Hydrologic process inference in large-scale models under human impacts

1 Motivation and Research Objectives

In October 2022, and for the first time in over 20 years, the U.S. Geological Survey released a new water cycle diagram featuring humans at its core (Duncombe, 2022). This is a significant step forward, as it demonstrates that the human domination of the global hydrologic cycle has finally started to permeate the public perception (Abbott et al., 2019). The representation of human actions in large-scale hydrologic studies has followed a similar pathway: in the early 1990s, scientists started to compare water availability and use at the continental and global scales (e.g., Falkenmark, 1989, 1997; Shiklomanov, 1997), but only later took on the challenge of modelling the impact of human actions on terrestrial water fluxes (Nazemi & Wheater, 2015a; Wada et al., 2017). The growing extent of the human footprint has further emphasized this need (Vörösmarty et al., 2003; Döll et al., 2009; Haddeland et al., 2014). To date, there are three complementary categories of models representing human-water interactions across large spatial domains (from multi-basin to global scale): land surface models, hydrologic models, and dynamic vegetation models (Bierkens, 2015). Hereafter, we refer to them as large-scale hydrologic models.

As outlined in a number of reviews (Nazemi & Wheater, 2015a, 2015b; Pokhrel et al., 2016; Wada et al., 2017), the representation of human actions in large-scale hydrologic models has typically focused on withdrawals for irrigative and non-irrigative demand as well as the operations of three main types of water infrastructure: reservoirs and artificial lakes, streamflow diversions, and groundwater wells. Among these, water reservoirs play a paramount role simply because of their preponderance and impact on river discharge—in the continental U.S. alone, for instance, there are more than 52,000 dams providing storage capacity for ∼75% of the mean annual runoff (Steyaert et al., 2022). Yet, dam operations are still represented rather simplistically in hydrologic models (Pokhrel et al., 2016) with a number of fragmented efforts (Telteu et al., 2021), leading to a lack of standards that further complicates model evaluations and inter-model comparisons. This fragmentation has several causes, including lack of accurate data, coarse spatial resolution (∼10-100 km), and the intrinsic complexity of human actions (Simon, 1957). There is also a legacy issue: large-scale hydrologic models were originally conceived to study the behavior of natural systems, so updating them to account for human actions is an effort that comes with substantial theoretical and computational challenges. Finally, the field of large-scale hydrology has yet to integrate the advances of other disciplines that develop computational models of human behaviour—*in primis*, water resources systems analysis.

Recent research efforts have underscored the importance of a correct representation of dams: Hodgkins et al., 2023, for example, appraised multiple national-scale hydrologic models (none of which explicitly account for reservoir storage) and found that their accuracy decreases with increasing reservoir storage. Similar findings were reported by Ghimire et al., 2023 when analyzing the new Dayflow dataset (simulated river discharge in ∼2.7 million reaches in the CONUS). However, the issue is deeper than model accuracy alone: the misrepresentation of dam operations may also affect the parameterization of hydrologic processes (Dang et al., 2020a). We thus hypothesize that the representation of reservoirs in large-scale hydrologic models is a major source of structural uncertainty that confounds the inference of hydrologic processes. This overarching hypothesis leads us to the following research questions:

1. What is the structural uncertainty associated with different models of reservoir release and storage? How does such uncertainty vary across dams (with different design features and operating objectives) and river basins (with different topographic characteristics and hydro-meteorological conditions)?

2. How does the uncertainty in the representation of dam operations impact the parameterization of large-scale hydrologic models? What are the conditions under which issues of *equifinality* are amplified?
3. How do these forms of uncertainty affect the representation of hydrologic processes (e.g., infiltration and subsurface flow, evaporation and transpiration) and fluxes of water (e.g., in soils, aquifers, and streams)?
4. What are the steps that hydrologists should take to characterize this uncertainty and, where possible, curb it? How do these steps vary with the study site at hand?

To answer these questions, we propose to build on the convergence of four research domains, namely remote sensing, large-scale hydrology, catchment hydrology, and water resources systems analysis. By combining elements from these domains, we will develop and test a computational framework that (1) characterizes the uncertainty associated with the adoption of different reservoir regulation models, (2) explains how such structural uncertainty propagates to the parameterization of large-scale hydrologic models, and (3) depicts the impact of these forms of uncertainty on the simulation of hydrologic processes. The framework will be applied to 18 large basins within the CONUS encompassing a broad spectrum of environmental and reservoir regulation conditions, on the basis of which we will (4) devise best modelling practices for the representation of dams in large-scale hydrologic models. Our findings will finally be upscaled by extending the framework to the Colorado and Columbia river basins.

There are multiple reasons why this is a timely moment for tackling the proposed research. First, the past decade has highlighted the challenge of implementing hydrologic and land surface models at much higher spatial resolution (from $\sim$1 km to $\sim$100 m), often referred to as the "hyper-resolution" (Wood et al., 2011). This effort has started to yield the first results (e.g., O'neill et al., 2021; Hanasaki et al., 2022; Hoch et al., 2023), and provides further impetus to model human impacts at the same resolution. Second, large-scale hydrologic models increasingly serve a broad spectrum of downstream modelling applications, all requiring an accurate representation of dam operations. Examples include streamflow forecasts (Yuan et al., 2013; Greuell et al., 2018), generation of reanalysis datasets (Shin et al., 2019), ecological impact assessments (Galelli et al., 2022), or studies of the interaction between multiple socio-economic sectors (Kanyako et al., 2023). In this regard, it should be stressed that several streamflow re-analysis datasets available at the CONUS scale were generated by models that do not explictly account for the presence of dams and their operations (Livneh et al., 2013; Ghimire et al., 2023; Johnson et al., 2023). Third, recent advances in remote sensing and data collection have started to provide information on dam storage and release with unprecedented granularity (Steyaert et al., 2022; Hou et al., 2022). There remains a need for a transdisciplinary effort to exploit these advances, with the ultimate goal of improving both prediction accuracy and process inference in large-scale hydrologic models.

2 Background

The proposed work draws on three related lines of literature: (1) remote sensing to create observational datasets of flow regulation; (2) representation of reservoir regulation in hydrologic models; (3) diagnostic evaluation of hydrologic models. Here, we provide a brief overview of the state-of-the-art in each area and highlight the knowledge gaps that are functional to our research questions.

2.1 Remote sensing to create observational datasets of flow regulation

The chief infrastructure supporting flow regulation and diversion are dams, which impound 600 km^3 of water in the US and 8,000 km^3 globally (Grill et al., 2019; Steyaert et al., 2022). To date, we have fairly detailed information on their location, technical specifications, and, at times, operating purposes (e.g., flood control, hydropower production), all encapsulated in databases commonly used by the large-scale hydrology community. Examples include the Global Reservoir and Dam Database (GRanD), the World Register of Dams (WRD), the Global Dam Tracker (GDAT) (Zhang & Gu, 2023), and the Global River Obstruction Database (GROD) (X. Yang et al., 2022). Importantly, these databases do not provide timeseries data on reservoir inflow, storage, and release, which are needed to assess the differential impact of each dam in a river basin. In turn, this explains why large-scale hydrologic models typically rely on routing schemes that estimate reservoir operational decisions on the basis of the only information globally available, such as operating purposes or storage capacity (Section 2.2). This raises key questions about how accurate these generic rules are and the impact of their inclusion on process inference and model reliability.

Advances in remote sensing allow us to circumvent this issue by monitoring dynamic reservoir level and storage. Specifically, time series of water level are created by directly measuring surface water elevations using radar and laser altimeters (Gao, 2015), which, importantly, are not affected by clouds and other disturbances. However, water level observations are available only for a small number of reservoirs (at the global scale) and span, at best, the last two decades (e.g., Birkett et al., 2011; Schwatke et al., 2015). Storage variations over time can be estimated by combining (1) remotely-sensed elevation (from radars or altimeters) with area (from imagery), or (2) remotely-sensed area (or elevation) with information on reservoir bathymetry (Gao, 2015). Approaches based on satellite imagery are particularly appealing because images are available for virtually any reservoir over the past four decades, though the area estimates must be adjusted for clouds (Pekel et al., 2016). Relying on the data generated by different missions (e.g., Landsat, Jason, Sentinel), several recent studies have produced databases of reservoir storage at either global (e.g., Busker et al., 2019; Biswas et al., 2021; Hou et al., 2022) or regional scale (e.g., Bonnema & Hossain, 2017; Vu et al., 2022; Shen et al., 2022). Yet, two major research gaps still need to be addressed. First, remotely-sensed data on reservoir operations differ in terms of temporal coverage, spatio-temporal resolution, accuracy, and monitored phenomena, so we lack a synergistic approach to consolidate the existing data (as well as the methods used to generate them) over specific regions of interest. Second, the validation of remotely-sensed data is typically challenged by the lack of in situ data—an issue often explained by lack of capacity, poor coordination among agencies, or secrecy. Researchers focusing on the U.S. are uniquely positioned to tackle these gaps, since databases including operational records have been recently released (e.g., Steyaert et al., 2022). This project will create a database of dam storage and release for all 18 large basins of interest (Section 3). We will first consolidate existing data (observed and inferred) and then complement them with additional information retrieved from remotely-sensed products (Section 4.1).

2.2 Representation of reservoir regulation in hydrologic models

The incorporation of reservoir regulation into large-scale hydrologic models has been achieved in several ways, yet remains challenged by the tradeoff between accuracy and transferability. This mirrors a broader and long-recognized tradeoff in modelling natural hydrologic processes (e.g., Bastidas et al., 2006; Hogue et al., 2006). Ideally, a model representation will also be interpretable in the context of available theory (Khatami et al., 2019; Knüsel et al., 2019). Most approaches to modelling reservoir regulation center around the concept of a *control policy*—a functional relationship mapping observed conditions to release decisions. Typically (though not always), the most accurate approach for a specific river basin is to model the reservoir control policies defined by the operating agencies, where such documentation exists. However, such rules are not only poorly documented but also, by definition, not transferable across study sites, which prevents their use for hydrologic process inference. The state-of-the-art can thus be divided into three macro categories of reservoir control policies: generic, data-driven, and optimization-based.

Generic control policies. Large-scale hydrologic models, such as PCR-GLOBWB 2 (Sutanudjaja et al., 2018), WaterGAP 2 (Müller Schmied et al., 2021), MOSART-WM (Voisin et al., 2017), and LEAF-Hydro-Flood (Shin et al., 2019), among others, employ generic operational schemes to determine reservoir releases. While details vary, all approaches rely on estimates of water demand to define seasonal releases, stemming from the foundational work of Hanasaki et al., 2006 and Haddeland et al., 2006. Generic control policies are highly transferable, but contain several limitations. First, the release rules typically contain free parameters, which are either calibrated against observations of release and / or storage (when available) or set uniformly to avoid relying on local observations (Turner et al., 2020). Second, they require global estimates of water demand at a relatively coarse spatial resolution, which introduces additional uncertainty (Masaki et al., 2017). Additionally, the focus on seasonal releases for water supply neglects the important flood control operations of large dams, which occur on shorter timescales but are critical for hydrologic prediction (Yassin et al., 2019). (See Nazemi and Wheater, 2015b; Pokhrel et al., 2016; Wada et al., 2017; Veldkamp et al., 2018 for additional details regarding the advantages and limitations of generic control policies.) However, the widespread use of generic approaches highlights the value of transferability and the potential for these rules to benefit from local parameterizations.

Data-driven control policies. In contrast to generic schemes, reservoir control policies can also be inferred directly from observed storage and release time series, where available. This approach relates to the problem described above of calibrating the parameters of generic control policies (e.g., Zhao et al., 2016; Yassin et al., 2019; Turner et al., 2020; Dang et al., 2020a; Tefs et al., 2021), but may also adopt a fully empirical approach where the control policy is represented by a data-driven model, such as an artificial neural network (e.g., Ehsani et al., 2016; Coerver et al., 2018; Ouyang et al., 2021). The former offers a lower risk of overfitting, and a stronger interpretability for the parameters (Turner et al., 2020; Dang et al., 2020a), which can also be supported by sensitivity analysis and other diagnostic tools to ensure physical plausibility (Pianosi et al., 2016; Gupta & Razavi, 2018; Dang et al., 2020b). Data-driven methods are especially promising because they provide basin-scale accuracy with a general functional form, without the need to model ad hoc rules for each reservoir. However, in order for data-driven control policies to generalize in ungauged areas, the at-site calibrations must be regionalized by understanding the relationship between control policy parameters and basin-scale features.

Optimization-based. The last category stems from water resources systems analysis and builds on the assumption that the decision-making problem of determining dam release can be modelled by formulating and solving an optimal control problem. In other words, dam operators are modelled as rational decision-makers whose target is to optimize the value of one or multiple objectives, such as hydropower production or flood control (Turner & Galelli, 2016). This approach has found successful application in a number of recent studies (e.g., Ng et al., 2017; Cáceres et al., 2022; Abeshu et al., 2023), and is particularly advantageous when the modeller has a clear understanding of the objectives, or targets, characterizing the reservoir operations. The other advantage is that this approach does not necessarily require data on storage and release, since only information on dam design specifications and downstream demand (if any) are necessary to formulate the optimal control problem. Like all other approaches, also this one suffers from a few limitations. First, solving optimization problems is easy when dealing with a single reservoir, but computational requirements quickly increase with the dimensionality of the problem at hand (Giuliani et al., 2021). Second, operating objectives are not constant across time, as they are affected by a variety of drivers, such as changes in legislation and downstream demand or the construction of new infrastructure (Vu et al., 2022). Where historical release data are available, it is possible to infer a target release pattern to meet via optimization without calibrating a control policy or assuming an economic objective function downstream.

Overall, the diversity in the approaches here described suggests that there could be substantial structural uncertainty in the reservoir operations simulated by hydrologic models (Masaki et al., 2017). This project will characterize this uncertainty as well as its cascading impact on the simulation of hydrologic processes and water fluxes.

2.3 Calibration of large-scale hydrologic models with reservoirs

The calibration of large-scale hydrologic models is a challenging exercise, owing to the number of free parameters involved, the lack of adequate data, and the computational requirements that arise when coupling models with the sophisticated calibration methods developed by the catchment hydrology community (e.g., Duan et al., 1992; Vrugt et al., 2003; Tang et al., 2006). Because of these reasons, large-scale hydrologic models are often uncalibrated, while their implementations on a smaller number of basins is usually performed using only observed discharge (Greve et al., 2020). We argue that the impact of dam operations on hydrologic processes could further convolute the calibration process. One potential pitfall is represented by parameter interactions and *equifinality* (Beven, 1993): for example, parameters in one part of the model may be assigned unrealistic values to compensate for a poor representation of water infrastructures, leading us to the 'right' results for the wrong reasons (Kirchner, 2006; Clark et al., 2021).

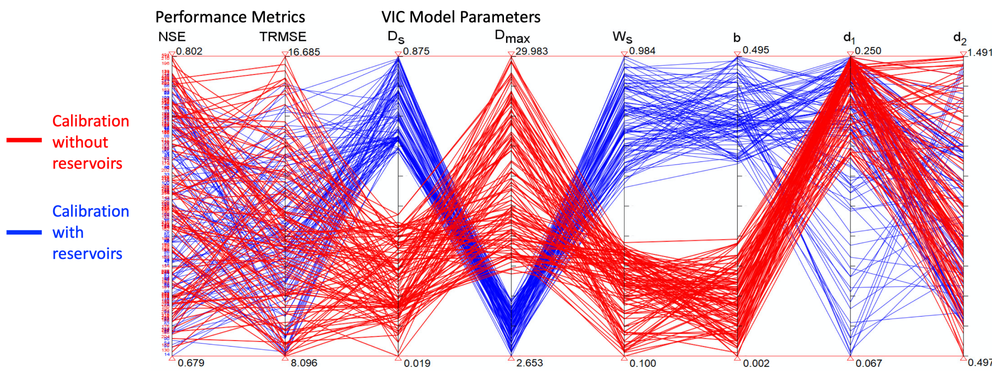

Figure 1. Parallel coordinate plot illustrating the values of two performance metrics (Nash–Sutcliffe Efficiency and Transformed Root Mean Square Error) and six parameters of the VIC model (D_s, D_{max}, W_s, b, d_1 and d_2) obtained by calibrating the model for the Upper Mekong River basin. Each line connecting the axes represents a parameterization, along with the corresponding model performance. Blue and red lines denote parameterizations obtained with and without reservoirs. Adapted from Dang et al., 2020a.

This challenge is exemplified by a recent study from the PI's group (Dang et al., 2020a, Figure 1), who set up two instances of the Variable Infiltration Capacity model (Liang et al., 2014) for the Upper Mekong River basin. One instance was coupled with the Lohmann's routing scheme (conceived for pristine catchments; Lohmann et al. (1996, 1998)), while the other adopted a variant of the routing scheme to explicitly represent reservoir operations for 12 reservoirs in the basin. As shown in Figure 1, the calibration exercise showed that both models can achieve the same accuracy in reproducing river discharge at the catchment outlet (metrics NSE and TRMSE). Yet, the model without reservoirs does so through 'optimal' soil parameters that compensate for the structural error of neglecting dams by 'creating' a thicker soil layer (parameter d_1) with higher storage capacity and generation of base flow in the lower layer (parameters D_s, D_{max}, W_s, and b)—ultimately biasing the representation of surface runoff, infiltration, and base flow. The example underscores the need to analyze how hydrologic process representation and model calibration are conceived in regions heavily affected by human interventions (Müller Schmied et al., 2014).

To date, we still have limited understanding of how different approaches to reservoir operations modelling influence the simulation of terrestrial water fluxes and storages across space and time. There is therefore a need for diagnostic approaches to analyze and target a variety of matters, such as structural uncertainty, flawed parameterizations, and equifinality. We speculate that such explorations are particularly challenging when dealing with models requiring several free parameters to characterise human actions, which often leads to a higher degree of uncertainty.

3 Study sites

We propose to focus on a spatial domain that strikes a reasonable balance between project feasibility and the need of working on a domain that (1) is sufficiently large and thus (2) carries variability in dam design features and operating objectives as well as topographic and hydro-meteorological conditions. The planned data collection, modelling, and synthesis efforts aim to inform future expansion to all basins within the CONUS, along with the ∼52,000 dams that punctuate them.

We select 18 large river basins according to the following procedure. Starting from the GAGES-II dataset (Falcone, 2011), basins were filtered to identify sites that have a complete record over the period 1990-2009 and that remained active beyond 2009. Next, the largest basin in each HUC2 region was selected to obtain a sample of basins with broad variability in hydro-meteorological conditions. Note that most of the 18 selected basins, shown in Figure 2, are larger than 30,000 km^2 and represent portions of some of the most intensively managed basins in the U.S.—the Colorado, Columbia, Sacramento, and Apalachicola-Chattahoochee-Flint, among others. We also note that the gage records will be extended through 2020 in this project; if the updated data are not available, the basins shown in Figure 2 will be slightly modified using a nearby gage.

Gage ID	Name	Area (km2)	# Reservoirs in GRaND	% Capacity in ResOpsUS
1184000	CONNECTICUT RIVER AT THOMPSONVILLE, CT.	25049	19	56
1554000	SUSQUEHANNA RIVER AT SUNBURY, PA	47364	10	0
2359170	APALACHICOLA RIVER NR SUMATRA,FLA.	49802	12	90
3341500	WABASH RIVER AT TERRE HAUTE, IND.	31881	7	82
3566000	HIWASSEE RIVER AT CHARLESTON, TN	5970	6	100
4193500	MAUMEE RIVER AT WATERVILLE OH	16409	4	0
6477000	JAMES R NEAR FORESTBURG SD	47157	7	71
5054000	RED RIVER OF THE NORTH AT FARGO, ND	17196	17	14
5330000	MINNESOTA RIVER NEAR JORDAN, MN	42041	10	38
6877600	SMOKY HILL R AT ENTERPRISE, KS	49592	6	100
7227500	CANADIAN RV NR AMARILLO, TX	49264	3	99
8066500	TRINITY RV AT ROMAYOR, TX	44412	27	92
8396500	PECOS RIVER NEAR ARTESIA, NM	46492	3	97
9163500	COLORADO RIVER NEAR CO-UT STATE LINE	46274	17	100
9469500	GILA RIVER BELOW COOLIDGE DAM, AZ.	33546	1	100
10335000	HUMBOLDT RV NR RYE PATCH, NV	41271	10	0
11390500	SACRAMENTO R NR GRIMES CA	36188	16	94
13081500	SNAKE R NR MINIDOKA ID	48830	9	91

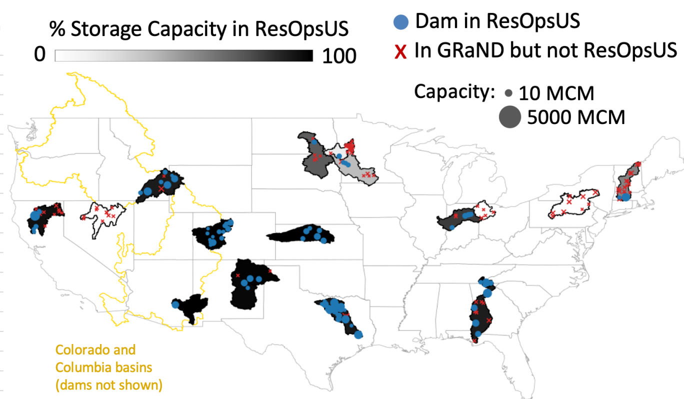

Figure 2. Selected basins from GAGES-II (Falcone, 2011), including the corresponding drainage area, number of dams, and percentage of reservoir capacity included in the ResOpsUS database. Each basin represents the gage with the largest basin area in each HUC2 region with a complete data record from 1990-2009 that remained active in 2009. The map illustrates the spatial distribution of these basins, along with the large dams.

The reservoirs in each selected basin are defined by the GRaND database as having storage larger than 10 million m^3 (MCM) and/or height larger than 15 meters (Lehner et al., 2011). As shown in Figure 2, the number of large reservoirs in each basin ranges from 1 (Gila) to 27 (Trinity). Many of these are also included in the ResOpsUS dataset (Steyaert et al., 2022), which contains daily records of inflow, release, and storage over the period 1980-2020. Importantly, the basins have excellent coverage in this dataset (above 90% of storage capacity); however, there are also some basins, such as the Susquehanna, Maumee, and Humboldt, with 0% of storage capacity included in ResOpsUS. This poses a regionalization challenge that will be the focus of Task 1 and 2 (Section 4.1 and 4.2).

Building on these 18 basins and 184 large dams, we will study how the uncertainty in the representation of dam operations affects hydrologic process inference. However, focusing on these basins only may limit our ability to fully appreciate the impact of human action representation in large-scale hydrologic models as well as their downstream applications. For this reason, we will conclude the project by expanding the proposed framework to the Colorado and Columbia river basins. By doing so, we will upscale the findings from the initial 18 basins to two full HUC2 regions, demonstrate the impact on large-scale hydrology, and provide concrete examples of how ill-posed process inference may lead to inaccurate modelling applications (e.g., overestimation of drought propagation under future climate forcings). In sum, our project will consider 18 basins and 184 dams in the first stage, and then two basins and 200 dams in the second stage (116 for Columbia, 84 for Colorado). In all of these locations, but particularly in the Western U.S., hydrologic process representation is critical for projecting water supply and flood risk in a changing climate, and for informing the storage decisions that make process inference difficult in the first place.

4 Proposed Research

The proposed research is organized into four tasks. The first one is aimed at laying the foundations for this work, namely creating a dataset of reservoir storage, inflow, and release for all dams within our study site. In the second and third tasks, we will characterize the uncertainty in the representation of dam operations (**Research Question 1**) and then study how such uncertainty impacts process inference and accuracy in large-scale hydrologic models (**Research Questions 2-3**). In the fourth task, we will derive modelling guidelines and upscale our findings to the Columbia and Colorado basins (**Research Question 4**).

4.1 Expanding observational datasets of flow regulation (Task 1)

The first challenge in our project stands in the creation of a dataset containing time series of daily records of inflow, release, and storage over the period 1990-2020. As explained in Section 4.2, each category of dam operation models (i.e., generic, data-driven, and optimization-based) has different data requirements; creating a comprehensive dataset containing these three variables will thus guarantee that any dam operation model can be setup, calibrated, and validated. The starting point for this data gathering exercise will be ResOpsUS dataset (Steyaert et al., 2022). However, over half of the GRaND reservoirs in our study sites are not included in ResOpsUS. This type of data gap represents a common obstacle to hydrologic modelling in human-dominated basins. To fill in this gap, we will proceed by leveraging remotely sensed data to infer reservoir storage. As for inflow and release, we will rely on existing observational records, combined with the estimated storage to infer missing components of the mass balance.

Beginning with storage, existing datasets provide multi-decadal time series of storage at the global scale (Section 2.1). Key examples are the datasets released by Hou et al., 2022 (6,695 reservoirs, 1984-2015) and Li et al., 2023 (7,245 reservoirs, 1999-2018), which rely on water surface data inferred from Landsat images and area-storage relationships (or hypsometric curves) retrieved from global databases and models (Messager et al., 2016; Li et al., 2020; Yigzaw et al., 2018). Setting for a moment aside the mismatch between our temporal domain (1990-2020) and the one provided by these datasets, the main challenge we foresee here stands in the accuracy of these datasets. Because global storage datasets have been used to characterize variability and trends in dam storage across the globe, validation efforts are typically limited to a few dams. Inaccuracies in storage estimates are thus not considered critical, as they are (rightfully so) likely not to affect research conclusions. However, such inaccuracies may be critical in hydrologic modelling applications, particularly in basins that are heavily affected by dam operations (Otta et al., 2023; Vu et al., 2023). A case in point is provided in Figure 3, where we compare observed and inferred storage for Shasta Reservoir (from Steyaert et al., 2022 and Li et al., 2023, respectively): the comparison shows that the time series created by analyzing satellite images correctly tracks the observed variability in storage, but also underestimates storage (by about 1 km^3); a mismatch that may be due to either the estimates of water surface or the hypsometric curves.

To address these issues, we propose to improve the accuracy of storage estimates by focusing on the two major modelling steps, namely the estimation of (1) water surface and (2) hypsometric curves. For the first aspect, the PI's group has recently introduced a new algorithm that improves our ability to remove the effect of clouds and other disturbances from Landsat images (Vu et al., 2022). Moreover, the availability of storage observations (from ResOpsUS) for about half of the dams within the study site will give us a great opportunity to properly calibrate the cloud removal and image processing algorithm—something that is prevented in ungauged basins, where these tools are typically applied. As for the hypsometric curves, we will begin by comparing the curves for each reservoir provided by all global databases (Messager et al., 2016; Yigzaw et al., 2018; Li et al., 2020; Khazaei et al., 2022; Hao et al., 2024). Reservoirs with a high degree of inconsistency across databases will be manually inspected. The hypsometric curves will be corrected by combining time series of area (retrieved from satellite images), water level (e.g., Birkett et al., 2011, Schwatke et al., 2015), and the specific topographic characteristics of the reservoir site. We note that this endeavour will be further supported by concurrent observations of water surface and level provided by the SWOT mission (Biancamaria et al., 2016). Preliminary results illustrating the potential of the proposed approach are reported in Figure 3.

The collection of inflow and release data is a somewhat simpler task, which will rely on the collection of time series from USGS at the nearest gages upstream/downstream of each reservoir. Gages are not likely to be available near all dams; for these cases, we will estimate the inflow (release) by combining observed release (inflow) with storage data. All time series so obtained will be inspected to identify outliers, and post-processed to the same daily time step.

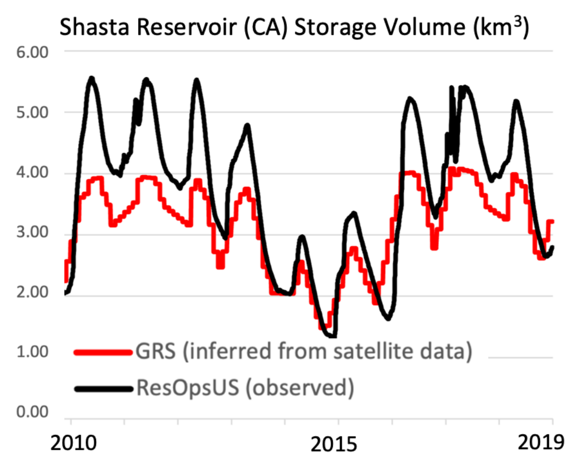

Figure 3. Comparison between observed (black line) and inferred storage from remote sensing (red line) for Shasta Reservoir, CA. Observed storage data (with daily resolution) were retrieved from ResOpsUS (Steyaert et al., 2022), while inferred data (with monthly resolution) were retrieved from Li et al., 2023. The green line represents the storage inferred using the methodology introduced by Vu et al., 2022.

4.2 Characterizing the uncertainty in the representation of dam operations (Task 2)

Task 1 will complete the dataset of storage, inflow, and release at all reservoirs in the study basins. Task 2 will leverage this dataset to study the uncertainty associated with dam operation models (**Research Question 1**). We will answer this question using the following steps. First, we will calibrate reservoir release policies in three categories: generic, data-driven, and optimization-based. The data-driven policies will directly use the observed and estimated storage, release, and inflow data from Task 1. The generic and optimization categories do not require observed release data, but instead require estimates of downstream demand, including monthly demand for irrigation. We will obtain these estimates for the chosen basins from the GAGES-II dataset (Falcone, 2011), supplemented by a monthly dataset of irrigation withdrawals at the HUC12 scale (Haynes et al., 2023)—recognizing that the choice of demand dataset represents an additional source of uncertainty.

The policies we plan to test are listed in Table 1. Generic policies assume a release rule based on downstream demand data and the current (simulated) storage, following Hanasaki et al., 2006. The approaches vary based on objective priority, use of natural versus impaired flow for deriving the releases, and use of consumptive use versus withdrawals for representing water demand (Voisin et al., 2013). The data-driven approaches allow fitting a policy to observed release and storage, such as the dataset we plan to create in Task 1. The structure of these rules varies from a parameterized rule (Turner et al., 2021) to more complex machine learning models (Ehsani et al., 2016; Coerver et al., 2018), but in all cases the parameters are fitted using observed releases and storage. Last, the optimization approaches assume objective functions and their priority. For example, Haddeland 2006 develop objective functions for irrigation, flood control, hydropower, and water supply, based on the operating purpose of the reservoir. Reservoirs with more than one operating purpose are assumed to prioritize irrigation. The model assumes 12-month perfect foresight to optimize the releases, which is improved by optimization strategies that account for future uncertainty, such as stochastic dynamic programming (SDP) and direct policy search (DPS) (e.g., Turner et al., 2017; Giuliani et al., 2016, respectively).

Generic	**Data-Driven**	**Optimization**
H06 (Haddeland et al., 2006)	Parameterized Rule (Turner et al., 2021)	Perfect Foresight (Haddeland et al., 2006)
H06 with Irrigation (Döll et al., 2009; Biemans et al., 2011; Voisin et al., 2013)	Decision Tree (T. Yang et al., 2016; J. D. Herman & Giuliani, 2018)	Stochastic Dynamic Programming (Turner et al., 2017)
Natural Lakes benchmark (Döll et al., 2003)	Neural Network (Ehsani et al., 2016; Coerver et al., 2018)	Direct Policy Search (Giuliani et al., 2016)

Table 1. Alternative model structures for reservoir regulation.

Each of the policy structures in Table 1 (9 total) will be calibrated and validated for all reservoirs in the study basins, leveraging the datasets created in Task 1. The split between calibration and validation data will be challenged by the fact that the policies themselves may have changed during the period 1990-2020. Therefore, we will use a 5-fold cross-validation to test on successive 6-year periods held out of calibration. The accuracy will be tested by comparing the modeled releases and storage to the observed values (or estimated values from Task 1). The models will be developed at a daily timestep; the generic reservoir policies designed for monthly timesteps will be downscaled by assuming a daily average release.

The structural uncertainty in reservoir regulation will be characterized by the ensemble of policies resulting from this step, after removing policies that perform poorly out-of-sample (e.g., KGE $\leq 1-\sqrt(2)$, Knoben et al., 2019). This ensemble will represent an empirical distribution of plausible predictions. We theorize that these alternative reservoir regulation schemes, when embedded in a hydrologic model, will result in significantly different calibration of hydrologic parameters and processes (Task 3).

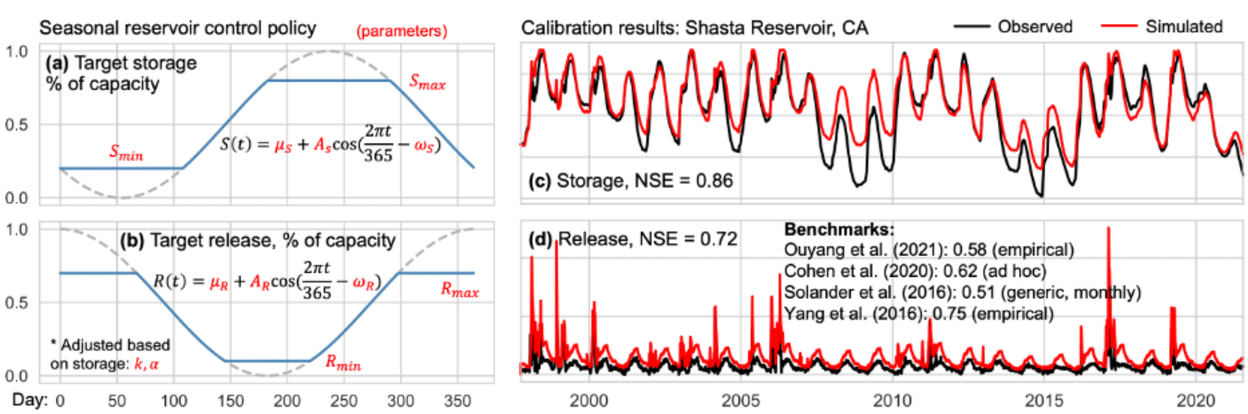

Figure 4. a-b) Target storage and release rules, with 12 parameters to calibrate. (c-d) Preliminary calibration results for Shasta Reservoir, California. The release NSE compares favorably to available benchmarks.

A preliminary result for a data-driven parameterized rule is shown in Figure 4 for Shasta Reservoir in the Sacramento River basin (Figure 2). Reservoir releases are determined by a 12-parameter rule similar to Turner et al. 2021. When storage exceeds the target, a fraction k of the excess is added to the release; if storage is less than the target, the release is multiplied by the hedging rule $(S/S_{\text{target}})^{\alpha}$. The calibration is performed with Differential Evolution (Storn & Price, 1997). The accuracy of simulated releases compares favorably to recent benchmarks from a combination of empirical and generic models for this specific reservoir, though clearly shows the difficulty of capturing flood control releases on a daily timescale. Yang et al. 2016 calibrated to 3 years of data, which may explain the slight improvement in the reported NSE as it does not account for sampling uncertainty (Clark et al., 2021).

4.3 Understanding the cascading impacts on large-scale hydrologic models (Task 3)

In Task 3, we will integrate the dam operation models into the hydrologic model, and then study how the uncertainty in the representation of reservoir operations impact hydrologic model parameterization and process inference (**Research Questions 2-3**).

4.3.1 Large-scale hydrologic model

To simulate the spatio-temporal variability of surface water, we will use a large-scale, semi-distributed, hydrologic model and coupled it with a flow routing model. The first modelling component is VIC, or Variable Infiltration Capacity (Liang et al., 1994), which simulates the water and energy fluxes governing the terrestrial water cycle on a spatial grid. In particular, the model divides each grid cell into one vegetation and two (or three) soil layers, for which it calculates evapotranspiration, infiltration, runoff, and base flow at a daily or sub-daily time step. The base flow and runoff from VIC will be processed with a flow routing model developed by the PI's Lab, VIC-Res (Dang et al., 2020b). VIC-Res builds on the routing model introduced by Lohmann et al., 1996, 1998 (with routing based on a flow direction matrix and a linearized version of the Saint-Venant equations) and augments its capabilities by including an explicit representation of dam operations. This is achieved by implementing a number of cells in which the storage dynamics are calculated, and then adopting control rules that determine the release (as detailed in Section 2.2 and 4.2).

There are four main reasons behind the choice of these models. First, VIC is one of the most widely-adopted large-scale hydrologic models at the catchment, regional, and global scale (see Melsen et al., 2016, Schaperow et al., 2021, and references therein) and, importantly, has been used to create streamflow re-analysis products at the global and CONUS scale (Maurer et al., 2002; Livneh et al., 2013; Lin et al., 2019; Ghimire et al., 2023). Second, Galelli's group has been leading the development of VIC-Res, which has found successful applications in a variety of regions, including Southeast Asia (Galelli et al., 2022), India (Vegad & Mishra, 2022), and Southern Africa (Chowdhury et al., 2022). Third, VIC-Res is integrated with multiple libraries for sensitivity analysis and optimization (Dang et al., 2020b), which will support the activities described next. Fourth, these models can also account for other forms of water management interventions, namely groundwater withdrawals and irrigation. In this regard, we note that these human interventions are not the focus of this proposal; that being said, major interventions in the study basins will be explicitly accounted for using the approach described by Eldardiry et al., 2022.

4.3.2 Model setup

VIC requires several input datasets that characterize meteorological forcing as well as vegetation, land cover, soil, and elevation features. During the past two decades, VIC's user base has produced a number of such datasets at both global (e.g., Nijssen et al., 2001; Adam et al., 2009; Lin et al., 2019; Schaperow et al., 2021) and CONUS-scale (e.g., Maurer et al., 2002; Livneh et al., 2013; Oubeidillah et al., 2014; Bohn and Vivoni, 2019). Among these different alternatives, we plan to use the input data prepared by Oubeidillah et al., 2014, since they were recently adopted to generate the most recent streamflow reanalysis dataset for the CONUS (Ghimire et al., 2023). Specifically, the input dataset includes:

- MODIS leaf area index (1-km resolution; Myneni et al., 2015);
- Land Cover Classification (1-km resolution; Hansen et al., 2000);
- Soil information retrieved from CONUS-SOIL (1-km resolution; Miller and White, 1998);
- USGS National Elevation Dataset (10-m resolution; USGS, 2018);
- Daymet meteorological forcing (Thornton et al., 2021) (1-km resolution from 1980-present), which includes daily precipitation, wind speed, and maximum and minimum temperature.

As in Oubeidillah et al., 2014 and Ghimire et al., 2023, these data will be processed to a spatial resolution of 1/24° (∼4 km) for simulations with VIC and, subsequently, with VIC-Res. Since our study includes 18 basins (Section 3), we will setup 18 independent instances of the model, all for the simulation period 1990-2020.

The VIC-Res routing model requires base flow and runoff data from VIC, a flow direction map, and the design specifications of each reservoir. The flow direction map will be generated by processing digital elevation models (e.g., USGS, 2018). The reservoir design specifications (e.g., year of commission, total storage capacity, dead storage capacity) will be retrieved from the the Global Reservoir and Dam Database (GRanD). VIC-Res also includes the hypsometric curves and reservoir control policy of each dam, which will be developed in Task 1 and 2, respectively. In particular, Task 2 will provide nine alternative model structures for reservoir regulation, all of which will be implemented in VIC-Res. To these, we will add one model implementation that excludes the representation of water reservoirs, thereby following the approach of Ghimire et al., 2023 as well as other studies that generated other recent streamflow re-analysis products (Lin et al., 2019). This means that we will setup 180 instances of VIC-Res (18 basins × 10 model structures).

4.3.3 Process inference and diagnostic evaluation

This part of Task 3 will test the overarching hypothesis of this project by analyzing how different representations of dam operations impacts process inference in large-scale hydrologic modelling. This leads us to two specific experiments.

Sensitivity analysis. The existing body of literature on VIC calibration and sensitivity analysis gives us two elements worth considering: one one hand, it is common practice to calibrate VIC with respect to ∼5-6 parameters that are 'often' found to control the model ability to simulate streamflow (e.g., Dan et al., 2012; Xue et al., 2015; Wi et al., 2017; Dang et al., 2020a); on the other hand, we know that the sensitivity is influenced by multiple factors, such as the vegetation type, soil characteristics, or hydro-climatic conditions of the basin (Demaria et al., 2007). Because of this second matter, we plan to begin this task by setting up a sensitivity analysis experiment for each of the 180 VIC + VIC-Res instances described above. Inspired by previous studies (e.g., Demaria et al., 2007; Chaney et al., 2015), we will consider ∼10 key parameters (e.g., base flow parameters, thickness of the soil layers, saturated hydraulic conductivities, and infiltration shape parameter), and explore the sensitivity of simulated river discharge to these parameters. We will consider 1990-2010 as simulation period, with 1990-1991 used for model spin-up (and thus excluded from the analysis). Observed river discharge will be retrieved from the stations described in Section 3. Model performance will be quantified by multiple error metrics to capture the model ability to simulate different nuances of river discharge, such as accuracy during low flow or high flow periods, or the ability to match the long-term variability of flows (Galelli et al., 2022). All experiments will be carried out linking VIC-Res with the open-source SALib package for global sensitivity analysis developed by co-PI Herman (J. Herman & Usher, 2017). Overall, this experiment will show how the sensitivity of the hydrologic model parameters and associated equifinality vary with changes in the structure of the reservoir operating policy (**Research Question 2**).

Calibration and inter-model comparisons. An important byproduct of the sensitivity analysis is the identification of the key parameters for each model instance, which will reduce the dimension of the model calibration process. Specifically, we will select the most sensitive parameters for each model instance (by relying, for instance, on total sensitivity indices; Pianosi et al., 2016), and calibrate them for each of the 180 VIC + VIC-Res instances. To this purpose, we will couple the hydrologic model with an optimization algorithm that calibrates the model parameters with respect to one or multiple metrics of performance (Reed et al., 2013). We will calibrate each model over the period 1990-2010 (with 1990-1991 for model spin-up) and then validate it over the period 2011-2020. In sum, this second experiment will yield an ensemble of 10 different model parameterizations and corresponding hydrologic simulations for each basin.

The ensemble of model calibrations (10 members for each basin) will be used to address **Research Question** 3. First, we will investigate how the parameterization and accuracy of simulated river discharge depends on the simulation of dam operations. To answer this question, we will compare our simulated river discharge against the time series of observed discharge retrieved from the GAGES-II dataset (Section 3). Through such comparison, we will learn whether different structural assumptions on dam operations have an impact on hydrologic model accuracy and parameterization. Second, we will investigate how the uncertainty in the representation of human actions impacts the representation of hydrologic processes. We will begin by extracting key fluxes (e.g., evaporation, surface runoff, base flow) and state variables (e.g., soil moisture, snow water equivalent) from the calibration ensemble. These will be compared across the nine models (for each basin) that explicitly account for dam operations against the benchmark model that excludes dam operations. This comparison will indicate (1) how different dam operation models affect the representation of hydrologic processes, (2) how the lack of human representation in hydrologic models curbs our ability to capture states and fluxes, and (3) how these effects (1 and 2) vary with changes in the natural environment. We will then corroborate our findings by comparing against ground observations, which can be retrieved for a few variables of interest—simulated total runoff (base flow plus surface runoff), for instance, can be compared against the USGS WaterWatch runoff, as suggested by Oubeidillah et al., 2014. This analysis is designed to identify pitfalls in the model conceptualization and calibration process—e.g., instances in which model calibration compensates for an incorrect representation of relevant processes, in particular those arising from human activity.

4.4 Deducting practical guidelines and generalizing insights (Task 4)

The availability of the data generated in Tasks 1-3 represents an excellent opportunity for deducting practical guidelines that hydrologists could follow when accounting for dam operations in large-scale hydrologic models (**Research Question 4**). Given the large amount of data, we plan to rely on statistical learning to identify relevant patterns. To this purpose, we will prepare a dataset consisting of (1) dam design features and operating objectives, (2) topographic and hydro-meteorological conditions of each basin, and (3) performance of each operating policy. When defining performance, we will not limit ourselves to the ability of a dam operation model to reproduce discharge and storage (Task 2) as well as the ability of the hydrologic model to simulate discharge (Task 3), since model accuracy can be confounded by parameterization issues. We will use a broader definition of 'performance', explicitly accounting, for instance, for the effect that a given dam operation model may have on the amplification of equifinality. With multiple metrics of performance at hand, we will finally rely on simple and interpretable statistical models (e.g., Classification and Regression Trees; James et al., 2013) to map dam design features, operating objectives, topographic and hydro-meteorological conditions into the performance of the dam operation models. The statistical models could thus be seen as a sequence of semi-quantitative guidelines that suggest the best modelling approach given the characteristics of the study site at hand.

In the final part our project, we will expand the findings to two major basins within the US, the Colorado and Columbia, both of which have been extensively modelled by the VIC community (e.g., Chegwidden et al., 2019; Turner et al., 2020; Eldardiry et al., 2022; Whitney et al., 2023). We plan, in particular, to carry out three experiments. First, we will implement the framework described in Tasks 2 and 3 to both basins; a modelling exercise that will yield 10 model instances for each basin and demonstrate how the (mis)representation of dams impacts process inference across large modelling domains. Second, we will use the two case studies to validate the practical guidelines described above. Finally, we will run our model instances under a set of future climate projections (i.e., CMIP6 multi-model ensemble; Eyring et al., 2016; O'Neill et al., 2016) to determine how structural uncertainty in reservoir regulation (and thus ill-posed hydrologic process inference) compares in the context of climate uncertainty and natural variability. Large-scale hydrologic models are commonly applied to estimate future water supply and flood risks, and this step aims to show the contribution of these structural assumptions to the total uncertainty.

5 Intellectual Merit

The overarching hypothesis of our project is that the representation of dam operations is a major source of structural uncertainty that curbs our ability to study hydrologic processes and fluxes of water in basins affected by human regulations. This hypothesis is well-aligned with the NSF HS Program along three main dimensions. First, it focuses on how hydrologic process inference is altered by human activity, and aims to show how hydrologic responses change not only with climate forcings, but also with changes in how the operations of large reservoirs are modeled. Importantly, these research needs emerged in a recent report by the National Academies of Sciences, Engineering, and Medicine 2020, which emphasized the need of understanding how the Earth's water cycle is changing is response to multiple drivers, including the effect of human activity (Sivapalan et al., 2014). Second, our methodological approach builds on the convergence of hydrologic modelling with remote sensing and water resources systems analysis. As explained in the aforementioned report, "observations from space will be increasingly vital for quantifying volumetric and temporal changes of different parts of the water cycle": our goal is to look at remotely-sense observations (of water reservoirs) through the lens of systems analysis tools, study the uncertainty in the representation of dam operations, and understand how such uncertainty reflects on large-scale hydrologic models. These efforts will help hydrologic models keep pace with the rapid changes in the Earth's water cycle. Finally, our project aims to advance both curiosity-driven and use-inspired basic research—a feature of Next Generation Earth Systems Science that was identified in another recent report (National Academies of Sciences & Medicine, 2022). Large-scale hydrologic models serve a variety of modelling applications, so characterizing their limits and improving their skill is a first, fundamental, step towards more informed and accurate hydrologic modelling efforts.

6 Broader Impacts

We believe the skills of the research team, the themes of the project, and its expected results will help us positively impact a few important domains.

Enhance infrastructure for research. The availability of datasets for large-sample studies has given our community the opportunity to test hypotheses, benchmark models, and improve our understanding of hydrologic processes. A case in point is CAMELS (Catchment Attributes and MEteorology for large-sample Studies), which provides hydro-meteorological time series together with the attributes for 671 catchments within the CONUS (Newman et al., 2015; Addor et al., 2017). However, CAMELS, like other datasets, was created for basins that are largely unaffected by human actions: we thus lack the underlying data needed to study hydrologic processes in human-altered basins. Our work will fill in this gap and contribute a large-sample dataset containing hydro-meteorologic forcings, dam design specifications, inflow / storage / release time series, and catchment attributes for 18 large basins containing $\sim$200 reservoirs. Our contribution to the infrastructure for research will be complemented by (1) simulated datasets of dam operations and hydrologic processes (Section 4.2 and 4.3) and (2) open-source toolboxes and models for inferring storage time series (Section 4.1) and simulating dam operations within hydrologic models (Section 4.3). Overall, these data, modelling tools, and practical guideline (Section 4.4) will expand the toolbox that is currently available to the large-scale hydrologic modelling community.

Broadly disseminate to enhance scientific understanding and engagement. We believe that a fundamental challenge preventing hydrologists from focusing on the representation of human actions (in hydrologic models) is the multi-disciplinary nature of such modelling effort—which requires skills in remote sensing, water system analysis, and, potentially, high-performance computing. To overcome this barrier and broadly disseminate the products described above, we propose to complement our publication and data management plans with two Scientific Workshops that will be organized during the 2025 and 2026 AGU Fall Meeting (corresponding to Year 2 and 3 of this project). These workshops will focus on three topics, namely (1) inferring storage time series from satellite images (Year 2), (2) modelling dam operations (Year 2), and (3) studying hydrologic processes in catchments affected by dam operations (Year 3). The reason behind the choice of this specific venue is that the AGU Fall Meeting arguably provides the best opportunity for engaging with the user base of the data and models we will develop. We also note that the participation at the AGU Fall Meeting is part of our budget and allows to organize Scientific Workshops at no cost. Participation to the workshop will be solicited through the AGU Fall Meeting website as well as the PIs' network. A key role will be played by the interaction with the leadership of specific AGU Hydrology Technical Committees (e.g., Catchment Hydrology, Hydrologic Uncertainty, Remote Sensing, Hydrology Section Student Committee).

Promote teaching, training, and learning. This project will provide a unique research training experience for two Postdoctoral researchers and one graduate student across the fields of large-scale and catchment hydrology, remote sensing, and water systems analysis. As described in Section 7 and in the Postdoctoral Mentoring Plan, the team members will follow a tailored training and will author peer-reviewed publications, datasets, and models. One undergraduate student (at Cornell) will be directly supported by this project, and others will be involved through opportunities provided by Cornell and UC Davis. At Cornell, Dr. Galelli currently mentors two undergraduate students via Engineers for a Sustainable World and the Atkinson Center for Sustainability. At UC Davis, Dr. Herman currently mentors two undergraduate students through a complementary research project on forecast-informed reservoir operations. Finally, the project will provide an opportunity to broaden the diversity of scholars engaged in hydrologic modelling. In this regard, we note that both Dr. Galelli's lab (previously in Singapore and now in Ithaca) and Dr. Herman's lab have an established history of mentoring international students and students from underrepresented groups.

7 Project Management Plan

7.1 Personnel

Stefano Galelli, Lead PI, is a tenured Associate Professor at Cornell University (School of Civil and Environmental Engineering). Dr. Galelli studies the interactions between climate and interconnected infrastructure systems, with particular focus on power grids, rivers, and the services they provide. His recent work has been focusing on combining knowledge from large-scale hydrology and water systems analysis to advance the capabilities and accuracy of hydrologic models.

Jon Herman, Co-PI, is is a tenured Associate Professor at UC Davis in the Department of Civil and Environmental Engineering. His research group focuses on water resources systems analysis, specifically the development of computational methods to support water planning and management under uncertainty. Recent work includes forecast-based reservoir operations, adaptation to climate change, and sensitivity analysis of hydrologic models.

Two Postdoctoral Research Fellows will be responsible for the estimation of the storage time series from remote sensing (Task 1), the collection of inflow and release data (Task 1), and the calibration and testing of the dam operations models for all dams within our study site (Task 2). The postdocs will also support Tasks 3-4.

One Graduate Student Researcher will implement the dam operations models in VIC + VIC-Res, and will lead the sensitivity analysis, calibration, and inter-model comparisons (Task 3). The Graduate Student Researcher will also lead the activities in Task 4 (deduction of practical guidelines and test on Colorado and Columbia River basins).

One Undergraduate Researcher will assist with the core of the project, namely the implementation and analysis of VIC + VIC-Res on multiple river basins. We therefore expect the student to pick-up research expertise in large-scale hydrology and sensitivity analysis.

7.2 Research Schedule and PI Responsibilities

All personnel listed above will be co-authors on relevant manuscripts resulted from this project.We expect publications in major journals to appear during Year 2-3. Presentations will be made at the AGU Fall Meeting and other conferences from Year 1. Public lectures at our respective institutes and via web-based platforms will be pursued concurrently.

Table 2. Project schedule and PI responsibilities.

Task	Activity	YR 1	YR 2	YR 3	PI
1	Estimation of storage time series	✓			SG
1	Collection of inflow and release data	✓			SG,JH
1	Consolidation of the {storage, inflow, release} dataset	✓			SG,JH
2	Calibration and validation of dam operations models	✓	✓		JH
3	Setup of VIC + VIC-Res		✓		SG
3	Sensitivity analysis		✓		SG,JH
3	Calibration and inter-model comparisons		✓	✓	SG,JH
4	Deduction of practical guidelines			✓	SG, JH
4	Large-scale test on Colorado and Columbia			✓	SG

8 Results from Prior NSF Support

8.1 S. Galelli: no prior NSF support.

8.2 J. Herman: Award: CBET-2041826. *Others in Current & Pending.*

Title: CAREER: Dynamic adaptation of water resources systems to navigate uncertain hydrologic and human stressors. **PI:** J. Herman. **Amount:** $509,527. **Period:** 03/21-02/26. **Intellectual Merit:** Framework to detect increases in flood and drought severity under climate change and respond with adaptive design and operation of reservoir infrastructure. **Broader Impacts:** Feedbacks between research, education, and professional practice on the topic of computational skills in civil engineering. Supported two graduate students and 1 postdoctoral scholar. **Products:** Cohen and Herman (2021), Steinschneider et al. (2023), Chen and Herman (2024).

References

Abbott, B. W., Bishop, K., Zarnetske, J. P., Minaudo, C., Chapin, F., Krause, S., Hannah, D. M., Conner, L., Ellison, D., Godsey, S. E., et al. (2019). Human domination of the global water cycle absent from depictions and perceptions. *Nature Geoscience*, *12*(7), 533–540.

Abeshu, G. W., Tian, F., Wild, T., Zhao, M., Turner, S., Chowdhury, A., Vernon, C. R., Hu, H., Zhuang, Y., Hejazi, M., et al. (2023). Enhancing the representation of water management in global hydrological models. *Geoscientific Model Development*, *16*(18), 5449–5472.

Adam, J. C., Hamlet, A. F., & Lettenmaier, D. P. (2009). Implications of global climate change for snowmelt hydrology in the twenty-first century. *Hydrological Processes: An International Journal*, *23*(7), 962–972.

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, *21*(10), 5293–5313.

Bastidas, L., Hogue, T., Sorooshian, S., Gupta, H., & Shuttleworth, W. (2006). Parameter sensitivity analysis for different complexity land surface models using multicriteria methods. *Journal of Geophysical Research: Atmospheres*, *111*(D20).

Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Wwater Resources*, *16*(1), 41–51.

Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT mission and its capabilities for land hydrology. In *Remote Sensing and Water Resources* (pp. 117–147). Springer.

Biemans, H., Haddeland, I., Kabat, P., Ludwig, F., Hutjes, R. W., Heinke, J., von Bloh, W., & Gerten, D. (2011). Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resources Research*, *47*(3).

Bierkens, M. F. (2015). Global hydrology 2015: State, trends, and directions. *Water Resources Research*, *51*(7), 4923–4947.

Birkett, C., Reynolds, C., Beckley, B., & Doorn, B. (2011). From research to operations: The USDA global reservoir and lake monitor. *Coastal Altimetry*, 19–50.

Biswas, N. K., Hossain, F., Bonnema, M., Lee, H., & Chishtie, F. (2021). Towards a global Reservoir Assessment Tool for predicting hydrologic impacts and operating patterns of existing and planned reservoirs. *Environmental Modelling & Software*, *140*, 105043.

Bohn, T. J., & Vivoni, E. R. (2019). MOD-LSP, MODIS-based parameters for hydrologic modeling of North American land cover change. *Scientific Data*, *6*(1), 144.

Bonnema, M., & Hossain, F. (2017). Inferring reservoir operating patterns across the Mekong Basin using only space observations. *Water Resources Research*, *53*(5), 3791–3810.

Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., Pekel, J.-F., & Cottam, A. (2019). A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrology and Earth System Sciences*, *23*(2), 669–690.

Cáceres, A. L., Jaramillo, P., Matthews, H. S., Samaras, C., & Nijssen, B. (2022). Potential hydropower contribution to mitigate climate risk and build resilience in africa. *Nature Climate Change*, *12*(8), 719–727.

Chaney, N., Herman, J., Reed, P., & Wood, E. F. (2015). Flood and drought hydrologic monitoring: The role of model parameter uncertainty. *Hydrology and Earth System Sciences*, *19*(7), 3239–3251.

Chegwidden, O. S., Nijssen, B., Rupp, D. E., Arnold, J. R., Clark, M. P., Hamman, J. J., Kao, S.-C., Mao, Y., Mizukami, N., Mote, P. W., et al. (2019). How do modeling decisions affect the spread among hydrologic climate change projections? Exploring a large ensemble of simulations across a diversity of hydroclimates. *Earth's Future*, *7*(6), 623–637.

Chen, M., & Herman, J. D. (2024). Detection Time for Nonstationary Reservoir System Performance Driven by Climate and Land-Use Change. *Journal of Water Resources Planning and Management*, *150*(4), 04024004.

Chowdhury, A. K., Deshmukh, R., Wu, G. C., Uppal, A., Mileva, A., Curry, T., Armstrong, L., Galelli, S., & Ndhlukula, K. (2022). Enabling a low-carbon electricity system for Southern Africa. *Joule*, *6*(8), 1826–1844.

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J., Tang, G., Gharari, S., Freer, J. E., Whitfield, P. H., Shook, K. R., et al. (2021). The abuse of popular performance metrics in hydrologic modeling. *Water Resources Research*, *57*(9), e2020WR029001.

Coerver, H. M., Rutten, M. M., & Van De Giesen, N. C. (2018). Deduction of reservoir operating rules for application in global hydrological models. *Hydrology and Earth System Sciences*, *22*(1), 831–851.

Cohen, J. S., & Herman, J. D. (2021). Dynamic adaptation of water resources systems under uncertainty by learning policy structure and indicators. *Water Resources Research*, *57*(11), e2021WR030433.

Dan, L., Ji, J., Xie, Z., Chen, F., Wen, G., & Richey, J. E. (2012). Hydrological projections of climate change scenarios over the 3H region of China: A VIC model assessment. *Journal of Geophysical Research: Atmospheres*, *117*(D11).

Dang, T. D., Chowdhury, A. K., & Galelli, S. (2020a). On the representation of water reservoir storage and operations in large-scale hydrological models: Implications on model parameterization and climate change impact assessments. *Hydrology and Earth System Sciences*, *24*(1), 397–416. https://doi.org/10.5194/hess-24-397-2020

Dang, T. D., Vu, D. T., Chowdhury, A. K., & Galelli, S. (2020b). A software package for the representation and optimization of water reservoir operations in the VIC hydrologic model. *Environmental Modelling & Software*, *126*, 104673.

Demaria, E. M., Nijssen, B., & Wagener, T. (2007). Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model. *Journal of Geophysical Research: Atmospheres*, *112*(D11).

Döll, P., Fiedler, K., & Zhang, J. (2009). Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. *Hydrology and Earth System Sciences*, *13*(12), 2413–2432.

Döll, P., Kaspar, F., & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: Model tuning and validation. *Journal of Hydrology*, *270*(1-2), 105–134.

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water resources research*, *28*(4), 1015–1031.

Duncombe, J. (2022). Not your childhood water cycle. *Eos*, *103*.

Ehsani, N., Fekete, B. M., Vörösmarty, C. J., & Tessler, Z. D. (2016). A neural network based general reservoir operation scheme. *Stochastic environmental research and risk assessment*, *30*, 1151–1166.

Eldardiry, H., Zhou, T., Huang, M., & Chegwidden, O. S. (2022). The Role of Groundwater Withdrawals on River Regulation: Example From the Columbia River Basin. *Water Resources Research*, *58*(6), e2020WR028955.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958.

Falcone, J. A. (2011). *GAGES-II: geospatial attributes of gages for evaluating streamflow* (tech. rep.). US Geological Survey.

Falkenmark, M. (1989). The massive water scarcity now threatening Africa: why isn't it being addressed? *Ambio*, 112–118.

Falkenmark, M. (1997). Meeting water requirements of an expanding world population. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1356), 929–936.

Galelli, S., Dang, T. D., Ng, J. Y., Chowdhury, A., & Arias, M. E. (2022). Opportunities to curb hydrological alterations via dam re-operation in the Mekong. *Nature Sustainability*, *5*(12), 1058–1069.

Gao, H. (2015). Satellite remote sensing of large lakes and reservoirs: From elevation and area to storage. *Wiley Interdisciplinary Reviews: Water*, *2*(2), 147–157.

Ghimire, G. R., Hansen, C., Gangrade, S., Kao, S.-C., Thornton, P. E., & Singh, D. (2023). Insights From Dayflow: A Historical Streamflow Reanalysis Dataset for the Conterminous United States. *Water Resources Research*, *59*(2), e2022WR032312.

Giuliani, M., Lamontagne, J., Reed, P., & Castelletti, A. (2021). A state-of-the-art review of optimal reservoir control for managing conflicting demands in a changing world. *Water Resources Research*, *57*(12), e2021WR029927.

Giuliani, M., Castelletti, A., Pianosi, F., Mason, E., & Reed, P. M. (2016). Curses, tradeoffs, and scalable management: Advancing evolutionary multiobjective direct policy search to improve water reservoir operations. *Journal of Water Resources Planning and Management*, *142*(2), 04015050.

Greuell, W., Franssen, W. H., Biemans, H., & Hutjes, R. W. (2018). Seasonal streamflow forecasts for Europe–Part I: Hindcast verification with pseudo-and real observations. *Hydrology and Earth System Sciences*, *22*(6), 3453–3472.

Greve, P., Burek, P., & Wada, Y. (2020). Using the Budyko framework for calibrating a global hydrological model. *Water Resources Research*, *56*(6), e2019WR026280.

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., Babu, S., Borrelli, P., Cheng, L., Crochetiere, H., et al. (2019). Mapping the world's free-flowing rivers. *Nature*, *569*(7755), 215–221.

Gupta, H. V., & Razavi, S. (2018). Revisiting the basis of sensitivity analysis for dynamical earth system models. *Water Resources Research*, *54*(11), 8692–8717.

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M., Ludwig, F., Masaki, Y., Schewe, J., et al. (2014). Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences*, *111*(9), 3251–3256.

Haddeland, I., Skaugen, T., & Lettenmaier, D. P. (2006). Anthropogenic impacts on continental surface water fluxes. *Geophysical Research Letters*, *33*(8).

Hanasaki, N., Kanae, S., & Oki, T. (2006). A reservoir operation scheme for global river routing models. *Journal of Hydrology*, *327*(1-2), 22–41.

Hanasaki, N., Matsuda, H., Fujiwara, M., Hirabayashi, Y., Seto, S., Kanae, S., & Oki, T. (2022). Toward hyper-resolution global hydrological models including human activities: application to Kyushu island, Japan. *Hydrology and Earth System Sciences*, *26*(8), 1953–1975.

Hansen, M. C., DeFries, R. S., Townshend, J. R., & Sohlberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, *21*(6-7), 1331–1364.

Hao, Z., Chen, F., Jia, X., Cai, X., Yang, C., Du, Y., & Ling, F. (2024). GRDL: A New Global Reservoir Area-Storage-Depth Data Set Derived Through Deep Learning-Based Bathymetry Reconstruction. *Water Resources Research*, *60*(1), e2023WR035781.

Haynes, J. V., Read, A. L., Chan, A. Y., Martin, D. J., Regan, R. S., Henson, W., Niswonger, R., & Stewart, J. S. (2023). Monthly crop irrigation withdrawals and efficiencies by HUC12 watershed for years 2000-2020 within the conterminous United States. https://doi.org/10.5066/P9LGISUM

Herman, J., & Usher, W. (2017). Salib: An open-source python library for sensitivity analysis. *Journal of Open Source Software*, *2*(9), 97.

Herman, J. D., & Giuliani, M. (2018). Policy tree optimization for threshold-based water resources management over multiple timescales. *Environmental modelling & software*, *99*, 39–51.

Hoch, J. M., Sutanudjaja, E. H., Wanders, N., van Beek, R. L., & Bierkens, M. F. (2023). Hyper-resolution PCR-GLOBWB: opportunities and challenges from refining model spatial resolution to 1 km over the European continent. *Hydrology & Earth System Sciences*, *27*(6).

Hodgkins, G. A., Over, T. M., Dudley, R. W., Russell, A. M., & LaFontaine, J. H. (2023). The consequences of neglecting reservoir storage in national-scale hydrologic models: An appraisal of key streamflow statistics. *JAWRA Journal of the American Water Resources Association*.

Hogue, T. S., Bastidas, L. A., Gupta, H. V., & Sorooshian, S. (2006). Evaluating model performance and parameter behavior for varying levels of land surface model complexity. *Water Resources Research*, *42*(8).

Hou, J., van Dijk, A. I., Beck, H. E., Renzullo, L. J., & Wada, Y. (2022). Remotely sensed reservoir water storage dynamics (1984–2015) and the influence of climate variability and management at a global scale. *Hydrology and Earth System Sciences*, *26*(14), 3785–3803.

James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Johnson, J. M., Fang, S., Sankarasubramanian, A., Rad, A. M., Kindl da Cunha, L., Jennings, K. S., Clarke, K. C., Mazrooei, A., & Yeghiazarian, L. (2023). Comprehensive analysis of the NOAA National Water Model: A call for heterogeneous formulations and diagnostic model selection. *Journal of Geophysical Research: Atmospheres*, *128*(24), e2023JD038534.

Kanyako, F., Lamontagne, J., Baker, E., Turner, S., & Wild, T. (2023). Seasonality and trade in hydro-heavy electricity markets: A case study with the West Africa Power Pool (WAPP). *Applied Energy*, *329*, 120214.

Khatami, S., Peel, M. C., Peterson, T. J., & Western, A. W. (2019). Equifinality and flux mapping: A new approach to model evaluation and process representation under uncertainty. *Water Resources Research*, *55*(11), 8922–8941.

Khazaei, B., Read, L. K., Casali, M., Sampson, K. M., & Yates, D. N. (2022). GLOBathy, the global lakes bathymetry dataset. *Scientific Data*, *9*(1), 36.

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*(3).

Knoben, W. J., Freer, J. E., & Woods, R. A. (2019). Inherent benchmark or not? comparing nash–sutcliffe and kling–gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331.

Knüsel, B., Zumwald, M., Baumberger, C., Hirsch Hadorn, G., Fischer, E. M., Bresch, D. N., & Knutti, R. (2019). Applying big data beyond small problems in climate research. *Nature Climate Change*, *9*(3), 196–202.

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

Li, Y., Gao, H., Zhao, G., & Tseng, K.-H. (2020). A high-resolution bathymetry dataset for global reservoirs using multi-source satellite imagery and altimetry. *Remote Sensing of Environment*, *244*, 111831.

Li, Y., Zhao, G., Allen, G. H., & Gao, H. (2023). Diminishing storage returns of reservoir construction. *Nature Communications*, *14*(1), 3203.

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres*, *99*(D7), 14415–14428.

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (2014). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres*, *99*(D7), 14415–14428. https://doi.org/10.1029/94jd00483

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., et al. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resources Research*, *55*(8), 6499–6516.

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., Maurer, E. P., & Lettenmaier, D. P. (2013). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *Journal of Climate*, *26*(23), 9384–9392.

Lohmann, D., Nolte-Holube, R., & Raschke, E. (1996). A large-scale horizontal routing model to be coupled to land surface parametrization schemes. *Tellus A: Dynamic Meteorology and Oceanography*, *48*(5), 708–721. https://doi.org/10.3402/tellusa.v48i5.12200

Lohmann, D., Raschke, E., Nijssen, B., & Lettenmaier, D. P. (1998). Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model. *Hydrological Sciences Journal*, *43*(1), 131–141. https://doi.org/10.1080/02626669809492107

Masaki, Y., Hanasaki, N., Biemans, H., Schmied, H. M., Tang, Q., Wada, Y., Gosling, S. N., Takahashi, K., & Hijioka, Y. (2017). Intercomparison of global river discharge simulations focusing on dam operation—multiple models analysis in two case-study river basins, Missouri–Mississippi and Green–Colorado. *Environmental Research Letters*, *12*(5), 055002.

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., & Nijssen, B. (2002). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous united states. *Journal of Climate*, *15*(22), 3237–3251.

Melsen, L. A., Teuling, A. J., Torfs, P. J., Uijlenhoet, R., Mizukami, N., & Clark, M. P. (2016). HESS opinions: The need for process-based evaluation of large-domain hyper-resolution models. *Hydrology and Earth System Sciences*, *20*(3), 1069–1079.

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, *7*(1), 13603.

Miller, D. A., & White, R. A. (1998). A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. *Earth Interactions*, *2*(2), 1–26.

Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., et al. (2021). The global water resources and use model WaterGAP v2. 2d: Model description and evaluation. *Geoscientific Model Development*, *14*(2), 1037–1079.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., & Döll, P. (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences*, *18*(9), 3511–3538.

Myneni, R., Knyazikhin, Y., & Park, T. (2015). OD15A2H MODIS/Combined Terra+Aqua Leaf Area Index/FPAR Daily L4 Global 500m SIN Grid [Accessed: February 1, 2024].

National Academies of Sciences, E., & Medicine. (2020). *A Vision for NSF Earth Sciences 2020-2030: Earth in Time*. The National Academies Press. https://doi.org/10.17226/25761

National Academies of Sciences, E., & Medicine. (2022). *Next Generation Earth Systems Science at the National Science Foundation*. The National Academies Press. https://doi.org/10.17226/26042

Nazemi, A., & Wheater, H. S. (2015a). On inclusion of water resource management in Earth system models–Part 1: Problem definition and representation of water demand. *Hydrology and Earth System Sciences*, *19*(1), 33–61.

Nazemi, A., & Wheater, H. S. (2015b). On inclusion of water resource management in Earth system models–Part 2: Representation of water supply and allocation and opportunities for improved modeling. *Hydrology and Earth System Sciences*, *19*(1), 63–90.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J., et al. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, *19*(1), 209–223.

Ng, J. Y., Turner, S. W., & Galelli, S. (2017). Influence of El Niño Southern Oscillation on global hydropower production. *Environmental Research Letters*, *12*(3), 034010.

Nijssen, B., Schnur, R., & Lettenmaier, D. P. (2001). Global retrospective estimation of soil moisture using the Variable Infiltration Capacity land surface model, 1980–93. *Journal of Climate*, *14*(8), 1790–1808.

O'Neill, B. C., Tebaldi, C., Vuuren, D. P. v., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., et al. (2016). The scenario model intercomparison project (scenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482.

O'neill, M. M., Tijerina, D. T., Condon, L. E., & Maxwell, R. M. (2021). Assessment of the ParFlow–CLM CONUS 1.0 integrated hydrologic model: evaluation of hyper-resolution water balance components across the contiguous United States. *Geoscientific Model Development*, *14*(12), 7223–7254.

Otta, K., Müller Schmied, H., Gosling, S. N., & Hanasaki, N. (2023). Use of satellite remote sensing to validate reservoir operations in global hydrological models: a case study from the CONUS. *Hydrology and Earth System Sciences Discussions*, *2023*, 1–30.

Oubeidillah, A. A., Kao, S.-C., Ashfaq, M., Naz, B. S., & Tootle, G. (2014). A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US. *Hydrology and Earth System Sciences*, *18*(1), 67–84.

Ouyang, W., Lawson, K., Feng, D., Ye, L., Zhang, C., & Shen, C. (2021). Continental-scale streamflow modeling of basins with reservoirs: Towards a coherent deep-learning-based strategy. *Journal of Hydrology*, *599*, 126455.

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232.

Pokhrel, Y. N., Hanasaki, N., Wada, Y., & Kim, H. (2016). Recent progresses in incorporating human land–water management into global land surface models toward their integration into Earth system models. *Wiley Interdisciplinary Reviews: Water*, *3*(4), 548–574.

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., & Kollat, J. B. (2013). Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in water resources*, *51*, 438–456.

Schaperow, J. R., Li, D., Margulis, S. A., & Lettenmaier, D. P. (2021). A near-global, high resolution land surface parameter dataset for the variable infiltration capacity model. *Scientific Data*, *8*(1), 216.

Schwatke, C., Dettmering, D., Bosch, W., & Seitz, F. (2015). DAHITI–an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences*, *19*(10), 4345–4364.

Shen, Y., Liu, D., Jiang, L., Nielsen, K., Yin, J., Liu, J., & Bauer-Gottwein, P. (2022). High-resolution water level and storage variation datasets for 338 reservoirs in China during 2010–2021. *Earth System Science Data*, *14*(12), 5671–5694.

Shiklomanov, I. A. (1997). Comprehensive assessment of the freshwater resources of the world. Stockholm Environ. Inst.

Shin, S., Pokhrel, Y., & Miguez-Macho, G. (2019). High-resolution modeling of reservoir release and storage dynamics at the continental scale. *Water Resources Research*, *55*(1), 787–810.

Simon, H. A. (1957). A behavioral model of rational choice. *Models of man, social and rational: Mathematical essays on rational human behavior in a social setting*, 241–260.

Sivapalan, M., Konar, M., Srinivasan, V., Chhatre, A., Wutich, A., Scott, C., Wescoat, J., & Rodríguez-Iturbe, I. (2014). Socio-hydrology: Use-inspired water sustainability science for the Anthropocene. *Earth's Future*, *2*(4), 225–230.

Steinschneider, S., Herman, J. D., Kucharski, J., Abellera, M., & Ruggiero, P. (2023). Uncertainty decomposition to understand the influence of water systems model error in climate vulnerability assessments. *Water Resources Research*, *59*(1), e2022WR032349.

Steyaert, J. C., Condon, L. E., WD Turner, S., & Voisin, N. (2022). ResOpsUS, a dataset of historical reservoir operations in the contiguous United States. *Scientific Data*, *9*(1), 34.

Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, *11*, 341–359.

Sutanudjaja, E. H., Van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H., Drost, N., Van Der Ent, R. J., De Graaf, I. E., Hoch, J. M., De Jong, K., et al. (2018). PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. *Geoscientific Model Development*, *11*(6), 2429–2453.

Tang, Y., Reed, P., & Wagener, T. (2006). How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? *Hydrology and Earth System Sciences*, *10*(2), 289–307.

Tefs, A., Stadnyk, T., Koenig, K., Dery, S. J., MacDonald, M., Slota, P., Crawford, J., & Hamilton, M. (2021). Simulating river regulation and reservoir performance in a continental-scale hydrologic model. *Environmental Modelling & Software*, *141*, 105025.

Telteu, C.-E., Müller Schmied, H., Thiery, W., Leng, G., Burek, P., Liu, X., Boulange, J. E. S., Andersen, L. S., Grillakis, M., Gosling, S. N., et al. (2021). Understanding each other's models: an introduction and a standard representation of 16 global water models to support intercomparison, improvement, and communication. *Geoscientific Model Development*, *14*(6), 3843–3878.

Thornton, P. E., Shrestha, R., Thornton, M., Kao, S.-C., Wei, Y., & Wilson, B. E. (2021). Gridded daily weather data for North America with comprehensive uncertainty quantification. *Scientific Data*, *8*(1), 190.

Turner, S. W., Doering, K., & Voisin, N. (2020). Data-driven reservoir simulation in a large-scale hydrological and water resource model. *Water Resources Research*, *56*(10), e2020WR027902.

Turner, S. W., & Galelli, S. (2016). Water supply sensitivity to climate change: An R package for implementing reservoir storage analysis in global and regional impact studies. *Environmental Modelling & Software*, *76*, 13–19.

Turner, S. W., Ng, J. Y., & Galelli, S. (2017). Examining global electricity supply vulnerability to climate change using a high-fidelity hydropower dam model. *Science of the Total Environment*, *590*, 663–675.

Turner, S. W., Steyaert, J. C., Condon, L., & Voisin, N. (2021). Water storage and release policies for all large reservoirs of conterminous united states. *Journal of Hydrology*, *603*, 126843.

USGS. (2018). USGS 3D Elevation Program (3DEP) [Accessed: February 1, 2024].

Vegad, U., & Mishra, V. (2022). Ensemble streamflow prediction considering the influence of reservoirs in Narmada River Basin, India. *Hydrology and Earth System Sciences*, *26*(24), 6361–6378.

Veldkamp, T. I. E., Zhao, F., Ward, P. J., De Moel, H., Aerts, J. C., Schmied, H. M., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., et al. (2018). Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: A multi-model validation study. *Environmental Research Letters*, *13*(5), 055008.

Voisin, N., Hejazi, M. I., Leung, L. R., Liu, L., Huang, M., Li, H.-Y., & Tesfa, T. (2017). Effects of spatially distributed sectoral water management on the redistribution of water resources in an integrated water model. *Water Resources Research*, *53*(5), 4253–4270.

Voisin, N., Liu, L., Hejazi, M., Tesfa, T., Li, H., Huang, M., Liu, Y., & Leung, L. (2013). One-way coupling of an integrated assessment model and a water resources model: Evaluation and implications of future changes over the us midwest. *Hydrology and Earth System Sciences*, *17*(11), 4555–4575.

Vörösmarty, C. J., Meybeck, M., Fekete, B., Sharma, K., Green, P., & Syvitski, J. P. (2003). Anthropogenic sediment retention: Major global impact from registered river impoundments. *Global and Planetary Change*, *39*(1-2), 169–190.

Vrugt, J. A., Gupta, H. V., Bouten, W., & Sorooshian, S. (2003). A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, *39*(8).

Vu, D. T., Dang, T. D., Galelli, S., & Hossain, F. (2022). Satellite observations reveal 13 years of reservoir filling strategies, operating rules, and hydrological alterations in the Upper Mekong River basin. *Hydrology and Earth System Sciences*, *26*(9), 2345–2364.

Vu, D. T., Dang, T. D., Pianosi, F., & Galelli, S. (2023). Calibrating macro-scale hydrological models in poorly gauged and heavily regulated basins. *Hydrology and Earth System Sciences Discussions*, 1–28.

Wada, Y., Bierkens, M. F., De Roo, A., Dirmeyer, P. A., Famiglietti, J. S., Hanasaki, N., Konar, M., Liu, J., Müller Schmied, H., Oki, T., et al. (2017). Human–water interface in hydrological modelling: Current status and future directions. *Hydrology and Earth System Sciences*, *21*(8), 4169–4193.

Whitney, K. M., Vivoni, E. R., Bohn, T. J., Mascaro, G., Wang, Z., Xiao, M., Mahmoud, M. I., Cullom, C., & White, D. D. (2023). Spatial attribution of declining Colorado River streamflow under future warming. *Journal of Hydrology*, *617*, 129125.

Wi, S., Ray, P., Demaria, E. M., Steinschneider, S., & Brown, C. (2017). A user-friendly software package for VIC hydrologic model development. *Environmental Modelling & Software*, *98*, 35–53.

Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L., Bierkens, M. F., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., et al. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring earth's terrestrial water. *Water Resources Research*, *47*(5).

Xue, X., Zhang, K., Hong, Y., Gourley, J. J., Kellogg, W., McPherson, R. A., Wan, Z., & Austin, B. N. (2015). New multisite cascading calibration approach for hydrological models: Case study in the Red River basin using the VIC model. *Journal of Hydrologic Engineering*, *21*(2), 05015019.

Yang, T., Gao, X., Sorooshian, S., & Li, X. (2016). Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme. *Water Resources Research*, *52*(3), 1626–1651.

Yang, X., Pavelsky, T. M., Ross, M. R., Januchowski-Hartley, S. R., Dolan, W., Altenau, E. H., Belanger, M., Byron, D., Durand, M., Van Dusen, I., et al. (2022). Mapping flow-obstructing structures on global rivers. *Water Resources Research*, *58*(1), e2021WR030386.

Yassin, F., Razavi, S., Elshamy, M., Davison, B., Sapriza-Azuri, G., & Wheater, H. (2019). Representation and improved parameterization of reservoir operation in hydrological and land-surface models. *Hydrology and Earth System Sciences*, *23*(9), 3735–3764.

Yigzaw, W., Li, H.-Y., Demissie, Y., Hejazi, M. I., Leung, L. R., Voisin, N., & Payn, R. (2018). A new global storage-area-depth data set for modeling reservoirs in land surface and earth system models. *Water Resources Research*, *54*(12), 10–372.

Yuan, X., Wood, E. F., Roundy, J. K., & Pan, M. (2013). CFSv2-based seasonal hydroclimatic forecasts over the conterminous United States. *Journal of Climate*, *26*(13), 4828–4847.

Zhang, A. T., & Gu, V. X. (2023). Global Dam Tracker: A database of more than 35,000 dams with location, catchment, and attribute information. *Scientific Data*, *10*(1), 111.

Zhao, G., Gao, H., Naz, B. S., Kao, S.-C., & Voisin, N. (2016). Integrating a reservoir regulation scheme into a spatially distributed hydrological model. *Advances in Water Resources*, *98*, 16–31.

Publications from co-PI Herman's CBET-2041826

Cohen, J. S., & Herman, J. D. (2021). Dynamic adaptation of water resources systems under uncertainty by learning policy structure and indicators. *Water Resources Research*, *57*(11),

e2021WR030433.
Steinschneider, S., Herman, J. D., Kucharski, J., Abellera, M., & Ruggiero, P. (2023). Uncertainty decomposition to understand the influence of water systems model error in climate vulnerability assessments. *Water Resources Research*, *59*(1), e2022WR032349.
Chen, M., & Herman, J. D. (2024). Detection Time for Nonstationary Reservoir System Performance Driven by Climate and Land-Use Change. *Journal of Water Resources Planning and Management*, *150*(4), 04024004.