025 | **Revised:** 21 November 2025 | **Accepted:** 27 January 2026

Keywords: community hydrologic modeling | hydrologic modeling framework | model interoperability | Next Generation Water Resources Modeling (NextGen)

ABSTRACT

Hydrologic science lacks a comprehensive theory of stormflow generation, preventing the development of a general hydrologic model. Studies show that models focusing on dominant local processes often outperform general models that rely on parameter tuning, leading to higher confidence solutions. For continental-scale hydrologic and hydraulic prediction, regional mosaics of models may outperform a single-model approach. However, variations in model inputs, programming languages, solvers, and discretizations hinder interoperability and comparisons. To address these challenges, we developed the Next Generation Water Resources Modeling Framework (NextGen): a model-agnostic, standards-based architecture for model interoperability and evaluation. Two standards enable the Framework: (1) the Basic Model Interface (BMI) version 2.0, for model control, coupling, and querying; and (2) the Open Geospatial Consortium WaterML 2.0 part 3 Hydrologic Features (HY_Features) conceptual data model to describe the "hydrofabric" of surface water hydrologic and hydraulic features. In the NextGen Framework, models retain their unique solution methods while becoming interoperable through BMI variable exchange tied to a common hydrofabric. The Framework enables scientific evaluation of water prediction models that simulate diverse hydrologic and hydraulic processes. Its design supports models written in multiple programming languages and runs on laptops, cloud and distributed memory supercomputers.

1 | Introduction

The motivation for the developments reported in this paper comes from the need to provide accurate water resources modeling results for a variety of purposes, including flood forecasting, reservoir operations, and drought prediction. While research in hydrologic and water resources science often involves development of specialized codes to simulate single or coupled processes, these models seldom advance to a "technical readiness level" (NOAA 2025) suitable for operational implementation. Typically, that function falls on national-level institutes, laboratories, or government agencies. Historically this has resulted in creation of a set of model-specific requirements and tools aimed at bolstering the capabilities of individual models or enabling specific coupled model configurations. This path does not address model interoperability and can impede robust testing and comparison with other models.

In the United States (US) and in many other countries, models used to predict and forecast water resources often originate from or are endorsed by one or more federal agencies. In the US, agencies are directed by Congress to develop and operate capabilities that allow them to assess, forecast and manage water resources. Federal agencies produce timely information on the state of water resources and the risks of extreme hydrologic events over time scales ranging from hours to decades, under a variety of future scenarios. This has resulted in a proliferation of different agency- and mission-specific models and modeling approaches. Various federal, state, and local regulatory authorities often require the use of specific federally developed or certified models because of their perceived unbiased development, scientific rigor and robustness.

For example, the US Department of Commerce, National Oceanic and Atmospheric Administration (NOAA), National Weather Service (NWS), Office of Water Prediction (OWP) is transforming its hydrologic service delivery and its provision of impact-based decision support services (IDSS) in its mission to protect lives and property and promote the national economy. A 2012 study by the National Research Council found that the hydrologic forecasting process in place required too much intervention by human forecasters. The report recommended that the NWS hydrology program upgrade forecasting technology to place the forecaster "above the loop" in forecast generation (National Research Council 2012). In 2015, Congress responded to this recommendation by appropriating resources to the OWP to develop and demonstrate an advanced continental scale forecasting capability to operate on NOAA's operational supercomputer.

In response, NOAA/NWS developed and deployed the National Water Model (NWM) to provide stormflow forecasts on over 3.6 million miles of rivers. Versions 3.1 and earlier of the NWM are based on the hydrologic extension to the Weather Research and Forecasting Model Hydrological modeling system (WRF-Hydro), which uses direct data assimilation of USGS streamflow observations to remove the forecaster from the loop (Cosgrove et al. 2024). NOAA-NWS also develops and operates models to predict the occurrence and intensity of drought, compound flooding in coastal regions and along the shores of the US Great Lakes, and subseasonal, seasonal, and annual water supplies. With passage of the Bipartisan Infrastructure Law in 2021, Congress further directed OWP to improve coastal and inland streamflow forecasts through the development of a next generation water resources prediction capability, and to produce and publicly disseminate flood inundation forecasts that serve nearly 100% of the US population by 2026; the public availability of this flood inundation mapping information is a Key Performance Indicator tracked by the US Department of Commerce (US Public Law 117–58 2021 TITLE II—COMMERCE, JUSTICE, SCIENCE, AND RELATED AGENCIES). This highlights the need to accurately predict stormflow nationwide, particularly peak river discharges as they are highly correlated with flood inundation.

Other federal water prediction agencies use hydrologic and hydraulic models to predict water supply and availability, operate reservoirs, and manage water quality and associated risks to water supplies and aquatic habitat. These demands are only growing. For example, Congress has directed the US Army Corps of Engineers (USACE) to consider comprehensive flood risk for the United States. Section 8106(a) of the Water Resources Development Act (WRDA) 2022 requires that when performing feasibility studies for flood risk management, USACE be prepared to account for both isolated and compound risk due to the effects of (among others) riverine discharge; rainfall events; inundation, wave attack, and erosion due to hurricanes or coastal storms; tidally driven flooding in coastal, riverine, or estuarine areas; seasonal variation in water levels; sea level rise; and groundwater emergence (Conor 2024). These and other requirements provide additional impetus for the continued investments into hydrologic and hydraulic models with capabilities to help guide flood risk and coastal storm risk mitigation design decisions.

The US Congress assigned the US Geological Survey (USGS) the responsibility of performing regular and thorough evaluations of water availability in the United States under Subtitle F of the Omnibus Public Land Management Act of 2009, also referred to as the SECURE Water Act. A preliminary report (Evenson et al. 2018) highlighted the necessity for enhancing modeling capabilities to better estimate water availability across the country. Recent advancements in these capabilities were applied in the US Geological Survey's Integrated Water Availability Assessment (USGS 2025). These developments included models for calculating water withdrawals for agricultural irrigation, thermoelectric power generation, and municipal supply (Luukkonen et al. 2025; Medalie et al. 2025), enabling water use estimates with significantly improved temporal and spatial detail compared to past methods. The assessment also employed two national hydrologic model codes: the USGS National Hydrologic Model using the Precipitation Runoff Modeling System (Regan et al. 2018), and WRF-Hydro (Rafieeinasab et al. 2024). In ongoing assessments of water resource availability at various scales—national, regional, and local—the USGS is enhancing its suite of model codes, accompanying data, and workflows to boost the accuracy, reliability, and efficiency of simulations, as well as the reproducibility of scientific methods. Like other Federal agencies, the USGS is embracing open science principles and is actively partnering with collaborators on various aspects, including coding and the geospatial framework used to model the states of landscape features.

The Department of Energy (DOE) develops and applies a diverse suite of hydrologic and hydro-biogeochemical models to support its diverse nonregulatory mission in water and energy security for the nation. Targeting fundamental processes, the Office of Science (SC) focuses on advancing our understanding of system function with respect to water quantity and water quality and leads the effort to utilize DOE's high performance computing facilities with process-driven models. This work underpins and supports the Water Power Technologies Office assessment of global climate change risk on hydroelectric power generation consistent with the SECURE Water Act, and the Office of Environmental (EM) Management in its mission to predict the fate of legacy waste in ground and surface water. For example, DOE-EM leverages new modeling capabilities from the DOE-SC to better understand and explore water quality issues at legacy waste sites while relying on a suite of established operational and legacy codes from the DOE and other agencies for its regulatory work. The diversity of applications and relevant codes in DOE has generated significant activity in software ecosystems and standards-driven approaches such as those described here.

In the US, NOAA has responsibility for streamflow prediction and warning at hourly to seasonal timescales, while responsibilities for management of water resources are shared across multiple federal, state, and local agencies. Each of these agencies has unique research, operational, and design requirements that they strive to meet. These unique needs have led to the creation of extensive data and modeling capabilities, but these capabilities are neither integrated nor employed in a way that allows them to have a national impact. This paper addresses the question: "How can we better develop water prediction models to promote model interoperability in a way that speeds advances and allows scientific evaluation to identify and select the best performing models?"

1.1 | Why Use Hydrologic Models?

There are myriad reasons why one may use a hydrologic model. Perhaps most importantly, models allow us to account for unobserved, uncertain hydrologic processes. While some watersheds can be treated as "simple dynamical systems" (Kirchner 2009), hydrologists seldom know the path taken by water that becomes streamflow in most catchments and how long the water took to traverse the path. To make matters more challenging, the relative importance of different pathways varies spatially and temporally (e.g., Blöschl et al. 2019; McDonnell et al. 2007), while different runoff generation mechanisms often exhibit similar behaviors that prevent diagnosis using streamflow observations alone (Ameli et al. 2015). As a result of these complexities, the literature contains many different hypotheses regarding streamflow response to rainfall and snowmelt, tested using hydrography and sometimes aided by streamflow geochemistry (Sklash and Farvolden 1979; Rice and Hornberger 1998), specific conductance (Kunkle 1965; Pinder and Jones 1969; Pellerin et al. 2008; Lott and Stewart 2013), introduced and naturally occurring tracers such as light stable and radioactive isotopes (Kendall and McDonnell 2012; Cartwright et al. 2017; Cook and Herczeg 2000), geophysical observations (Robinson et al. 2008), and applied rainfall and irrigation experiments (Montgomery et al. 1997; Polyakov et al. 2018).

To address the sizable gaps in our understanding, hydrologists, and water resources engineers complement field studies by employing or developing computational models that represent the full hydrologic balance or components thereof (e.g., Horton et al. 2022; Clark, Hendrikx, et al. 2011; Clark, Kavetski, and Fenicia 2011), relying on various structures and paradigms (Beven and Young 2013). Such computational models come in various forms, including conceptual, physical-conceptual, physically based, and statistical. Developers of such hydrologic models apply an assumed set of valid water flux estimators and integrate calculated water fluxes over time using a variety of discretizations to compute approximations to the mathematical representations of the processes of interest (Clark et al. 2017; Ogden 2021).

Recently, the development and application of data-driven machine learning techniques have added a new dimension to modeling paradigms. A common approach is long short-term memory (LSTM) networks (Kratzert et al. 2018) that are applicable at hourly to daily time steps (Gauch, Kratzert, et al. 2021), can incorporate data assimilation (Nearing et al. 2021), and enforce conservation laws (Frame et al. 2023). More recent advances in AI/ML have involved "differentiable models" consisting of functions that are mathematically differentiable. Differentiability is crucial because it allows practitioners to efficiently use AI techniques to train the model's parameters using gradient-based optimization methods such as backpropagation or gradient descent. Tests of this approach have demonstrated skill in improving the conceptual understanding of deep learning approaches (Wang et al. 2023; Shen et al. 2023; Bindas et al. 2024).

The above information means researchers, water resources managers, and operational forecasters now have their pick of model types, structures, and paradigms. The ability of researchers to develop new model types has unfortunately not been matched by a commensurate increase in understanding of processes and model appropriateness criteria. To this latter point, the hydrologic modeling community continues to struggle with the appropriate choice and interpretation of performance metrics. The lack of consistent standards with regard to metrics and the use of different metrics hampers model intercomparison.

1.2 | Model Selection and Performance

Despite the diversity of model types, assumptions, and suitability for different applications, the literature suggests that researchers do not often select models for valid reasons (Dooge 1986; Kirchner 2006). Addor and Melsen (2019) analyzed published hydrological modeling studies to ascertain why particular models were used in those studies. They found that hydrologists seemed to favor familiarity over appropriateness in selecting a model to meet their study objectives. In other words, researchers preferred to use models they know how to use rather than for their applicability or appropriateness. Addor and Melsen (2019) also detected regional preferences in model use, with certain models being consistently preferred by researchers with current or former affiliation with particular research universities or institutes. The influence of model adequacy on selection was less clear. A major reason modelers tend to rely on familiar models is the time and effort required to learn the details of setting up and

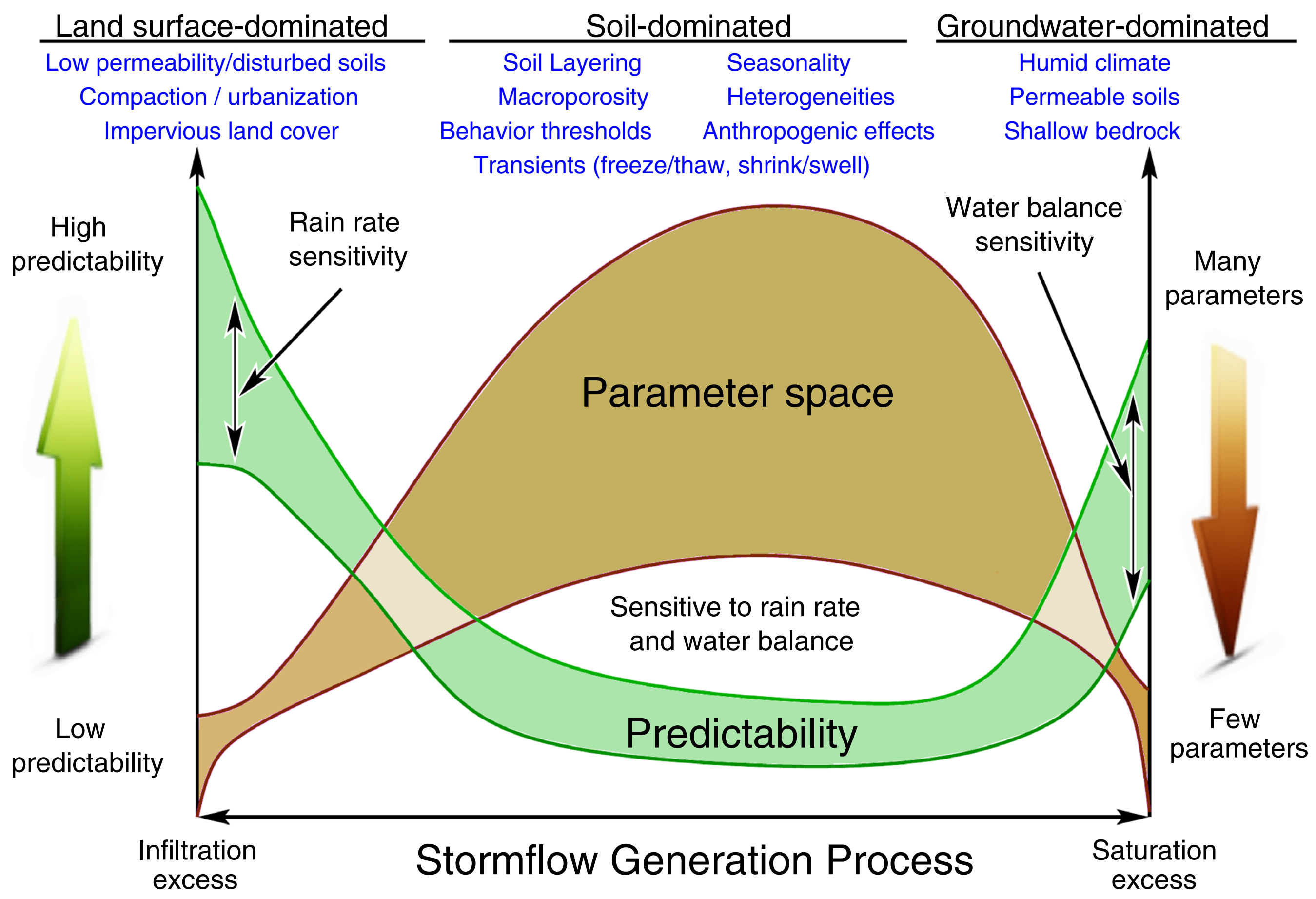

FIGURE 1 | Conceptual model showing the relative size of the parameter space required (brown band) and resulting predictability (green band) for modeling stormflow generation as a function of land surface-, soil-, and groundwater-dominated mechanisms. This conceptual model considers infiltration–excess and saturation–excess storm flow generation mechanisms as highly predictable end members.

running an unfamiliar model. Working with a new model also increases the likelihood of making mistakes. Together, these factors create a strong incentive to "stick with what you know" rather than explore alternative model formulations. Therefore, any approach that lowers the barrier to using multiple models is appealing.

Model selection depends on a number of factors, including the hydrologic processes simulated and the quantity and quality of observed or modeled forcing inputs used for parameter estimation and forecasting skill evaluation. Given the focus of this manuscript on computational models and not observational data, we pay more attention to the former than the latter. In terms of stormflow generation, for example, there exist two highly predictable situations. The first occurs when potential infiltration is small or nearly zero and the majority of precipitation and snowmelt runs off, often referred to as the "infiltration excess" runoff generation mechanism (Horton 1933). The second occurs when all rainfall and snowmelt infiltrates into the soil; in humid climates or situations where the water table is near the land surface, this can result in the "saturation excess" runoff generation mechanism (Dunne and Black 1970). These two processes, saturation-excess and infiltration-excess, might be considered as end members of a complex set of stormflow generation behaviors. Between these two highly predictable end members, infiltration processes are determined by particular properties of the soil or land cover, which is often complicated by hard-to-observe vertical (Cuthbert and Tindimugaya 2010) and lateral (Beven and Germann 1982, 2013; Cheng et al. 2017) preferential flow paths.

Figure 1 illustrates how the stormflow generation process affects both the predictability and the number of parameters required for process-based modeling. The land surface-dominated infiltration–excess and groundwater-dominated saturation–excess mechanisms are highly predictable using a small number of parameters. Between these two end members lie the complexities of soil-dominated runoff generation, where the properties of the soil profile and many other factors such as freeze/thaw, shrink/swell, tillage practices, and soil compaction can dominate. In the case of soil-dominated runoff generation, difficult-to-observe phenomena such as soil macropores and land-use history can play important roles. In this case, hydrologic science lacks a comprehensive theory of stormflow generation. This has led to an over-reliance on infiltration theory and perhaps on topographic wetness index theory as well (Beven 2021), because these end members represent the most mature theoretically based techniques for partitioning rainfall/snowmelt at the land surface and require only a small number of parameters.

At the left-hand limit of the plot shown in Figure 1 lies land-surface-dominated stormflow generation, which occurs in the

FIGURE 2 | Photograph showing complexities of soil-dominated hydrology, looking down a fence line after an intense springtime rain in Iowa. At the time that this photo was taken, the field on the left side of the fence was farmed using standard tillage practices while the field on the right was farmed using minimum tillage practices. Photo credit: David Keusel, used with permission, previously published in chapter by Ogden et al. (2021).

case of impervious surfaces, or those with a small potential infiltration rate due to disturbance, compaction, or surface sealing by wind-blown clays. This situation offers high predictability when modeled using infiltration–excess models with a small number of parameters yet exhibits high sensitivity to rainfall rate. At the right-hand limit of the plot on Figure 1, soils with high infiltrability allow all rainfall or snow melt to infiltrate, and stormflow is produced on areas where the soil is saturated from below by the saturation–excess mechanism. Topmodel (Beven and Kirkby 1979) and other topographic wetness index models such as TOPKAPI (Ciarapica and Todini 2002) and the so-called "One Parameter Model" (Pradhan and Ogden 2010) exhibit considerable skill at simulating saturation–excess stormflow generation using a small number of parameters, albeit with considerable sensitivity to water balance errors. Some studies suggest a correlation between saturation–excess stormflow generation and curve number-like approaches as well (Steenhuis et al. 1995).

In the case of soil-dominated hydrology, stormflow generation is dominated by a host of phenomena occurring below the land surface that are difficult to observe and that evade quantification and representation in models (e.g., McDonnell et al. 2010; Blöschl et al. 2019). These include soil layering (Phillips and Lorz 2008), preferential flow paths (Zhang et al. 2018), trampling by hooves and compaction by farm equipment (Mulholland and Fullen 1991; Alaoui et al. 2018; Baumhardt et al. 2017), freeze/thaw (Sjöberg et al. 2021), shrink/swell processes (Stewart et al. 2016), shallow groundwater table dynamics (Scaife et al. 2020), evapotranspiration over different land cover and terrain types (Wang and Dickinson 2012). Figure 2 illustrates the challenge associated with modeling areas where land use practices influence soil hydrologic behavior.

Furthermore, it is not just soil hydrology that complicates model representations. Cold-region processes, for example, comprise another complex set of phenomena that are difficult to observe. These include energy-driven rapid snowmelt plus the storage and release of liquid water during rain-on-snow events (Marks et al. 1998; Rössler et al. 2014), the distribution of snow in complex terrain (López-Moreno et al. 2013; Clark, Hendrikx, et al. 2011; Clark, Kavetski, and Fenicia 2011), snow sublimation in windy environments (Sexstone et al. 2016), and river ice formation and breakup, among others.

Understanding the challenges associated with model selection and how they relate to hydrologic processes, Spieler et al. (2020) proposed calibration of model structure as an alternative solution to the model selection problem. Using calibration routines to identify effective model structures eliminates a priori identification of model structure. Use of this approach can be straightforward with regard to lumped/conceptual models (Chlumsky et al. 2021) but becomes much more difficult when considering models that employ some kind of spatial discretization such as physics-based or semi-distributed models.

Alternatively, model intercomparison projects (e.g., Smith et al. 2004, 2012; Reed et al. 2004; Rutter et al. 2009; Xia et al. 2012; Essery et al. 2013; Cai et al. 2014; Maxwell et al. 2014; Tegegne et al. 2017; Kollet et al. 2017; Krinner et al. 2018; Baroni et al. 2019; Tijerina et al. 2021) can provide valuable information at the level of model structure, but often do not provide information about the advantages offered by individual components. They may reveal that one model performs better than others in a particular setting, but not necessarily why. Scientific evaluation to get at the answer to this question requires controlled comparisons at the level of individual process models/modules, ideally in an environment where as much as possible is equal in every regard outside of the process under evaluation (Clark, Hendrikx, et al. 2011; Clark, Kavetski, and Fenicia 2011). This need places the emphasis on interoperable and swappable models and modules, which is promoted by standardization.

Recently, Knoben et al. (2025) illustrated that uncertainties in model structure, parameters, and forcing can hamper the identification of optimal model formulations on a regional basis. For this reason, Knoben et al. (2025) suggest execution of multi-model ensembles, where a set of candidate formulations are applied over an entire domain of interest and evaluated regionally. Evaluation might determine weights to apply to different ensemble members in forming a prediction. This solution is significantly more computationally intensive and does not eliminate the uncertain step of model parameter regionalization.

In the above intercomparison projects and other modeling literature, there is ample evidence in support of three main points:

1. Hydrologic models specifically designed to simulate dominant local processes generally outperform models with a larger number of parameters that emphasize process through parameter tuning (Beven and Freer 2001; K. J. Beven 2005; Roberts 2007; Chlumsky et al. 2021). This is known in the literature as the "Uniqueness of Place Hypothesis" (K. J. Beven 2000).
2. Conceptual or physical–conceptual hydrologic models with an appropriately small number of parameters generally perform as well and with higher confidence compared to models having a larger number of parameters, while requiring less calibration effort and improving parameter identifiability (Stedinger and Lettenmaier 1996; McMahon et al. 2006; Perrin et al. 2001). This is known in the literature as the "Parsimonious Model Hypothesis."
3. There is also growing evidence from the machine learning literature suggesting improved predictive power from data-driven models using robust training on large catchment samples (Gauch, Mai, and Lin 2021; Xu and Liang 2021). This also extends to ungauged basins (Kratzert et al. 2019; Nearing et al. 2024), ungauged regions (Feng et al. 2023; Nearing et al. 2024), and extreme events (Frame et al. 2022). However, machine learning models are not exempt from the challenges of Uniqueness of Place (K. Beven 2020) and Parsimony requirements (de La Fuente et al. 2023). While these methods show promise for predictions of hydrologic quantities, rigorous comparisons and model evaluation must continue in parallel with process-based modeling, particularly when phenomena that are difficult to observe are important.

As evidenced by the above three points, there is no one perfect model. This fact suggests the need to consider and apply different models in different portions of a model domain and in service of different applications, which creates challenges related to model-specific computing and data needs. One possible solution to this is to identify a means to improve model interoperability in a way that also promotes model evaluations and intercomparisons in a consistent and fair manner.

1.3 | Community Modeling

The above paragraphs make clear that there is no lack of hydrologic models and that the preponderance of data models, model ecosystems, and proprietary development processes restrain the advancement of hydrologic modeling and understanding. This is not a new observation. The literature contains calls for unification of hydrologic modeling dating back to the 1990s. The thinking started out along the lines of "if we can all agree on a community model, we can advance and stop reinventing models/data models, etc." Jakeman and Hornberger (1993) proposed a framework for environmental model development and evaluation. Refsgaard et al. (1996) discussed the need for a structured approach to hydrologic model development and evaluation. By the end of the 1990s, the promise of spatial hydrology ushered in by the start of the GIS era in the late 1980s seemed exhausted. Both K. J. Beven (2001) and Singh and Woolhiser (2002) concluded that because of uncertainties in process, process representation through model structure, and parameters, the likelihood of identifying skillful model formulations from spatial information related to topography, physiography, and land-use/cover was very small.

Weiler and Beven (2015) asserted that there are currently too many hydrologic models, which has led to duplication of effort and a lack of agreement on process representations. They proposed a community model as a platform for testing different modeling concepts and promoting interdisciplinary applications. However, they acknowledged challenges such as agreeing on modeling concepts and securing long-term financial support for such an effort. The paper emphasized the benefits of collaboration, knowledge sharing, and the development of a unified framework for hydrological modeling.

A series of workshops involving over 200 hydrologic researchers sought to identify the most pressing needs in hydrologic modeling. That effort assessed research community interest in developing a "community model" (Blöschl et al. 2019). Accordingly, this issue was given low priority, with Blöschl et al. (2019) concluding that "… the context-dependence or uniqueness of place (K. J. Beven 2000) continues to be considered a relevant factor in hydrology, notwithstanding a range of modular models and model repositories that have been developed in the past decades (e.g., Clark et al. 2015; CSDMS 2019)."

1.4 | Modeling Frameworks

Given persistent constraints in getting researchers and institutions to agree on a single community model, modeling frameworks may provide a more likely avenue of collaboration. For example, Addor and Melsen (2019) suggested that adopting modular modeling frameworks can enhance model adequacy and foster a collaborative and responsive model development environment. This represents a paradigm shift in hydrologic modeling resulting from our deepening understanding over time of the range of valid options and strategies available for representing hydrologic phenomena. The global hydrologic modeling community has introduced a broad variety of models that vary in structure, complexity, process specification ("physics" or "parameterization"), physical realism, inputs, and parameter estimation approaches (Fatichi et al. 2016). Flexible modeling frameworks allow researchers to entertain competing hypotheses about the most suitable modeling approach to apply in a given region, yet without forcing the choice of a single model.

K. J. Beven (2001) discussed the concept of a six-part "Alternative Blueprint" for distributed modeling as a modeling framework. K. J. Beven (2001) also suggested using an uncertain or fuzzy landscape space to model space mapping as the basis for this framework. This framework included a priori rejection of model structures inconsistent with a perceptual (i.e., an observation-informed) model of the catchment, then testing of remaining models and rejection of those that produce unacceptable predictions. The Alternative Blueprint emphasized prior evaluation of models in terms of physical realism and the value of data in model rejection. It also emphasized the need for testable hypotheses and the potential for learning about different places through distributed modeling. While K. J. Beven (2001) focused on distributed process-based hydrological models, the proposed Alternative Blueprint seems equally applicable to conceptual models as well.

Wagener et al. (2001) similarly proposed the use of a hydrological modeling framework that balances the level of model complexity with the available data and the desired application. Their proposed framework would seek optimal use of the available data to identify model structure and parameters and allow for detailed analysis of model behavior. The proposed framework emphasized the importance of considering the modeling purpose, characteristics of the hydrological system, and available data in determining the appropriate level of model complexity.

Since 2001, the development of several model construction frameworks has moved towards more generalized modeling. David et al. (2002) introduced the Object Modeling System (OMS), a multi-modeling framework related to work by Leavesley et al. (2002) to develop a Modular Modeling System (MMS), which centered on the USGS Precipitation-Runoff Modeling System (PRMS). David et al. (2013) later discussed broader issues around the engineering and institutional requirements of such "environmental modeling frameworks" (EMFs). Clark et al. (2008) developed the Framework for Understanding Structural Errors (FUSE), which allowed for flexible selection of conceptual models (lacking a snow component) to diagnose differences in model structures and their effects on simulated streamflow. Others developed modeling systems that allow ready comparison of different model structures. Examples include SUPERFLEX (Fenicia et al. 2011), MARRMoT (Knoben et al. 2019), and Raven (Craig et al. 2020), which tend to apply modules formulated for HRU or catchment scales to construct models.

Clark and Kavetski (2010) noted that many commonly used models in hydrology were undermined by inadequate numerical implementations, with first order impacts on their performance. In part, this motivated Clark et al. (2015) to develop a flexible and process-agnostic framework called the Structure for Unifying Multiple Modeling Alternatives (SUMMA), in which all process schemes use an independent numerical solver. The separation of physics and numerics is more common in other branches of geosciences, for example, atmospheric and ocean science, than in surface hydrology (Keyes et al. 2013; Miller et al. 2013). The primary reason for this is that atmospheric and oceanic sciences tend to use solver-centric continuum models, some with adaptive grids and time-stepping, whereas hydrologic states, such as overland flow (Fiedler and Ramirez 2000), channel flow (Godsey and Kirchner 2014), permafrost, and perched groundwater, are often discontinuous in space and time and require an adaptive solution of some sort.

The approach used in SUMMA contrasts with other multi-modeling strategies that couple existing model codes, such as with simple "wrappers" (e.g., Leavesley et al. 2002; Pomeroy et al. 2007; Werner et al. 2013), tight "couplers" that control the time-stepping and flow of information between models (e.g., Craig et al. 2011; Hutton et al. 2020), or probability-based theoretical frameworks such as in Montanari and Koutsoyiannis (2012).

Notably, many frameworks often require new modules to be written in a single programming language, representing a challenge to interoperability. Wholesale re-writing of model codes into a different programming language increases the likelihood of "orphaned code," where the original author no longer recognizes their own creation and cannot provide support.

1.5 | The Challenges of Having Many Models and the Open-Source Solutions

Over time model developers have employed a variety of programming languages to create models, including Fortran, C, C++, R, Python, and others. Some models run easily on a small computer with a single CPU, while others require a supercomputer. Some models have self-contained source code, while others require a long list of external code dependencies—that is, other software stored in libraries that the user must install in order to run the model. New models often apply new mathematical representations of processes that describe fluxes, changes in storage, and resulting changes in residence times. However, significant advances regarding novel solutions of the fundamental equations or constructs occur infrequently. Instead, much of the reinvention around development of new models arises from peculiarities related to input/output data requirements and formatting or specialized use cases.

Nearly all hydrologic models require custom or heavily modified datasets with derivative, new, or adapted data models to represent the discretization of landscape and river components (Blodgett et al. 2021). Not only are there a variety of gridded, lumped, and graph-based (i.e., dendritic) data models, but even when the same conceptual landscape discretization is identified, differences can arise from issues related to data resolution, time of collection, and the software and settings used to identify and manipulate the modeling units. For example, the NOAA/NWS NWM version 3.1 and earlier uses a derivative of the National Hydrography Dataset Plus (NHDPlusV2; McKay et al. 2012) network within the contiguous US (CONUS) and a combination of NHDPlus and NHDPlus-High Resolution derivatives, and custom derived hydrography sets for the areas outside of CONUS (oCONUS), including the Canadian portion of the Laurentian Basin, tributaries originating in Mexico and flowing to the Rio Grande, Alaska, Hawaii, Puerto Rico and the U.S. Virgin Islands (Cosgrove et al. 2024).

Early versions of NOAA's continental scale Flood Inundation Mapping capabilities also began with the NHDPlusV2 as a source network (Maidment 2017; Johnson et al. 2019); however, as capabilities evolved, the underlying hydrography and landscape discretization was modified such that it yielded different realizations of the surface water drainage network inconsistent with that of the NWM. The long history of large modeling enterprises (e.g., USGS, NOAA, NCAR) making individual modifications to NHDPlusV2 without a means of contributing and rectifying these changes in the source dataset have led to the creation of the National Reference Network (Blodgett et al. 2023) and its supplementary activities (e.g., Modaresi Rad et al. 2024) used for water resources modeling efforts across the USGS and NOAA.

The existence of varied data models impedes interoperability and scientific evaluation of outputs because of mundane issues related to format, coordinate reference systems, discretization, data source, forcing requirements, and a litany of other preprocessing steps that drive many assumptions and the applicability of a given model. As such, model interoperability typically involves a series of steps to identify how one model represents the same area of the earth's surface, regardless of the underlying assumptions of the model and its inputs.

The variety of modeling formulations, programming languages, compilers, data models, and required inputs create unique "model silos" (Nativi et al. 2013), which require considerable effort to understand and apply. In practice, this means that the work that goes into preparing and running one model is often neither reproducible nor transferable to another model, even when run on the same compute platform (Tucker et al. 2022). This impedes model interoperability.

The complexity of model setup can be mitigated by development of model-specific user interfaces to reduce the burden placed on a model user to understand, configure, and generate the underlying inputs required to run a model. In some cases, government agencies and research institutes develop such interfaces along with software ecosystems that aid in the execution of their mission but impede interoperability with those of other groups. While such interfaces and model ecosystems can be powerful and greatly reduce the knowledge required to set up and run models supported by those interfaces, models outside those supported remain a challenge to understand, set up, and run. Furthermore, proprietary or closed-source models are frequently walled off from community improvements. All of these factors impede model development progress and the optimization of simulated outputs.

Community model development that leverages frameworks and standards and considers interoperability while supporting model development needs represents an alternative to model-centric development. The creation of open-science communities depends on shared foundations, including open-source software and community standards. These give users a shared understanding of the computational elements of a system, what they are called, and how they are referenced. This is central to developing core conceptual and logical models. Only with a common foundation can software and data models be written that support making data more findable, accessible, and ultimately more reusable and, most critically, interoperable. By using such a system, researchers and operations professionals can spend less time on the mundane—and often error-prone—pre-processing steps—and place more effort on model development, evaluation, and integration.

The open development of standards through entities such as the Open Geospatial Consortium (OGC) and Community Surface Dynamics Modeling System (CSDMS) community, along with widely accepted software development and version control systems like Git, allows for broader engagement. They also provide a record of the assumptions made by developers. Together, these features result in a process that can achieve the FAIR (findable, accessible, interoperable, reusable) data objective (Wilkinson et al. 2016).

1.6 | Experience of the NOAA/NWS National Water Model

One model application or use case for successful, interoperable frameworks is operational water prediction in support of extreme water event forecasting/mitigation and water resources management. The US National Weather Service (NWS), Office of Water Prediction (OWP) regularly produces operational streamflow guidance using the National Water Model (NWM) (Cosgrove et al. 2024). OWP made version 1.0 of the NWM operational in August 2016, based on a specific configuration of the WRF-Hydro model (Gochis et al. 2020). Configured as the NWM, WRF-Hydro represents land surface column physics using the Noah-MP point-process land surface model on a 1 km grid (Cosgrove et al. 2024). In addition, the NWM configuration of WRF-Hydro adds surface and lateral subsurface flow routing components, each running on a 250 m grid, plus a nonlinear reservoir groundwater model. All land surface model routing outputs and groundwater outflows

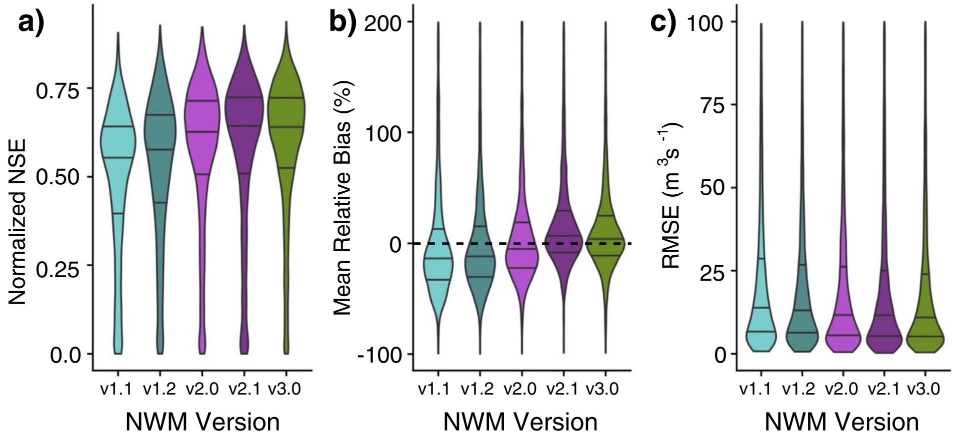

FIGURE 3 | Violin plots for normalized Nash-Sutcliffe Efficiency (NSE) (a), mean relative bias (b), and root mean squared error (RMSE) (c) for NWM versions 1.1 through 3.0 based on historical streamflow simulations from 2013-10-01 to 2016-10-01 compared to USGS streamflow observations at 4571 gages. The lines within each violin denote the 25th percentile, the median, and the 75th percentile. To better view the data, we limited the *y*-axis range of subpanel (b) to −100% to +200% (excluding 5.7% of the data set with outlying mean relative bias values) and the *y*-axis range of subpanel (c) to 0–100 $m^3 s^{-1}$ (excluding 8.2% of the dataset with outlying RMSE values).

TABLE 1 | Median error metrics for NWM versions 1.1 through 3.0 based on historical streamflow simulations from 2013-10-01 to 2016-10-01 compared to USGS streamflow observations at 4571 gages.

NWM version	Rel. bias (%)	Normalized NSE	RMSE $m^3 s^{-1}$
v1.1	−10.4	0.55	14.98
v1.2	−10.0	0.58	13.95
v2.0	−3.2	0.63	12.19
v2.1	9.3	0.64	12.35
v3.0	5.4	0.64	11.45

are input as lateral inflows to a Muskingum-Cunge hydrologic channel routing model. The channel routing model simulates water storage/flood control reservoirs using a level-pool "fill-and-spill" approximation, or using measured or forecast reservoir releases where available.

In the intervening years of NWM operations, the WRF-Hydro based NWM has undergone a series of upgrades. Each upgrade included a comprehensive assessment characterizing changes in model skill from version to version, ensuring that each version represents an improvement in streamflow simulation skill over the last. These assessments used a multi-year set of NWM streamflow outputs, generated as part of the overall NWM calibration, parameter regionalization, and validation process. While version-over-version gains have been achieved for every iteration of the NWM, in some regards, performance improvements through calibration and regionalization have diminished over time (Figure 3, Table 1). In both cases, while version-on-version improvements continue, the pace of improvements is beginning to plateau.

Another pattern emerges when analyzing the skill of NWM peak discharge across regions of the US. Figure 4 shows the median absolute error (MdAE) in peak discharge from analysis of NWM version 3.0 retrospective output (2013–2018) using an event detection algorithm during the warm season. Errors were calculated by comparing NWM output against observations at the USGS GAGES II network, selected because the influence of reservoirs and diversions was minimal (Falcone 2011). Figure 4 uses larger dots to show stations where the MdAE was less than 25%. The violin plot shows the distribution of event-scale peak discharge MdAE across the entire GAGES II network. This figure clearly shows that in terms of peak discharge, the current operational NWM formulation performs best in the Pacific Northwest and northern Rocky Mountains. These results provide support for the need for mosaic modeling capabilities as previously described. These patterns, combined with the ever-present need for improving hydrologic forecast accuracy, serve as one of the key drivers for considering the development described in this paper.

1.7 | Objective: A New Way Forward

The accuracy of hydrologic forecasts produced by the NWM and other operational systems depends on the accuracy of a series of models and a number of observations. These include observed and simulated meteorological forcings, earth observations and their assimilation. Advances in each of these areas improve hydrological predictions. For the work reported here, however, advancing hydrological modeling is our foundational focus. Here we rely on our experience in producing operational streamflow forecasts and a synthesis of the existing literature, which provides support for two important needs to advance hydrologic modeling (e.g., Clark et al. 2015) and move beyond the limiting assumption that one model formulation will work well everywhere. These objectives include:

1. Identify a common representation of the important surface water hydrology features of the landscape. The setup of any surface water hydrologic model involves a common set of steps. These include: application of contributing area thresholds to identify streams, catchments, and their connectivity; identifying lakes; reducing noise in underlying topographical data sets to eliminate "pits" or "digital

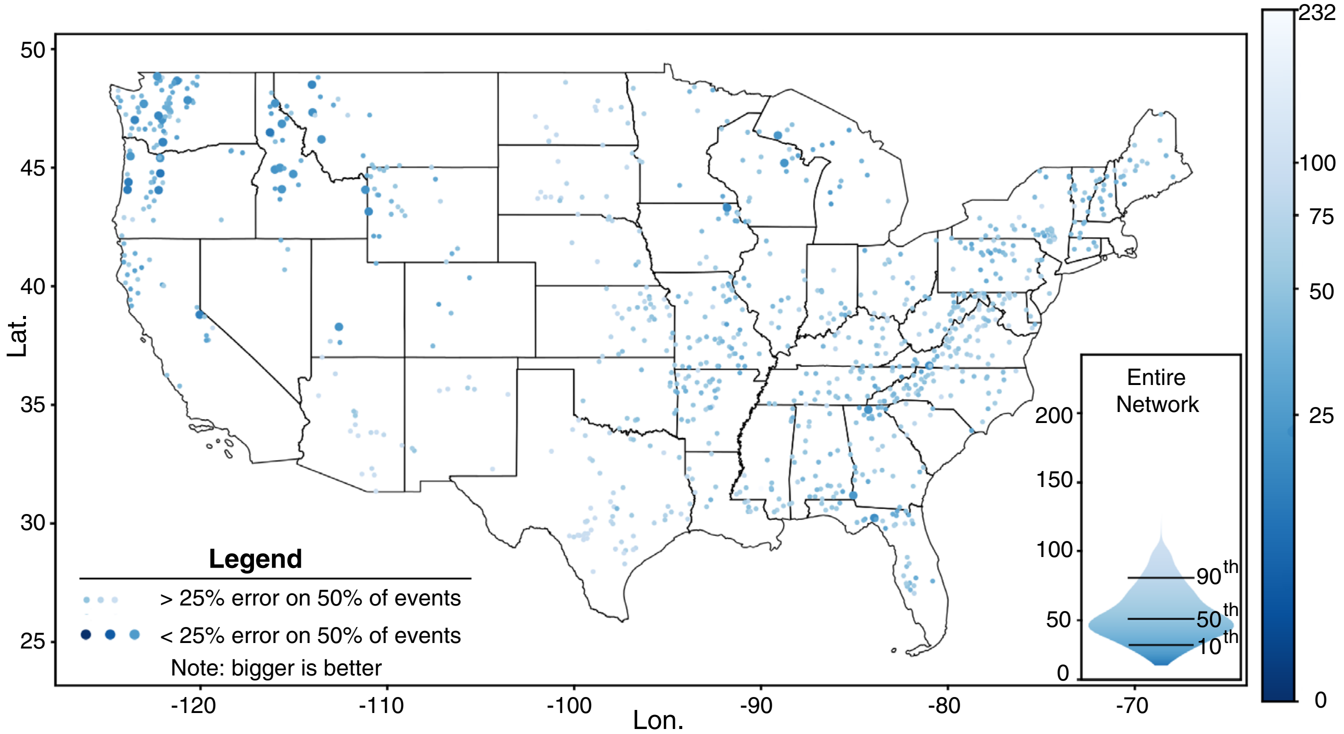

FIGURE 4 | Median absolute error (%) in event peak discharge during the warm season from the 1979–2023 NWM 3.0 retrospective evaluated against the USGS GAGES II network in the contiguous US.

dams" that artificially create regions of adverse (upslope) flow; etc. After identification of these "hydrofeatures," digital representations of catchment boundaries, stream locations, etc. allow sampling of other georeferenced data sets for estimation of model parameters. Most existing models have a unique workflow that performs this process and produces outputs in a model-specific format. In reality, transforming the same underlying topographical inputs into myriad model-specific creations represents a huge impediment to model interoperability. Identifying a common representation and developing the tools that create it will prevent a great deal of duplication of effort and promote model interoperability.

2. Identify the characteristics of a computational environment that provides the capability to test and scientifically evaluate different representations of hydrologic processes, using different discretizations to describe spatial variability and hydrologic connectivity, using different model structures and paradigms (e.g., conceptual vs. process-based vs. data-driven). Many existing modeling frameworks impose limitations on allowable model structures (e.g., spatial and temporal discretizations and resolutions, solvers, programming languages, operating environments). A framework that successfully overcomes these limitations in a way that allows a diversity of model formulations, resolutions, and discretizations and selection of different dominant fluxes will represent an advance.

Meeting these objectives requires identification of community needs. A successful framework based on these objectives promises to facilitate community contributions to its development. Over the past 20 years, the rise of open-source software in computer science and the education of a generation of computer and domain scientists to think in terms of collaborative, shared, and open frameworks have transformed large software development projects. These transformations have enabled advancements in other scientific fields. One prominent example is in the geospatial sciences, where researchers have built a proliferation of open-source data manipulation and display packages (e.g., QGIS), data formats (Cloud Optimized GeoTIFF [COG], Zarr, Geopackage), and functional programming language extensions (e.g., sf and terra in R, and fiona and geopandas in Python) using the open source GDAL library (GDAL/OGR contributors 2022). Each of these resulted from user-driven communities where companies such as Microsoft and Esri, individuals, and governments contributed to open science (https://makepath.com/history-of-open-source-gis/). A similar community in hydrologic modeling could be just as impactful, but such a community has not yet gained the same momentum and contributions. Possibly this results from a lack of a community-led, community-developed open-source, model-agnostic, standards-based, interoperable modeling framework.

2 | Advancing Community Hydrologic Modeling With the Next Generation Water Resources Modeling Framework

What would such a framework look like? Perhaps the challenge of developing the long-advocated-for-but-unrealized "community model" can be met by developing a "community modeling framework" designed to promote interoperability and allow scientific evaluation of different modeling approaches using standards for model coupling and geospatial data. This approach avoids challenges associated with the questions considered above regarding the starting-point, discretization, purpose, and temporal nature (e.g., run duration, output interval, timestep, etc.). The proposed solution would not constrain the simulation type or purpose; rather, researchers can bring their own models and assumptions to the "community modeling framework." In the context of this paper, "models" might be an assemblage of different coupled process simulation codes that simulate more than one process, or a well-defined module that simulates one process, or a "process module."

To incorporate the wide range of hydrologic/hydraulic models and their paradigms, we need a framework that is truly interoperable and model-agnostic, relying on standards similar to those underpinning the previously discussed geospatial science tools. This would provide a way past the previously described difficulties with community models and hydrologic models in general. The framework should seek to minimize built-in assumptions about its potential use cases. A domain scientist should be free to use the framework for whatever purpose they see fit, provided that the standards it employs are compatible with their codes. We call this framework the Next Generation Water Resources Modeling Framework (NextGen).

2.1 | Origins of the Next Generation Water Resources Modeling Framework

In response to NWM performance evaluations (e.g., Cosgrove et al. 2024) and evidence from the literature regarding the general challenges of hydrologic modeling, in 2019 OWP commissioned a review (Dunkman et al. 2019) of the NWM code by the 18F Group, which was a component of the U.S. Government's General Services Administration that helped federal agencies develop software and technology that better serves the public. That review recommended "refactoring," or essentially rewriting the code to eliminate unused features, clarify components, and improve usability with modern software design standards. The report suggested that refactoring the NWM code to improve modularity should treat coupling interfaces as "products" with the aim of increasing community involvement (Dunkman et al. 2019).

The first exploratory refactoring attempts sought to increase the modularity of WRF-Hydro to ease code swapping and scientific evaluation of different process modules. These attempts revealed several limitations with that approach. These largely arose from the evolution of WRF-Hydro from a research code to operations; the codebase contained relic code left over from its developmental history that slows interpretation by people trying to make improvements, which is known as "intellectual debt." Furthermore, as applied in the NWM configuration, the WRF-Hydro model applies a fixed, structured-grid discretization, which hampers consideration of hydrologic modeling approaches that use other discretizations. Finally, the code base is written in Fortran, making it difficult or in some cases impossible to couple with models written in other programming languages.

Discussions around refactoring led to clean-sheet thinking and meetings with other federal agencies about how to overcome these limitations by creating a new modeling framework with increased flexibility, modularity, and extensibility. The guiding question for the early stages of this process was: "How can we refactor the NWM to accomplish these desired outcomes?" We called the project NextGen in a nod to the evolution of operational water resources forecasting. As envisioned, NextGen is not a model. Rather, it is a model-agnostic, standards-based, interoperability software tool that enables construction and execution of water prediction models and evaluation of model outputs.

2.2 | Requirements

To meet the objectives and guide NextGen development, an initial meeting was held in August 2020 to discuss desired features in a refactored NWM. That meeting involved participants from NOAA and OWP, the US Dept. of Defense (US Army Corps of Engineers, Engineering Research and Development Center, USACE-ERDC), US Dept. of the Interior (US Bureau of Reclamation USBR, and US Geological Survey, USGS), and the National Center for Atmospheric Research (NCAR). The outcome of that meeting was a consensus that the solution would take the form of some kind of modular system that supported the needs of all federal water prediction agencies and promoted model interoperability. The federal agencies agreed to further meetings (Figure 5) to explore the concept.

A follow-on meeting focused on requirements was held in October 2020 involving only personnel from US water prediction agencies including: NOAA, USACE-ERDC, USBR, USGS, and the US Dept. of Energy. The outcome of that meeting was a set of requirements that the participants agreed represented desirable features of a water resources prediction model interoperability system:

1. Provide maximum flexibility. In short, do not make decisions outside of the standards selected that prevent someone from using the framework as they see fit. Furthermore, such flexibility will allow that as models, data sources, and modeling needs evolve, the framework can support those changes and additions.
2. The framework design shall be model agnostic. In principle, the framework might find uses outside of the traditional water prediction needs of hydrology/hydraulics.
3. Employ standards to establish a common architecture that helps avoid duplication and promotes interoperability. Identify or develop needed standards for coding, coupling, data and metadata, and model verification/validation and test data.
4. Develop the framework in an open-source repository. This will promote code reuse and development efficiency and ease/encourage participation by partners and the

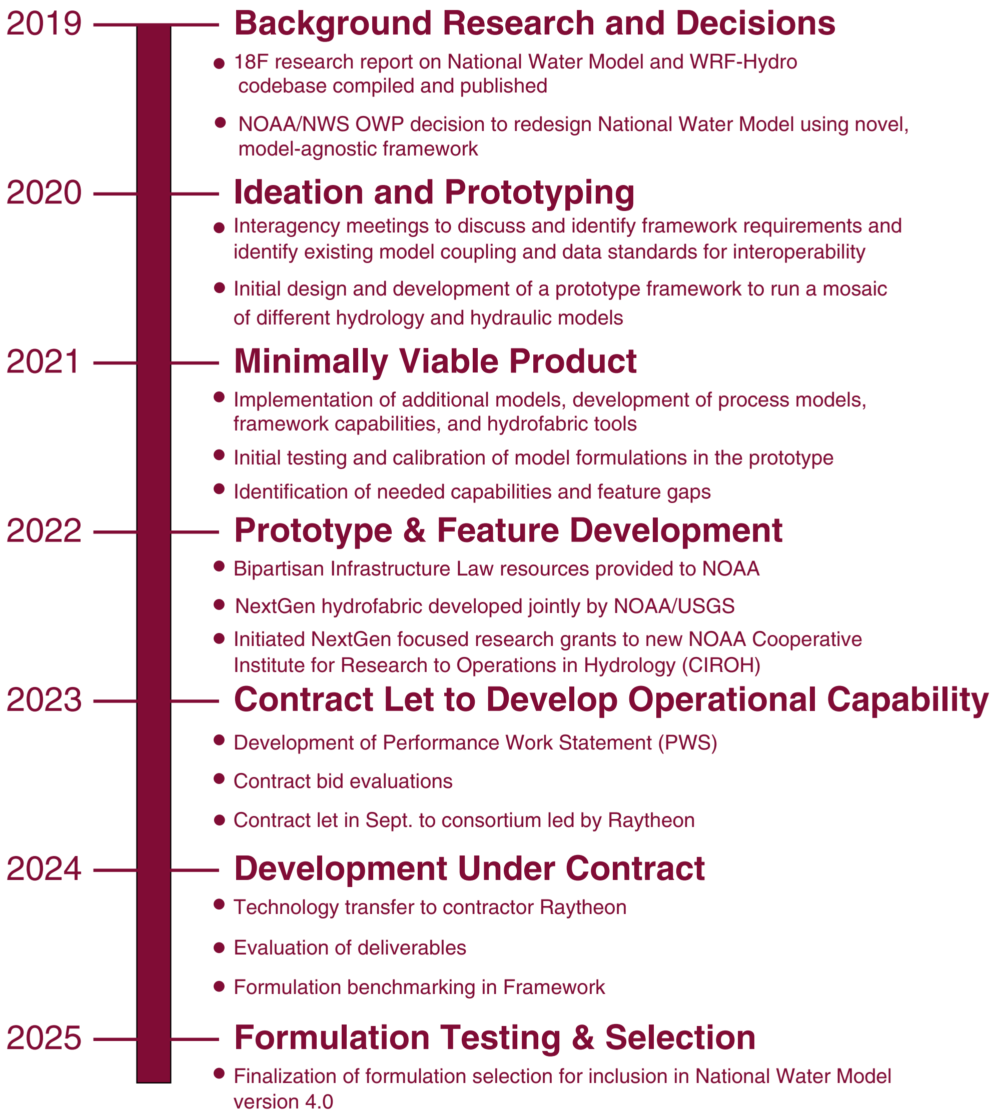

FIGURE 5 | Development timeline of the Next Generation Water Resources Modeling Framework.

community. It will also allow for creation of an authoritative repository for federal water models.

5. Provide user-friendly development paths to domain scientists and engineers. Implementation of models in the framework shall be as minimally invasive as possible so as to not require a rewriting of domain science code. The original author of a model should be able to recognize their code.
6. Commit to sharing model codes, evaluation tools, and datasets.
7. Establish and maintain a glossary and define terms to communicate clearly across disciplinary boundaries. This requires that definitions be established in a disciplinary context, because the same term often means different things in different disciplines.
8. Use mature open-source libraries where appropriate.
9. Provide support for multiple programming languages. The set identified included the following widely used scientific programming languages: C, C++, Fortran, and Python. This list is not exclusive but a consensus opinion of the present minimum set.
10. Support execution of models on diverse hardware ranging from laptops to supercomputers. Contemporary libraries for shared- and distributed-memory run time environments and container/cloud solutions represent enabling technologies.
11. Establish a 2-week target for a graduate student or new employees to add a model or module to the framework. This will require excellent documentation and step-by-step examples/tutorials. For accessibility to domain scientists, this will require that they possess programming skills but not a computer science background.

2.3 | Adopted Design Standards

A third meeting, held in February 2021, focused on the identification and selection of the most appropriate standards for widespread adoption and consistency with the framework requirements. The meeting identified the need for two different standards. The first was a standard providing a common description of surface water hydrologic and hydraulic features to unify model applications, because all models simulate some version of the same physical system, albeit with unique internal discretizations and scales of interest. The second was a standard method for coupling diverse hydrologic, hydraulic, and coastal models and promoting interoperability in a wide range of computing environments.

In February 2021, only one standard existed for describing the hydrologic and hydraulic features of the land surface and was deemed suitable to support modeling. That standard was the Open Geospatial Consortium (OGC) WaterML 2: Part 3—Surface Hydrology Features (HY_Features) conceptual model (Blodgett and Dornblut 2018). HY_Features identifies and defines the core concepts underpinning surface hydrologic and hydrographic data across a range of use cases and solidifies these concepts into a common system with codified associations. The group in the February 2021 meeting agreed this standard was sufficient to increase interoperability of surface water hydrology and hydraulic models.

As a conceptual model, the HY_Features standard defines relevant surface hydrologic and hydraulic features, but not the functional elements that support modeling applications and data creation. The set of surface water hydrologic and hydraulic features considered by the HY_Features conceptual data model is defined as a "hydrofabric." Since its introduction in 2018, the HY_Features conceptual model has been extended to support a logical model for dataset integration (Blodgett et al. 2021), a general data model for hydrofabric and river corridor information, and an explicit, interoperable data product that supports a range of large-scale modeling tasks (Bock et al. 2022; Johnson 2022). Questions still remain around the completeness of this paradigm for supporting other unique model types such as groundwater in aquifers and snow processes; however, the process for adopting, refining, and extending the standard is well documented.

Four different model coupling strategies were also discussed: the Basic Model Interface 2.0 (BMI) standard developed by the US NSF-funded Community Surface Dynamics Modeling System (CSDMS) (Peckham et al. 2013; Hutton et al. 2020; BMI 2023), the OpenMI shared-memory model coupling standard (Dagum and Menon 1998), the Coupling Approaches for Next Generation Architectures (CANGA) Tasked-Based standard developed by the US Dept. of Energy (Coon et al. 2019), and the Earth System Modeling Framework (ESMF) standard developed by NOAA (Hill et al. 2004). After discussing the requirements, the group identified the paramount importance of a "thin" standard offering maximum flexibility while being minimally invasive. This latter point was viewed as extremely important to attract community involvement. The assembled group unanimously agreed that the BMI standard is most appropriate to meet the previously stated design requirements.

Computational models in hydrology and other fields are numerous and exhibit considerable differences in terms of programming language, dimensionality, discretization of space and time (i.e., grid and time stepping scheme), input and output variable names, variable units, and data types. The BMI standard is a lightweight, standardized API (i.e., an Application Programming Interface or a wrapper or adapter) for computational models that enables models and even datasets to interoperate and share the values of their variables, despite all of these possible differences (Gan et al. 2023).

The letter "B" in BMI stands for "Basic." The BMI standard is not comprehensive, and other useful metadata describing models such as design minimum/maximum/optimal time step, spatial scale limitations, seasonality limitations, and design assumptions might not be available through the BMI interface. For this reason, an auxiliary model metadata description might be desired to help a coupling framework determine the appropriateness of coupling two models, or for comparison against a perceptual model (McMillan et al. 2023) to evaluate formulation completeness.

The BMI functions can be grouped into five categories: Model Control Functions (i.e., initialize, update and finalize), Model Information Functions, Variable Information Functions, Variable Getter and Setter Functions, and Discretization Information Functions. These standard BMI functions allow a centralized framework to start and stop a model; execute one or more timesteps; query model information; and get, set, and pass variable values, among other capabilities. By design, the BMI functions are noninvasive and employ the so-called "Hollywood principle" from computer science—"Don't call us, we'll call you." That is, a model coupling framework can query and control a BMI-compliant model, but the model does not make any calls to other components or tools and is not modified to use new data structures. Making a model BMI compliant introduces no new or external dependencies into a model so that it can still be used in a "stand-alone" manner or within a model-coupling framework. The framework and its mediators, not the models, are responsible for passing the values of variables between models and for performing any required transformations. This conforms to the computer science concept known as "separation of concerns."

The BMI functions are implementable in any programming language, including C, C++, Fortran (all years), Java, and Python. It is more straightforward in the case of memory address-aware languages, While BMI implementations exist for Javascript, Octave, Julia, Rust, R, and other languages, those were not called out in the NextGen requirements as being widely used by the agency personnel involved to develop hydrologic/hydraulic modeling codes. Even though some languages are object-oriented and support user-defined types (i.e., classes), the BMI functions use only simple (universal) data types. CSDMS provides templates to further simplify implementation in C, C++, Fortran, Java, and Python.

Once implemented using the BMI, a user may run a model as a single unit or couple it with one or more models. Given two coupled models, the framework (1) uses the BMI interface to get values of state variables or fluxes from the first model using its BMI get_value() function, (2) obtains information about each model (e.g., variables, units, grid(s), time step, etc.) via BMI calls, (3) passes this information to framework utilities, often called mediators, that perform transformations such as spatial interpolation/extrapolation, temporal interpolation/averaging, and unit conversions, and (4) sets values of variables in the second model using its BMI set_value() function.

Since model developers must do extra work to implement the BMI for their model, it is important for them to see an overall benefit in doing so. One such benefit is more widespread use of their model, partly stemming from the ability to easily couple it to other models. Another is that the framework can help provide suitable meteorological forcing data, or other input data for topography, land use, and soils. A third benefit is that the framework can access the values of output variables via BMI and then write them to files at any given sampling rate and in any framework-supported file format (e.g., netCDF), with supporting metadata. A fourth benefit is that it is much easier for a new user to understand, use, and modify the model, due to the standard layout and self-description. In addition, it has been shown (Peckham et al. 2014) that a BMI-compliant component can often be used in other model-coupling frameworks that use a different component interface (without change) by applying a general-purpose adapter that maps the BMI interface to some other interface.

2.4 | Identified Gaps

In a component-based, or plug-and-play approach to modeling, another question that arises is that of "granularity," which refers to the relative size of the source code modules (or components) that are wrapped to expose a coupling interface like BMI. While it is possible to wrap complete hydrologic (multi-process) models with BMI, this limits the options for coupling and reusability. In hydrology, these processes include stormflow generation, open channel flow, radiation balance, precipitation, evaporation, infiltration, snowmelt, etc. Therefore, process formulations are a natural "scale" for component granularity because different formulations of the same process may be more suitable in a given context, perhaps based on what input data is available. Moreover, much of the research and the resulting advances in our understanding involve new formulations, which may use new equations, algorithms, assumptions, approximations, or numerical methods for modeling a given process. The ability to easily "swap out" one process component for another while keeping everything else the same enables intercomparison studies, isolating problems, rapid prototyping, and accelerating the science.

Given the characteristics of the selected BMI and HY_Features standards, the group identified the following gaps where further development is needed to facilitate realization of the NextGen Framework:

1. Devise methods that employ the BMI standard to increase its utility on HPC systems. This would include serializing model states for checkpointing and hotstarting of models and for HPC load balancing. This can be performed using the current BMI standard provided that all variables required to entirely save the state of a model and reload it are exposed to the BMI interface.
2. Add component-level coupling, making it possible to construct mimic or novel models.
3. Enable support within the framework for adoption of "complete" model formulations or construction of "componentized" models. This will allow construction of mimic and novel model formulations in the framework.
4. Because it is impossible to anticipate all the different model formulations that might be adapted for use in the framework, model setup workflows must be provided by model developers. Where necessary, these workflows must interpret and use the HY_Features conceptual model for full interoperability.
5. In developing the BMI standard, CSDMS used the "adapter" and "mediator" patterns from computer science. Adapters correspond to "wrappers" or standardized interfaces like BMI (or mappings from one interface to another), that enable "plug-and-play" interoperability. Mediators, as framework utilities, perform transformations like unit conversions and space/time interpolation/extrapolation. It appears that a third type of entity, perhaps called Couplers, will sometimes be needed for more complex domain science codes that enable coupling between models having different states, fluxes, or dimensionalities at their interface (e.g., a river model coupled to an ocean model at a coastline). Where possible, including these in the framework will prove useful.
6. A thin interface like BMI is perceived as beneficial and desirable from a domain scientist perspective. Making a model work within the framework should require added functionality, but models should still run in a stand-alone fashion outside of the framework.
7. Preventing coupling of incompatible models is a framework responsibility. This will require establishment of a model definition standard that contains enough information to achieve this goal. Preliminary attempts at this include the CSDMS standard variable names (http://cdsdms.colorado.edu/wiki/CSN_Metadata_Names).
8. Sub-setting of the model domain by space, time, and model is an important need.

3 | NextGen Framework Prototype

3.1 | High Level Design Concept

Synthesis of the outcomes of the requirements and design meetings, coupled with the specific needs of the NOAA/NWS National Water Model use-case led to a modular design for the NextGen Framework. This design consists of components that each perform a singular function and can run as a subset for alternative configurations or uses. The design shown in Figure 6 resulted from multiple iterations within the NOAA-NWS Office of Water Prediction (OWP), which contains all the functionality needed to evaluate and optionally calibrate a mosaic of models. OWP is currently working with contractors to realize this design starting from an initial prototype development effort which lacks some of the functionality shown in Figure 6.

Specifically, the components that make up the NextGen Framework calibration/evaluation cycle shown in Figure 6 include:

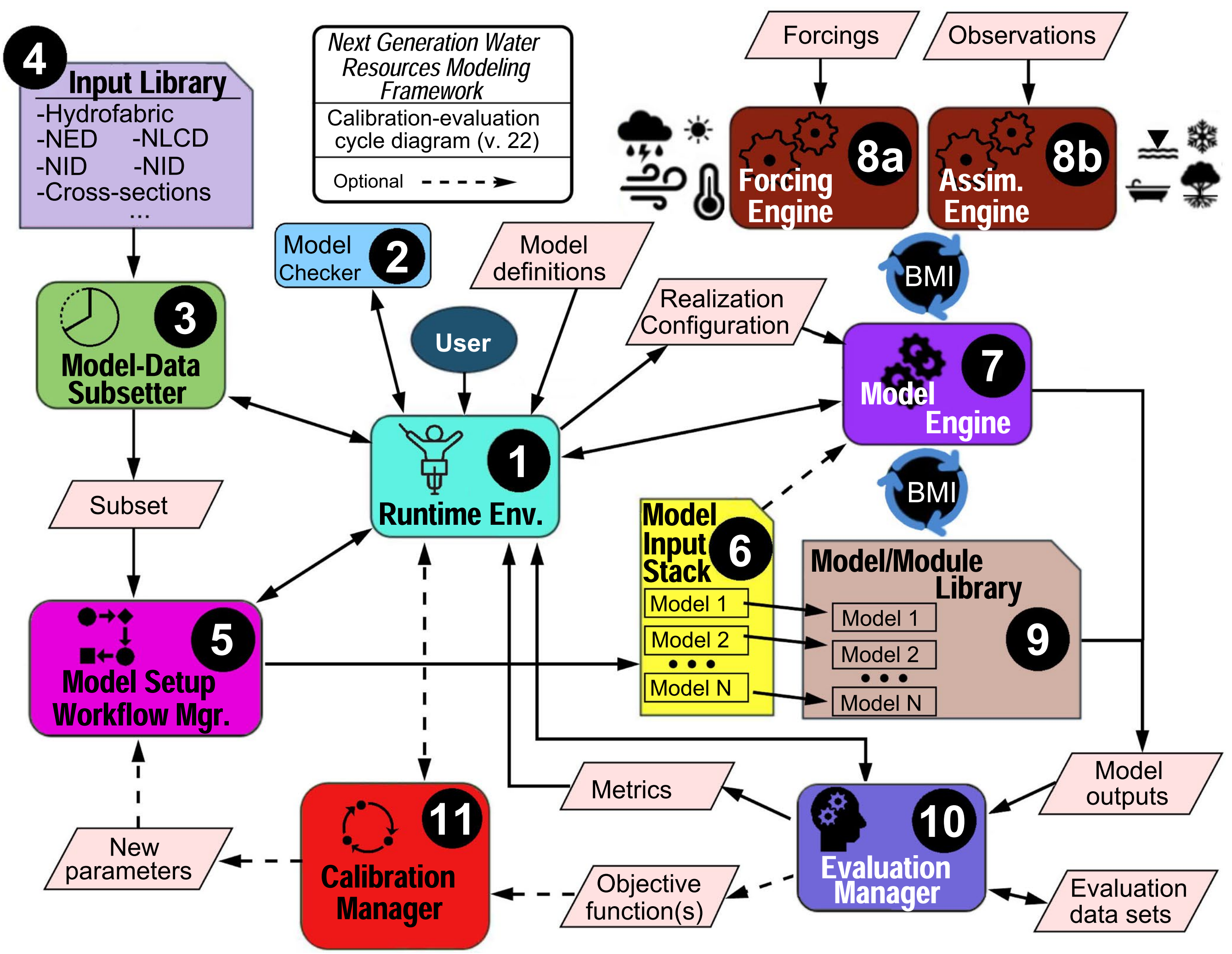

FIGURE 6 | High level design diagram of the Next Generation Water Resources Modeling Framework showing all components.

1. *Runtime Environment* (*RE*): This is the Framework "orchestrator." The user may interact with component 1 through a command line, graphical user, or application programming interface to select which models to run where, specify the source and time-span of the forcings data, which outputs are produced and evaluated using which metrics, and optionally, which automated calibration routines and objective functions shall be applied to calibrate which models. The RE reads a "Model Definition File" or (MDF) that provides model metadata and allows the framework to understand which models are available in the model library, their capabilities and limitations, input needs including the base physiographic data required to generate initial parameter values and state variables. The MDF also contains information on how to run the workflow scripts that produce initial parameter values and state variables. This MDF contains all metadata needed by the RE to evaluate formulation appropriateness and completeness (Ogden et al. 2024) in accordance with an optional perceptual model created by the user. Consistent with the BMI philosophy, the RE is the only component of the framework that is aware of all the other components in the framework; it launches all other components in sequence, and each of those components reports their completion back to it.

2. *Model Checker*: This component performs two important functions. First, it compares the user-specified formulations against a user-specified perceptual model to evaluate model completeness. Secondly, it uses metadata about models from the model definition file to verify that user-specified model formulations use compatible models/modules and that the exchange variables used are recommended and compatible, a check of model appropriateness.

3. *Model and Data Subsetter*: This component produces a spatial subset of all relevant geospatial data describing the physiography of the study area, coinciding with the relevant portions of the hydrofabric. The relevant physiographic data are those described in the MDF as required by setup workflows for each selected model that estimate model parameters and initial state variables for each model in the simulation mosaic. The purpose of the subsetter is to reduce the quantity of data processed by the model setup workflows.

4. *Model Input Library*: Includes the underlying user-generated hydrofabric and all geospatial data needed to estimate model parameters and initial state variables by the model setup workflows.

5. *Model Setup Workflow Manager*: This component executes scripts designed by model developers or model experts that read hydrofabric and physiographical data from the subset and estimate initial model parameters and initial state variables. On completion, it populates the model input stack (Component 6). If running in an optional calibration loop, it can read updated parameter values computed by the Automated Calibration Manager (Component 11) and substitute them back into the model input stack for further calibration runs. In that instance, it can also re-run model setup workflows to estimate new initial model state variable values if necessary.

6. *Model Input Stack*: The set of files that provide parameter and initial state variable values for the various models in the model mosaic. Models may directly read their own input files. Alternatively, these input files may be read by the model engine with values inserted into the models at startup through the BMI. Both cases are useful in different situations.

7. *Model Engine*: This component executes the simulation using the BMI to drive execution of models and modules. When started by the RE, it reads the "Realization Configuration" file containing all the details about the simulation in terms of domain, models, input stack contents and location, simulation period, forcing data source, and desired outputs. It reads forcing inputs using the BMI with the forcing engine, executes models using the BMI, and writes desired outputs. It also provides regular status updates to the RE, allowing it to inform the user of simulation progress and detect anomalies.

8a. *Forcing Engine*: This component is customized to read, process, and re-grid commonly used forcing data for the NWM and provide it to the model engine using the BMI interface.

8b. *Assimilation Engine*: This component accepts input observations for assimilation. Examples might include discharges, water levels, periodic updates to soil-moisture or snow extent, and maps of observations such as NDVI that affect ET parameters or burned areas that affect infiltration parameters.

9. *Model/Module Library*: This contains compiled versions of BMI compliant models/modules, having the capabilities described in the Model Definition File.

10. *Evaluation Manager*: The RE calls the evaluation manager to perform time-series analysis of observations or previous simulation results at multiple points in the hydrofabric. It also has the ability to perform point-to-grid, grid-to-point, and grid-to-grid time-series evaluations using re-gridding functionality provided by the Earth System Modeling Framework (ESMF) re-gridder.

11. *Automated Calibration Manager*: Applies specified parameter estimation routines for particular models using objective functions specified by the user to estimate new parameter values.

The design allows execution of *N* different models and evaluation of model outputs with optional automated calibration, showing major components as numbered and colored blocks. Execution proceeds along increasing component numbers. Parallelograms represent information exchanged between them. The model engine (Component 7) repeatedly interacts with components 8a, 8b, and 9 using the BMI standard until the simulation is complete. Dashed lines denote optional information flows.

The modular design of the NextGen Framework enables a variety of configurations, additions, and substitutions. Components not necessary for particular applications need not be included. For example, an operational forecasting configuration might include only components: 1, 6, 7, 8a, 8b, and 9, because the model input stack (Component 6) will be prepared in advance, and the calibration and evaluation components are not required for forecasting. Because the framework is model agnostic, models in the model library can discretize the land surface features described by a hydrofabric unit in any way, or not at all as in the case of catchment-based conceptual or machine learning models.

3.2 | Example Applications

The first step in setting up models in the NextGen framework is production of a HY_Features hydrofabric on the desired model domain. The hydrofabric tools provide the user with the ability to determine whether the focus is on catchment size distribution and channel reach lengths, or points of interest such as USGS stream gauge locations.

Figure 7 shows how the HY_Features data model allows execution of different model formulations simultaneously in the NextGen Framework at user desired resolutions. Starting from the original topographic data, catchment and stream delineation occur at using user-specified scale criteria. The National Hydrographic Dataset One definition might focus on points of interest such as stream gaging locations. Another might desire a catchment size distribution and/or minimum reach length. The reach length criterion is particularly useful to produce a computational stream network that is amenable to solution without requiring tiny timesteps as required by the Courant number.

The high-level design of the NextGen prototype provides flexibility in configuring and running model simulations. The simplest use case would involve setting up and running one hydrologic model in one basin without coupling to a channel routing module. A complex configuration would involve multiple linked models, including those of different paradigms (e.g., conceptual, physical, and machine learning), across multiple basins with a mixture of hydrologic and hydraulic routing modules. The three following example NextGen realizations illustrate how a user might take advantage of these capabilities.

In the example shown in Figure 8a, a user wishes to test/calibrate a snow ablation process module (Snow-17) in a single hydrofabric catchment by evaluating prediction of snow water equivalent (SWE). In this example, the NOAA Snow-17 temperature index model is linked to the Conceptual Functional Equivalent (CFE) to the National Water Model (described in: Araki et al. 2025), which provides partitioning and flow routing and a soil freeze/thaw (SFT) module to simulate frozen soil effects. The NextGen Framework calibration manager is used to optimize parameters applying a calibration objective function based on catchment average SWE.

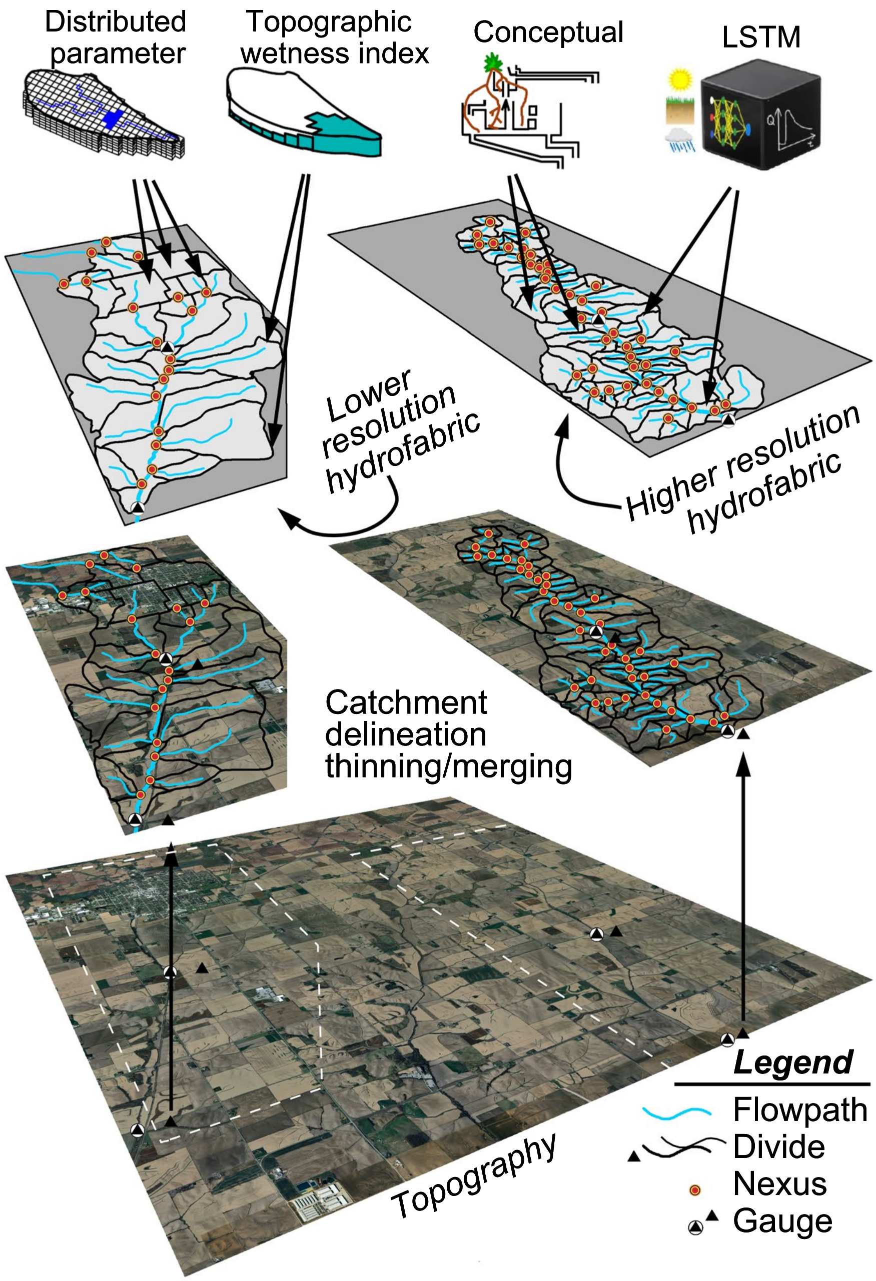

FIGURE 7 | Development of a hydrofabric that merges topographic information and user desired catchment sizes and nexus density for the application of different models.

In the second example shown in Figure 8b, another researcher wants to simulate more components of the hydrologic cycle over a larger spatial domain. They subset multiple contiguous basins from the NextGen hydrofabric with a USGS gage at the outlet point. They run a mixed configuration involving a variety of coupled modules in the various hydrofabric catchments and route the runoff output from each basin through the stream network using a hydrologic (Muskingum) routing module. To improve streamflow output, they deploy the ngen-cal calibration package and apply the calibrated model parameters after calibration.

In the third example shown in Figure 8c, water resources forecasters deploy a fully optimized model configuration over the continental US domain. They first subset multiple basins with USGS outlet gages from the NextGen hydrofabric and run ngen-cal in each larger basin grouping to calibrate model parameters for a large set of coupled models that may include land-surface routines such as Noah-OWP-Modular, Snow-17, and standard PET estimator, plus stormflow generation models such as CFE, TOPMODEL, HBV, and Sac-SMA. These catchment hydrologic models are coupled to the hydrofabric in a variety of ways. Hydrologic routing using the Muskingum-Cunge method is applied in low-order streams and rivers. In higher order streams, a diffusive-wave module allows simulation of backwater effects. Larger rivers are coupled to coastal models using a dynamic-wave channel routing module. Simulation outputs from different catchment hydrologic models are compared using multiple continuous, event, and categorical metrics against outputs from the previously trained long short-term memory (LSTM) model.

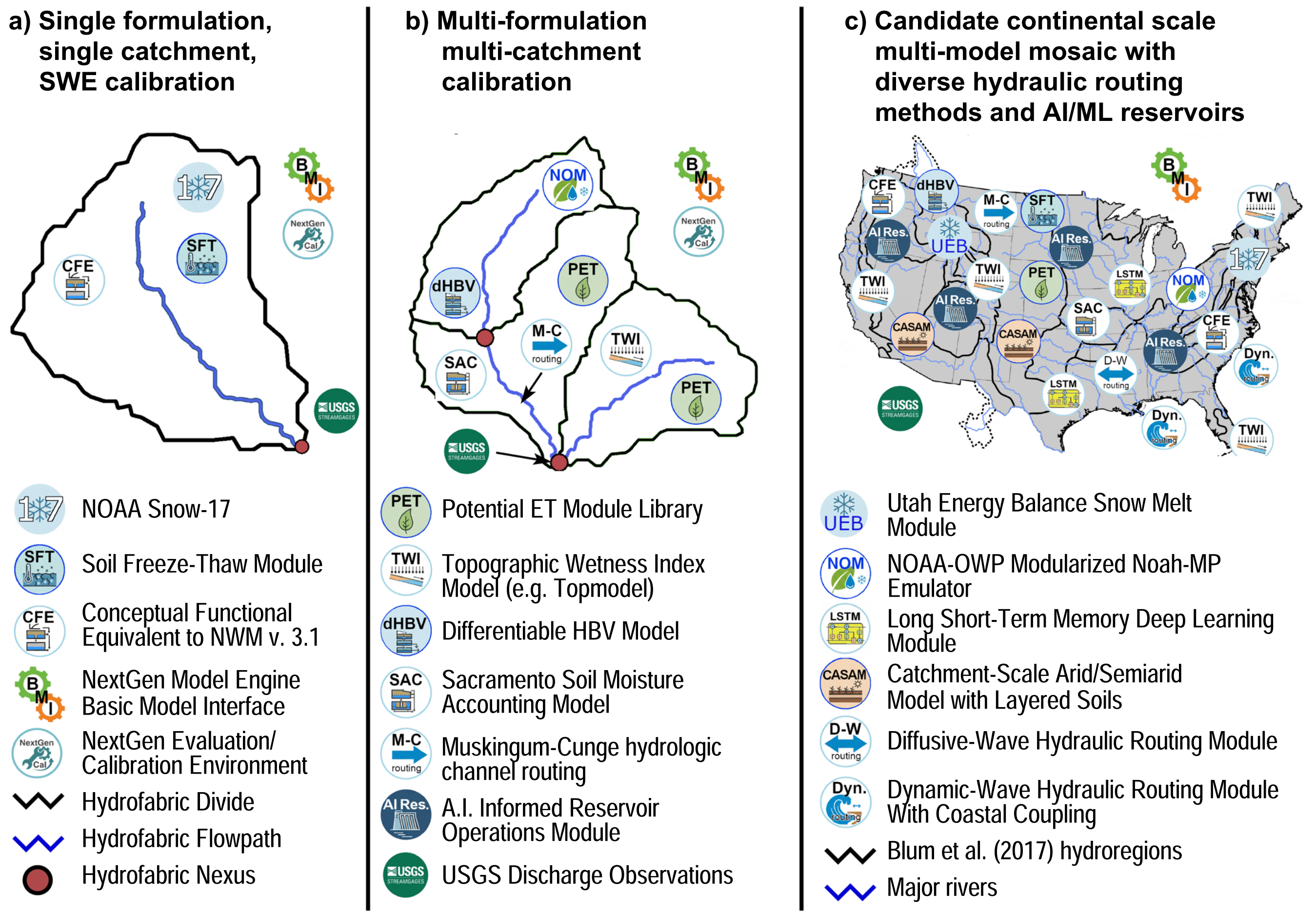

FIGURE 8 | Different use cases/configurations of the Next Generation Water Resources Modeling Framework showing increasing complexity. Panel (a) one hydrofabric catchment to calibrate one snow ablation model (e.g., Snow-17); (b) a coupled multi-model mosaic run across several hydrofabric catchments using hydrologic channel routing between hydrofabric nexuses; and (c) a variety of coupled models and modules running across the CONUS domain with either hydrologic or different hydraulic routing schemes depending on the importance of backwater effects and possibility of flow reversal.

This methodology might allow identification of the best performing model formulation on a regional basis.

4 | Conclusions

In the introduction to this paper, we detailed a series of challenges that researchers, developers, water managers, and operational forecasters all face when it comes to hydrologic prediction. In short, much of hydrology is concerned with hard-to-observe phenomena that necessitates computational models to overcome shortcomings in measurements, process understanding, and forecasting. Over time, the number of models, programming languages, and paradigms has grown, creating challenges in model selection and coupling, and evaluation of suitability and appropriateness. Facing this buffet of choices, researchers and practitioners often choose models based on familiarity instead of using the model that is most appropriate for their study domain and objective. To mitigate such heuristic issues, some researchers have proposed building community models that integrate modeling efforts. Unfortunately, these endeavors have largely failed as result of challenges in agreeing on allowed model structure, process representations, model domain discretization, input and output data requirements, parameter estimation, and calibration techniques, among others (e.g., getting researchers to agree on a final product).

In response, we propose a modular, standards-based, model-agnostic framework that works with existing models and standards so that researchers, students, and practitioners can experiment with a number of approaches and develop optimized solutions appropriate to their domains and objectives. We call this set of software tools the Next Generation Water Resources Modeling framework, or NextGen for short. NextGen applies an existing hydro-physiographic conceptual model, the OGC WaterML 2 HY_Features conceptual model, because it provides a means to represent relevant surface water hydrologic/hydraulic features at any desired resolution. This "hydrofabric" forms the basis of the development of standard hydro-feature descriptions. Model setup workflows that are compatible with this standard description offer a way forward and eliminate the need for new model developers to re-invent equivalent functionality.

The NextGen Framework also relies on the Basic Model Interface (BMI) model coupling standard to control model run-time and couple independently developed models written in different languages, such as C, C++, Fortran, and Python. The "B" in BMI, which stands for "Basic," means that the standard provides a minimum yet complete set of functions to allow for outside execution control, exchange of values using the widely available mechanisms that run on individual computers and HPC systems, and awareness of a wide variety of model discretizations. It also means that it represents a minimum set of required functionalities and helps to increase its appeal to the model development community. The BMI standard is model-agnostic. This allows the model developer to apply their solver and discretizations to the model domain. Extensions are possible. For example, the eXtended Model Interface extension (XMI), which allows tighter coupling between models/modules that require an iterative solution to improve simulation results (Hughes et al. 2022), provides a means to grow the capabilities of the framework as model coupling needs evolve.

Together, the use of the HY_Features and BMI standards enables several key NextGen capabilities, such as (1) the use of user-defined, not model-specific, geospatial data to define model domains, initialize parameter values and model states, and to build stream networks for flow routing, (2) the execution of various models and modules in the same framework using the same configuration system and execution commands, and (3) the production of model outputs in a standardized format over a consistent domain regardless of the chosen model(s) or module(s). We believe these NextGen capabilities promise to advance hydrologic and hydraulic modeling by providing a common operating environment that allows evaluation of different models.

The model-agnostic and standards-based Next Generation Water Resources Modeling Framework aims to maximize interoperability and promote scientific evaluation of different water prediction models in a common environment, leading to fair assessments and advancement of more capable models. The use of standards increases collaboration between the research community and federal agencies and speeds identification and operational implementation of superior modeling approaches. Enhanced interoperability means rapid coupling of verified domain-specific models in a way that allows users to take advantage of the expertise of others. Ultimately the benefits provided by such a standards-based modeling framework include an increased pace of discovery, more accurate and less redundant modeling as demonstrably superior models advance. It also supports the application of FAIR (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al. 2016) principles in water prediction.

Author Contributions

Fred L. Ogden: conceptualization, methodology, investigation, writing – original draft, writing – review and editing, formal analysis, visualization, validation. **Keith Jennings:** software, investigation, conceptualization, methodology, writing – original draft, writing – review and editing, validation, visualization. **Edward P. Clark:** conceptualization, methodology, supervision, writing – review and editing, project administration. **Ethan Coon:** conceptualization, methodology, writing – review and editing. **Brian Cosgrove:** conceptualization, methodology, supervision, project administration, writing – original draft, writing – review and editing. **Luciana Kindl da Cunha:** conceptualization, methodology, software, validation, investigation, writing – review and editing. **Matthew W. Farthing:** conceptualization, writing – review and editing. **Trey Flowers:** conceptualization, supervision, project administration. **Jonathan M. Frame:** conceptualization, methodology, software, investigation, validation, writing – review and editing. **Nels J. Frazier:** conceptualization, methodology, software, data curation, investigation, validation, writing – review and editing, writing – original draft. **Jessica L. Garrett:** conceptualization, methodology, software, data curation, investigation, validation, writing – review and editing, visualization. **Thomas M. Graziano:** conceptualization, methodology, supervision, project administration, writing – review and editing. **Joseph D. Hughes:** conceptualization. **J. Michael Johnson:** conceptualization, methodology, software, data curation, investigation, validation, visualization, writing – review and editing, writing – original draft. **Rachel McDaniel:** conceptualization, methodology, supervision, project administration, writing – review and editing, writing – original draft. **J. David Moulton:** conceptualization, methodology, writing – review and editing. **Scott D. Peckham:** writing – original draft, conceptualization, methodology, software, data curation, investigation, validation, writing – review and editing. **Fernando R. Salas:** conceptualization, investigation, validation, project administration, supervision. **Gaurav Savant:** conceptualization. **Roland Viger:** conceptualization, methodology, writing – review and editing. **Andy Wood:** conceptualization, methodology, investigation, validation, writing – review and editing.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The source code for the Next Generation Water Resources Modeling Framework, and the associated flood inundation mapping, hydrofabric processing, and hydrologic and hydraulic models and modules are open-source and available on GitHub: https://github.com/NOAA-OWP.

References

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.

Alaoui, A., M. Rogger, S. Peth, and G. Blöschl. 2018. "Does Soil Compaction Increase Floods? A Review." *Journal of Hydrology* 557: 631–642.

Ameli, A. A., J. R. Craig, and J. J. McDonnell. 2015. "Are All Runoff Processes the Same? Numerical Experiments Comparing a Darcy-Richards Solver to an Overland Flow-Based Approach for Subsurface Storm Runoff Simulation." *Water Resources Research* 51, no. 12: 10008–10028.

Araki, R., F. L. Ogden, and H. K. McMillan. 2025. "Testing Soil Moisture Performance Measures in the Conceptual-Functional Equivalent to the WRF-Hydro National Water Model." *Journal of the American Water Resources Association* 61: e70002. https://doi.org/10.1111/1752-1688.70002.

Baroni, G., B. Schalge, O. Rakovec, et al. 2019. "A Comprehensive Distributed Hydrological Modeling Intercomparison to Support Process Representation and Data Collection Strategies." *Water Resources Research* 55, no. 2: 990–1010. https://doi.org/10.1029/2018WR023941.

Baumhardt, R. L., G. L. Johnson, R. C. Schwartz, and D. K. Brauer. 2017. "Grazing and Tillage Effects on Soil Properties, Rain Infiltration, and Sediment Transport During Fallow." *Soil Science Society of America Journal* 81, no. 6: 1548–1556.

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. 2021. "The Era of Infiltration." *Hydrology and Earth System Sciences* 25, no. 2: 851–866. https://doi.org/10.5194/hess-25-851-2021.

Beven, K., and P. Germann. 1982. "Macropores and Water Flow in Soils." *Water Resources Research* 18, no. 5: 1311–1325.

Beven, K., and P. Germann. 2013. "Macropores and Water Flow in Soils Revisited." *Water Resources Research* 49, no. 6: 3071–3092.

Beven, K. J. 2000. "Uniqueness of Place and Process Representations in Hydrological Modelling." *Hydrology and Earth System Sciences* 4, no. 2: 203–213. https://doi.org/10.5194/hess-4-203-2000.

Beven, K. J. 2001. "How Far Can We Go in Distributed Hydrological Modelling?" *Hydrology and Earth System Sciences* 5, no. 1: 1–12.

Beven, K. J. 2005. "A New Paradigm for Environmental Predicting: Uncertainty, Complexity and Non-Linearity." *Hydrological Processes* 19, no. 3: 570–584.

Beven, K. J., and J. Freer. 2001. "Equifinality, Uncertainty and the Uniqueness of Place." *Hydrological Processes* 15, no. 3: 417–422.

Beven, K. J., and M. J. Kirkby. 1979. "A Physically-Based Variable Contributing Area Model of Basin Hydrology." *Hydrological Sciences Bulletin* 24: 43–69.

Beven, K. J., and P. Young. 2013. "A Guide to Good Practice in Modeling Semantics for Authors and Referees." *Water Resources Research* 49, no. 8: 5092–5098. https://doi.org/10.1002/wrcr.20393.

Bindas, T., W. P. Tsai, J. Liu, et al. 2024. "Improving River Routing Using a Differentiable Muskingum-Cunge Model and Physics-Informed Machine Learning." *Water Resources Research* 60, no. 1: e2023WR035337. https://doi.org/10.1029/2023WR035337.

Blodgett, D., and I. Dornblut. 2018. "OGC WaterML 2: Part 3-Surface Hydrology Features (HY_Features)-Conceptual Model. Version 1.0."

Blodgett, D., J. M. Johnson, and A. Bock. 2023. "Generating a Reference Flow Network With Improved Connectivity to Support Durable Data Integration and Reproducibility in the Coterminous US." *Environmental Modelling & Software* 165: 105726.

Blodgett, D., J. M. Johnson, M. Sondheim, M. Wieczorek, and N. J. Frazier. 2021. "Mainstems: A Logical Data Model Implementing Mainstem and Drainage Basin Feature Types Based on WaterML2 Part 3: HY Features Concepts." *Environmental Modelling & Software* 135: 104927.

Blöschl, G., M. F. P. Bierkens, A. Chambel, et al. 2019. "Twenty-Three Unsolved Problems in Hydrology (UPH)–A Community Perspective." *Hydrological Sciences Journal* 64, no. 10: 1141–1158.

BMI. 2023. "BMI Documentation (Version Stable)." https://bmi.readthedocs.io/en/stable/.

Bock, A. R., D. L. Blodgett, J. M. Johnson, M. Santiago, and M. E. Wieczorek. 2022. "PROVISIONAL: National Hydrologic Geospatial Fabric Reference and Derived Hydrofabrics: U.S. Geological Survey Data Release." https://doi.org/10.5066/P9NFPB5S.

Cai, X., Z.-L. Yang, Y. Xia, et al. 2014. "Assessment of Simulated Water Balance From Noah, Noah-MP, CLM, and VIC Over CONUS Using the NLDAS Test Bed." *Journal of Geophysical Research: Atmospheres* 119: 13751–13770. https://doi.org/10.1002/2014JD022113.

Cartwright, I., D. Cendón, M. Currell, and K. Meredith. 2017. "A Review of Radioactive Isotopes and Other Residence Time Tracers in Understanding Groundwater Recharge: Possibilities, Challenges, and Limitations." *Journal of Hydrology* 555: 797–811.

Cheng, Y., F. L. Ogden, and J. Zhu. 2017. "Earthworms and Tree Roots: A Model Study of the Effect of Preferential Flow Paths on Runoff Generation and Groundwater Recharge in Steep, Saprolitic, Tropical Lowland Catchments." *Water Resources Research* 53, no. 7: 5400–5419.

Chlumsky, R., J. Mai, J. R. Craig, and B. A. Tolson. 2021. "Simultaneous Calibration of Hydrologic Model Structure and Parameters Using a Blended Model." *Water Resources Research* 57, no. 5: e2020WR029229.

Ciarapica, L., and E. Todini. 2002. "TOPKAPI: A Model for the Representation of the Rainfall–Runoff Process at Different Scales." *Hydrological Processes* 16: 207–229.

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., J. Hendrikx, A. G. Slater, et al. 2011. "Representing Spatial Variability of Snow Water Equivalent in Hydrologic and Land-Surface Models: A Review." *Water Resources Research* 47, no. 7: W07539. https://doi.org/10.1029/2011WR010745.

Clark, M. P., and D. Kavetski. 2010. "Ancient Numerical Daemons of Conceptual Hydrological Modeling: 1. Fidelity and Efficiency of Time Stepping Schemes." *Water Resources Research* 46, no. 10. https://doi.org/10.1029/2009WR008894.

Clark, M. P., D. Kavetski, and F. Fenicia. 2011. "Pursuing the Method of Multiple Working Hypotheses for Hydrological Modeling." *Water Resources Research* 47, no. 9: WR009827. https://doi.org/10.1029/2010WR009827.

Clark, M. P., B. Nijssen, J. D. Lunquist, et al. 2015. "A Unified Approach for Process-Based Hydrologic Modeling: 1. Modeling Concept." *Water Resources Research* 51: 2498–2514. https://doi.org/10.1002/2015WR07198.

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. Vrugt, and H. V. Gupta. 2008. "Framework for Understanding Structural Errors (FUSE): A Modular Framework to Diagnose Differences Between Hydrological Models." *Water Resources Research* 44, no. 12: W00B02. https://doi.org/10.1029/2007WR006735.

Conor, M. L. 2024. "Implementation Guidance for Section 8106(a) of the Water Resources Development Act of 2022, Scope of Feasibility Studies." https://usace.contentdm.oclc.org/utils/getfile/collection/p16021coll5/id/37206.

Cook, P. G., and A. L. Herczeg. 2000. *Environmental Tracers in Subsurface Hydrology*. Springer Science+Business. https://doi.org/10.1007/978-1-4615-4557-6.

Coon, E. T., W. R. Elwasif, H. Pillai, P. E. Thornton, and S. L. Painter. 2019. "Exploring the Use of Novel Programming Models in Land Surface Models." In *2019 IEEE/ACM Parallel Applications Workshop, Alternatives to MPI (PAW-ATM)*, 1–10. IEEE.

Cosgrove, B., D. Gochis, T. Flowers, et al. 2024. "NOAA's National Water Model: Advancing Operational Hydrology Through Continental-Scale Modeling." *JAWRA Journal of the American Water Resources Association* 60: 247–272. https://doi.org/10.1111/1752-1688.13184.

Craig, A. P., M. Vertenstein, and R. Jacob. 2011. "A New Flexible Coupler for Earth System Modeling Developed for CCSM4 and CESM1." *International Journal of High Performance Computing Applications* 26: 31–42. https://doi.org/10.1177/1094342011428141.

Craig, J. R., G. Brown, R. Chlumsky, et al. 2020. "Flexible Watershed Simulation With the Raven Hydrological Modelling Framework." *Environmental Modelling and Software* 129: 104728. https://doi.org/10.1016/j.envsoft.2020.104728.

CSDMS. 2019. Accessed August 3, 2023. https://csdms.colorado.edu/wiki/Model_download_portal.

Cuthbert, M. O., and C. Tindimugaya. 2010. "The Importance of Preferential Flow in Controlling Groundwater Recharge in Tropical Africa and Implications for Modelling the Impact of Climate Change on Groundwater Resources." *Journal of Water and Climate Change* 1, no. 4: 234–245.

Dagum, L., and R. Menon. 1998. "OpenMP: An Industry Standard API for Shared-Memory Programming." *IEEE Computational Science and Engineering* 5, no. 1: 46–55.

David, O., J. C. Ascough II, W. Lloyd, et al. 2013. "A Software Engineering Perspective on Environmental Modeling Framework Design: The Object Modeling System." *Environmental Modelling & Software* 39: 201–213. https://doi.org/10.1016/j.envsoft.2012.03.006.

David, O., S. L. Markstrom, K. W. Rojas, L. R. Ahuja, and I. W. Schneider. 2002. "The Object Modeling System." In *Agricultural System Models in Field Research and Technology Transfer*, edited by L. Ahuja, L. Ma, and T. A. Howell, 317–331. Lewis Publishers, CRC Press LLC.

de La Fuente, L. A., M. R. Ehsani, H. V. Gupta, and L. E. Condon. 2023. "Towards Interpretable LSTM-Based Modelling of Hydrological Systems." *Hydrology and Earth System Sciences* 28: 945–971. https://doi.org/10.5194/hess-2023-252.

Dooge, J. C. I. 1986. "Looking for Hydrologic Laws." *Water Resources Research* 22, no. 9S: 46S–58S. https://doi.org/10.1029/WR022i09Sp0046S.

Dunkman, A., P. Burkholder, and K. Connolly. 2019. "NOAA Office of Water Prediction, National Water Model Path Analysis." 18F Group Report to NOAA/NWS Office of Water Prediction: 19.

Dunne, T., and R. D. Black. 1970. "An Experimental Investigation of Runoff Production in Permeable Soils." *Water Resources Research* 6, no. 2: 478–490. https://doi.org/10.1029/WR006i002p00478.

Essery, R., S. Morin, Y. Lejeune, and C. B Ménard. 2013. "A Comparison of 1701 Snow Models Using Observations From an Alpine Site." *Advances in Water Resources* 55: 131–148. https://doi.org/10.1016/j.advwatres.2012.07.013.

Evenson, E. J., S. A. Jones, N. L. Barber, et al. 2018. "Continuing Progress Toward a National Assessment of Water Availability and Use (No. 1440)." US Geological Survey. https://doi.org/10.3133/cir1440.

Falcone, J. 2011. "GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow: U.S. Geological Survey Data Release." https://doi.org/10.5066/P96CPHOT.

Fatichi, S., E. R. Vivoni, F. L. Ogden, et al. 2016. "An Overview of Current Applications, Challenges, and Future Trends in Distributed Process-Based Models in Hydrology." *Journal of Hydrology* 537: 45–60.

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.

Fenicia, F., D. Kavetski, and H. H. Savenije. 2011. "Elements of a Flexible Approach for Conceptual Hydrological Modeling: 1. Motivation and Theoretical Development." *Water Resources Research* 47, no. 11. https://doi.org/10.1029/2010WR010174.

Fiedler, F. R., and J. A. Ramirez. 2000. "A Numerical Method for Simulating Discontinuous Shallow Flow Over an Infiltrating Surface." *International Journal for Numerical Methods in Fluids* 32, no. 2: 219–239.

Frame, J. M., F. Kratzert, H. V. Gupta, P. Ullrich, and G. S. Nearing. 2023. "On Strictly Enforced Mass Conservation Constraints for Modeling the Rainfall-Runoff Process." *Hydrological Processes* 37: e14847. https://doi.org/10.31223/X5BH0P.

Frame, J. M., F. Kratzert, D. Klotz, et al. 2022. "Deep Learning Rainfall–Runoff Predictions of Extreme Events." *Hydrology and Earth System Sciences* 26, no. 13: 3377–3392. https://doi.org/10.5194/hess-26-3377-2022.

Gan, T., G. E. Tucker, E. W. Hutton, et al. 2023. "CSDMS Data Components: Data-Model Integration Tools for Earth Surface Processes Modeling." *Geoscientific Model Development Discussions* 2023: 1–31.

Gauch, M., F. Kratzert, D. Klotz, G. Nearing, J. Lin, and S. Hochreiter. 2021. "Rainfall-Runoff Prediction at Multiple Timescales With a Single Long Short-Term Memory Network." *Hydrology and Earth System Sciences* 25, no. 4: 2045–2062. https://doi.org/10.5194/hess-25-2045-2021.

Gauch, M., J. Mai, and J. Lin. 2021. "The Proper Care and Feeding of CAMELS: How Limited Training Data Affects Streamflow Prediction." *Environmental Modelling and Software* 135: 104926. https://doi.org/10.1016/j.envsoft.2020.104926.

GDAL/OGR contributors. 2022. *GDAL/OGR Geospatial Data Abstraction*. Software Library, Open Source Geospatial Foundation. https://doi.org/10.5281/zenodo.5884351.

Gochis, D. J., M. Barlage, R. Cabell, et al. 2020. *The WRF-Hydro Modeling System Technical Description (Version 5.1.1)*, 107. NCAR Technical Note. https://doi.org/10.5281/zenodo.3625238.

Godsey, S. E., and J. W. Kirchner. 2014. "Dynamic, Discontinuous Stream Networks: Hydrologically Driven Variations in Active Drainage Density, Flowing Channels and Stream Order." *Hydrological Processes* 28, no. 23: 5791–5803.

Hill, C., C. DeLuca, M. Suarez, and A. Da Silva. 2004. "The Architecture of the Earth System Modeling Framework." *Computing in Science & Engineering* 6, no. 1: 18–28.

Horton, P., B. Schaefli, and M. Kauzlaric. 2022. "Why Do We Have So Many Different Hydrological Models? A Review Based on the Case of Switzerland." *WIREs Water* 9, no. 1: e1574. https://doi.org/10.1002/wat2.1574.

Horton, R. E. 1933. "The Role of Infiltration in the Hydrologic Cycle." *Transactions - American Geophysical Union* 14: 446–460.

Hughes, J. D., M. J. Russcher, C. D. Langevin, E. D. Morway, and R. R. McDonald. 2022. "The MODFLOW Application Programming Interface for Simulation Control and Software Interoperability." *Environmental Modelling & Software* 148: 105257.

Hutton, E. W., M. D. Piper, and G. E. Tucker. 2020. "The Basic Model Interface 2.0: A Standard Interface for Coupling Numerical Models in the Geosciences." *Journal of Open Source Software* 5, no. 51: 2317.

Jakeman, A. J., and G. M. Hornberger. 1993. "How Much Complexity Is Warranted in a Rainfall-Runoff Model?" *Water Resources Research* 29, no. 8: 2637–2649.

Johnson, J. M. 2022. "National Hydrologic Geospatial Fabric (Hydrofabric) for the Next Generation (NextGen) Hydrologic Modeling Framework." HydroShare. http://www.hydroshare.org/resource/129787b468aa4d55ace7b124ed27dbde.

Johnson, J. M., D. Munasinghe, D. Eyelade, and S. Cohen. 2019. "An Integrated Evaluation of the National Water Model (NWM)–Height Above Nearest Drainage (HAND) Flood Mapping Methodology." *Natural Hazards and Earth System Sciences* 19, no. 11: 2405–2420.

Kendall, C., and J. J. McDonnell. 2012. *Isotope Tracers in Catchment Hydrology*. Elsevier.

Keyes, D. E., L. C. McInnes, C. Woodward, et al. 2013. "Multiphysics Simulations: Challenges and Opportunities." *International Journal of High Performance Computing Applications* 27, no. 1: 4–83. https://doi.org/10.1177/1094342012468181.

Kirchner, J. W. 2006. "Getting the Right Answers for the Right Reasons: Linking Measurements, Analyses, and Models to Advance the Science of Hydrology." *Water Resources Research* 42, no. 3: W03S04.

Kirchner, J. W. 2009. "Catchments as Simple Dynamical Systems: Catchment Characterization, Rainfall-Runoff Modeling, and Doing Hydrology Backward." *Water Resources Research* 45, no. 2: WR006912. https://doi.org/10.1029/2008WR006912.

Knoben, W. J., J. E. Freer, K. J. Fowler, M. C. Peel, and R. A. Woods. 2019. "Modular Assessment of Rainfall–Runoff Models Toolbox (MARRMoT) v1. 2: An Open-Source, Extendable Framework Providing Implementations of 46 Conceptual Hydrologic Models as Continuous State-Space Formulations." *Geoscientific Model Development* 12, no. 6: 2463–2480. https://doi.org/10.5194/gmd-12-2463-2019.

Knoben, W. J. M., A. Raman, G. J. Gründemann, et al. 2025. "How Many Models Do We Need to Simulate Hydrologic Processes Across Large Geographical Domains?" *Hydrology and Earth System Sciences* 29, no. 11: 2361–2375. https://doi.org/10.5194/hess-29-2361-2025.

Kollet, S., M. Sulis, R. M. Maxwell, et al. 2017. "The Integrated Hydrologic Model Intercomparison Project, IH-MIP2: A Second Set of Benchmark Results to Diagnose Integrated Hydrology and Feedbacks." *Water Resources Research* 53, no. 1: 867–890.

Kratzert, F., D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger. 2018. "Rainfall–Runoff Modelling Using Long Short-Term Memory (LSTM) Networks." *Hydrology and Earth System Sciences* 22, no. 11: 6005–6022.

Kratzert, F., D. Klotz, M. Herrnegger, A. K. Sampson, S. Hochreiter, and G. S. Nearing. 2019. "Towards Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning." *Water Resources Research* 55: 2019WR026065. https://doi.org/10.1029/2019WR026065.

Krinner, G., C. Derksen, R. Essery, et al. 2018. "ESM-SnowMIP: Assessing Snow Models and Quantifying Snow-Related Climate Feedbacks." *Geoscientific Model Development* 11, no. 12: 5027–5049. https://doi.org/10.5194/gmd-11-5027-2018.

Kunkle, G. R. 1965. *Computation of Ground-Water Discharge to Streams During Floods, or to Individual Reaches During Base Flow, by Use of Specific Conductance*, D207–D210. United States Geological Survey Professional Paper 525-D.

Leavesley, G. H., S. L. Markstrom, P. J. Restrepo, and R. J. Viger. 2002. "A Modular Approach to Addressing Model Design, Scale, and Parameter Estimation Issues in Distributed Hydrological Modelling." *Hydrological Processes* 16, no. 2: 173–187. https://doi.org/10.1002/hyp.344.

López-Moreno, J. I., S. R. Fassnacht, J. T. Heath, et al. 2013. "Small Scale Spatial Variability of Snow Density and Depth Over Complex Alpine Terrain: Implications for Estimating Snow Water Equivalent." *Advances in Water Resources* 55: 40–52. https://doi.org/10.1016/j.advwatres.2012.08.010.

Lott, D. A., and M. T. Stewart. 2013. "A Power Function Method for Estimating Base Flow." *Groundwater* 51, no. 3: 442–451. https://doi.org/10.1111/j.1745-6584.2012.00980.x.

Luukkonen, C. L., A. H. Alzraiee, D. M. Herbert, et al. 2025. "Harmonization of a Water Withdrawal Dataset for the Conterminous United States." *JAWRA Journal of the American Water Resources Association* 61, no. 6: e70054. https://doi.org/10.1111/1752-1688.70054.

Maidment, D. R. 2017. "Conceptual Framework for the National Flood Interoperability Experiment." *JAWRA Journal of the American Water Resources Association* 53, no. 2: 245–257. https://doi.org/10.1111/1752-1688.12474.

Marks, D., J. Kimball, D. Tingey, and T. Link. 1998. "The Sensitivity of Snowmelt Processes to Climate Conditions and Forest Cover During Rain-on-Snow: A Case Study of the 1996 Pacific Northwest Flood." *Hydrological Processes* 12, no. 10: 1569–1587.

Maxwell, R. M., M. Putti, S. Meyerhoff, et al. 2014. "Surface-Subsurface Model Intercomparison: A First Set of Benchmark Results to Diagnose Integrated Hydrology and Feedbacks." *Water Resources Research* 50, no. 2: 1531–1549. https://doi.org/10.1002/2013WR013725.

McDonnell, J. J., K. McGuire, P. Aggarwal, et al. 2010. "How Old Is Streamwater? Open Questions in Catchment Transit Time Conceptualization, Modelling and Analysis." *Hydrological Processes* 24, no. 12: 1745–1754. https://doi.org/10.1002/hyp.7796.

McDonnell, J. J., M. Sivapalan, K. Vaché, et al. 2007. "Moving Beyond Heterogeneity and Process Complexity: A New Vision for Watershed Hydrology." *Water Resources Research* 43, no. 7: WR005467. https://doi.org/10.1029/2006WR005467.

McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Rea. 2012. *NHDPlus Version 2: User Guide*, 745. US Environmental Protection Agency.

McMahon, T. A., V. K. Gupta, and V. P. Singh. 2006. "Using a Parsimonious Conceptual Model for Improving Hydrologic Predictions." *Water Resources Research* 42, no. 5: W05413.

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.

Medalie, L., A. E. Galanter, A. J. Martinez, et al. 2025. "Water Use Across the Conterminous United States, Water Years 2010–20 (No. 1894-D)." US Geological Survey. https://doi.org/10.3133/pp1894D.

Miller, C. T., C. N. Dawson, M. W. Farthing, et al. 2013. "Numerical Simulation of Water Resources Problems: Models, Methods, and Trends." *Advances in Water Resources* 51: 405–437. https://doi.org/10.1016/j.advwatres.2012.05.008.

Modaresi Rad, A., J. M. Johnson, Z. Ghahremani, J. Coll, and N. Frazier. 2024. "Enhancing River Channel Dimension Estimation: A Machine Learning Approach Leveraging the National Water Model, Hydrographic Networks, and Landscape Characteristics." *Journal of Geophysical Research: Machine Learning and Computation* 1, no. 4: e2024JH000173.

Montanari, A., and D. Koutsoyiannis. 2012. "A Blueprint for Process-Based Modeling of Uncertain Hydrological Systems." *Water Resources Research* 48: W09555. https://doi.org/10.1029/2011WR011412.

Montgomery, D. R., W. E. Dietrich, R. Torres, S. P. Anderson, J. T. Heffner, and K. Loague. 1997. "Hydrologic Response of a Steep, Unchanneled Valley to Natural and Applied Rainfall." *Water Resources Research* 33, no. 1: 91–109.

Mulholland, B., and M. A. Fullen. 1991. "Cattle Trampling and Soil Compaction on Loamy Sands." *Soil Use and Management* 7: 189–193.

National Research Council. 2012. *Weather Services for the Nation: Becoming Second to None*. National Academies Press. https://doi.org/10.17226/13429.

Nativi, S., P. Mazzetti, and G. N. Geller. 2013. "Environmental Model Access and Interoperability: The GEO Model Web Initiative." *Environmental Modelling & Software* 39: 214–228.

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., D. Klotz, A. K. Sampson, et al. 2021. "Technical Note: Data Assimilation and Autoregression for Using Near-Real-Time Streamflow Observations in Long Short-Term Memory Networks." *Hydrology and Earth System Sciences Discussions* 26: 5493–5513.

NOAA. 2025. "Research Transitions, Applications for Funded Research." Accessed January 14, 2025. https://wpo.noaa.gov/research-transitions/.

Ogden, F., B. Avant, R. Bartel, et al. 2021. *The Next Generation Water Resources Modeling Framework: Open Source, Standards Based, Community Accessible, Model Interoperability for Large Scale Water Prediction*. Vol. 2021, H43D-01. AGU Fall Meeting Abstracts. https://ui.adsabs.harvard.edu/abs/2021AGUFM.H43D.01O/abstract.

Ogden, F., N. Frazier, K. Jennings, et al. 2024. *Technical Guidance for the Inclusion of Models/Modules in the NextGen Water Resources Modeling Framework*. NOAA Cooperative Institute for Research to Operations, Working Group 2, Technical Report.

Ogden, F. L. 2021. "Geohydrology: Hydrological Modeling." In *Encyclopedia of Geology*, edited by D. Alderton and S. A. Elias, vol. 6, 2nd ed., 457–476. Academic Press.

Peckham, S. D., C. DeLuca, D. J. Gochis, et al. 2014. *EarthCube-Earth System Bridge: Spanning Scientific Communities With Interoperable Modeling Frameworks*. Vol. 2014, IN31D-3754. AGU Fall Meeting Abstracts.

Peckham, S. D., E. W. Hutton, and B. Norris. 2013. “A Component-Based Approach to Integrated Modeling in the Geosciences: The Design of CSDMS.” *Computers & Geosciences* 53: 3–12. https://doi.org/10.1016/j.cageo.2012.04.002.

Pellerin, B. A., W. M. Wollheim, X. Feng, and C. J. Vorosmarty. 2008. “The Application of Electrical Conductivity as a Tracer for Hydrograph Separation in Urban Catchments.” *Hydrological Processes* 22, no. 12: 1810–1818. https://doi.org/10.1002/hyp.6786.

Perrin, C., C. Michel, and V. Andréassian. 2001. “Does a Large Number of Parameters Enhance Model Performance? Comparative Assessment of Common Catchment Model Structures on 429 Catchments.” *Journal of Hydrology* 242, no. 3–4: 275–301.

Phillips, J. D., and C. Lorz. 2008. “Origins and Implications of Soil Layering.” *Earth-Science Reviews* 89, no. 3–4: 144–155.

Pinder, G. F., and J. F. Jones. 1969. “Determination of the Ground-Water Component of Peak Discharge From the Chemistry of Total Runoff.” *Water Resources Research* 5, no. 2: 438–445. https://doi.org/10.1029/WR005i002p00438.

Polyakov, V., J. Stone, C. Holifield Collins, et al. 2018. “Rainfall Simulation Experiments in the Southwestern USA Using the Walnut Gulch Rainfall Simulator.” *Earth System Science Data* 10, no. 1: 19–26.

Pomeroy, J. W., D. M. Gray, T. Brown, et al. 2007. “The Cold Regions Hydrological Process Representation and Model: A Platform for Basing Model Structure on Physical Evidence.” *Hydrological Processes* 21, no. 19: 2650–2667. https://doi.org/10.1002/hyp.6787.

Pradhan, N. R., and F. L. Ogden. 2010. “Development of a One-Parameter Variable Source Area Runoff Model for Ungauged Basins.” *Advances in Water Resources* 33, no. 5: 572–584.

Rafieeinasab, A., A. Mazrooei, T. Enzminger, et al. 2024. “A WRF-Hydro-Based Retrospective Simulation of Water Resources for US Integrated Water Availability Assessment.” *Hydrology and Earth System Sciences Discussions* 2024: 1–37. https://doi.org/10.5194/hess-2024-262.

Reed, S., V. Koren, M. Smith, et al. 2004. “Overall Distributed Model Intercomparison Project Results.” *Journal of Hydrology* 298, no. 1: 27–60. https://doi.org/10.1016/j.jhydrol.2004.03.031.

Refsgaard, J. C., J. Knudsen, and H. J. Henriksen. 1996. “Modelling Guidelines—Terminology and Guiding Principles.” *Journal of Hydrology* 173, no. 1–4: 523–539.

Regan, R. S., S. L. Markstrom, L. E. Hay, et al. 2018. “Description of the National Hydrologic Model for use with the Precipitation-Runoff Modeling System (PRMS).” U.S. Geological Survey Techniques and Methods, Book 6, Chapter B9, 38 p. https://doi.org/10.3133/tm6B9.

Rice, K. C., and G. M. Hornberger. 1998. “Comparison of Hydrochemical Tracers to Estimate Source Contributions to Peak Flow in a Small, Forested, Headwater Catchment.” *Water Resources Research* 34: 1755–1766. https://doi.org/10.1029/98WR00917.

Roberts, J. M. 2007. “Representing the Uniqueness of Place in Hydrological Models Using Bayesian Networks.” *Journal of Hydrology* 343, no. 1–4: 135–150.

Robinson, D. A., A. Binley, N. Crook, et al. 2008. “Advancing Process-Based Watershed Hydrological Research Using Near-Surface Geophysics: A Vision for, and Review of, Electrical and Magnetic Geophysical Methods.” *Hydrological Processes* 22, no. 18: 3604–3635.

Rössler, O., P. Froidevaux, U. Börst, R. Rickli, O. Martius, and R. Weingartner. 2014. “Retrospective Analysis of a Nonforecasted Rain-on-Snow Flood in the Alps—A Matter of Model Limitations or Unpredictable Nature?” *Hydrology and Earth System Sciences* 18, no. 6: 2265–2285. https://doi.org/10.5194/hess-18-2265-2014.

Rutter, N., R. Essery, J. Pomeroy, et al. 2009. “Evaluation of Forest Snow Processes Models (SnowMIP2).” *Journal of Geophysical Research: Atmospheres* 114: JD011063. https://doi.org/10.1029/2008JD011063.

Scaife, C. I., N. K. Singh, R. E. Emanuel, C. F. Miniat, and L. E. Band. 2020. “Non-Linear Quickflow Response as Indicators of Runoff Generation Mechanisms.” *Hydrological Processes* 34, no. 13: 2949–2964.

Sexstone, G. A., D. W. Clow, D. I. Stannard, and S. R. Fassnacht. 2016. “Comparison of Methods for Quantifying Surface Sublimation Over Seasonally Snow-Covered Terrain.” *Hydrological Processes* 30, no. 19: 3373–3389. https://doi.org/10.1002/hyp.10864.

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: 552–567. https://doi.org/10.1038/s43017-023-00450-9.

Singh, V. P., and D. A. Woolhiser. 2002. “The Uniqueness Problem in Conceptual Hydrology.” *Hydrological Processes* 16, no. 10: 2243–2259.

Sjöberg, Y., A. Jan, S. L. Painter, et al. 2021. “Permafrost Promotes Shallow Groundwater Flow and Warmer Headwater Streams.” *Water Resources Research* 57, no. 2: e2020WR027463.

Sklash, M. G., and R. N. Farvolden. 1979. “The Role of Groundwater in Storm Runoff.” *Developments in Water Science* 12: e65. https://doi.org/10.1016/S0167-5648(09)70009-7.

Smith, M. B., V. Koren, Z. Zhang, et al. 2012. “Results of the DMIP 2 Oklahoma Experiments.” *Journal of Hydrology* 418: 17–48.

Smith, M. B., D. J. Seo, V. I. Koren, et al. 2004. “The Distributed Model Intercomparison Project (DMIP): Motivation and Experiment Design.” *Journal of Hydrology* 298, no. 1–4: 4–26.

Spieler, D., J. Mai, J. R. Craig, B. A. Tolson, and N. Schütze. 2020. “Automatic Model Structure Identification for Conceptual Hydrologic Models.” *Water Resources Research* 56, no. 9: e2019WR027009. https://doi.org/10.1029/2019WR027009.

Stedinger, J. R., and D. P. Lettenmaier. 1996. “Hydrologic Model Calibration.” In *Hydrology: Processes, Resources and Management*, edited by J. C. Rodda, 181–200. Chapman and Hall.

Steenhuis, T. S., M. Winchell, J. Rossing, J. A. Zollweg, and M. F. Walter. 1995. “SCS Runoff Equation Revisited for Variable-Source Runoff Areas.” *Journal of Irrigation and Drainage Engineering* 121, no. 3: 234–238.

Stewart, R. D., M. R. Abou Najm, D. E. Rupp, and J. S. Selker. 2016. “Modeling Multidomain Hydraulic Properties of Shrink-Swell Soils.” *Water Resources Research* 52, no. 10: 7911–7930.

Tegegne, G., D. K. Park, and Y.-O. Kim. 2017. “Comparison of Hydrological Models for the Assessment of Water Resources in a Data-Scarce Region, the Upper Blue Nile River Basin.” *Journal of Hydrology: Regional Studies* 14: 49–66. https://doi.org/10.1016/j.ejrh.2017.10.002.

Tijerina, D., L. Condon, K. FitzGerald, et al. 2021. “Continental Hydrologic Intercomparison Project, Phase 1: A Large-Scale Hydrologic Model Comparison Over the Continental United States.” *Water Resources Research* 57, no. 7: e2020WR028931.

Tucker, G. E., E. W. Hutton, M. D. Piper, et al. 2022. “CSDMS: A Community Platform for Numerical Modeling of Earth Surface Processes.” *Geoscientific Model Development* 15, no. 4: 1413–1439.

US Public Law 117-58. 2021. “Title II—Commerce, Justice, Science, And Related Agencies.” https://www.congress.gov/117/plaws/publ58/PLAW-117publ58.pdf.

USGS. 2025. "National Water Availability Assessment Data Companion." Accessed January 18, 2025. https://www.usgs.gov/special-topics/integrated-water-availability-assessments/national-water-availability-assessment-0.

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian. 2001. "A Framework for Development and Application of Hydrological Models." *Hydrology and Earth System Sciences* 5, no. 1: 13–26. https://doi.org/10.5194/hess-5-13-2001.

Wang, K., and R. E. Dickinson. 2012. "A Review of Global Terrestrial Evapotranspiration: Observation, Modeling, Climatology, and Climatic Variability." *Reviews of Geophysics* 50, no. 2: RG2005. https://doi.org/10.1029/2011RG000373.

Wang, M., E. Hamidi, and D. Mccay. 2023. "Chapter 2: On Numerical Methods and Differentiable Modeling for Soil Process Representations in the Nextgen Framework in Arid Regions." In *National Water Center Innovators Program Summer Institute Report 2023*. CUAHSI Hydroshare. https://www.hydroshare.org/resource/709de9e2bc5640b281c7f022cecc593d/.

Weiler, M., and K. Beven. 2015. "Do We Need a Community Hydrological Model?" *Water Resources Research* 51, no. 9: 7777–7784. https://doi.org/10.1002/2014WR016731.

Werner, M., J. Schellekens, R. Gijsbers, M. van den Dijk, O. Akker, and K. Heynert. 2013. "The Delft-FEWS Flow Forecasting System." *Environmental Modelling & Software* 40: 65–77. https://doi.org/10.1016/j.envsoft.2012.07.010.

Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3: 160018. https://doi.org/10.1038/SDATA.2016.18.

Xia, Y., K. Mitchell, M. Ek, et al. 2012. "Continental-Scale Water and Energy Flux Analysis and Validation for the North American Land Data Assimilation System Project Phase 2 (NLDAS-2): 1. Intercomparison and Application of Model Products." *Journal of Geophysical Research. Atmospheres* 117: JD016048.

Xu, T., and F. Liang. 2021. "Machine Learning for Hydrologic Sciences: An Introductory Overview." *WIREs Water* 8, no. 5: e1533.

Zhang, Y., Z. Zhang, Z. Ma, et al. 2018. "A Review of Preferential Water Flow in Soil Science." *Canadian Journal of Soil Science* 98, no. 4: 604–618.