Hierarchical Temporal Scale Data-driven Reservoir Operation Modeling

Qianqiu Longyang[1], Ruijie Zeng[1*]

[1] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, 85281.

Corresponding author: Ruijie Zeng (ruijie.zeng.1@asu.edu)

Key Points:

- A hierarchical temporal scale framework is developed for data-driven reservoir operation modeling and tested across CONUS.
- The hierarchical temporal scale framework better captures reservoir release decisions under different operation targets and interactions of decisions across time scales than a single-scale model.
- The effects of decision variables on reservoir operations changes across time scales.

Abstract

Reservoirs are the key hydraulic infrastructure that regulates natural streamflow variability to fulfill various operation targets, including flood control, water supply, hydroelectricity generation and sustaining environmental flow. As an important anthropogenic interference on the hydrologic cycle, reservoir operation behavior remains challenging to be properly represented in hydrologic models, thus limiting the capability of predicting streamflow under the interactions between hydrologic variability and operational preferences. Data-driven models that utilize machine learning techniques provide a promising approach to represent reservoir operation rules by capturing relationships embedded in historical records. Similar to hydrologic processes vary across temporal scales, reservoir operation behaviors manifest themselves at different timescales, prioritizing different operation targets to mitigate streamflow variability at a given time scale. To capture the interaction of reservoir operation across time scales, we proposed a hierarchical temporal scale framework to investigate the behaviors of over 300 major reservoirs across the Contiguous United States with a wide range of streamflow conditions. Machine learning models were constructed to simulate reservoir operation at monthly, weekly, and daily scales, where decisions at short-term scales interact with long-term decisions. We found that the hierarchical temporal scale configuration better captures reservoir releases than models constructed at a single time scale, especially for reservoirs with multiple operation targets. Model-based sensitivity analysis shows that for more than one third of the studied reservoirs, the release schemes, as a function of decision variables, vary at different time scales, suggesting that operators are commonly faced with complicated trade-offs to serve multiple designed purposes. The proposed hierarchical temporal scale approach is flexible to incorporate various data-driven models and decision variables to derive reservoir operation rule, providing a robust framework to understand the feedbacks between natural streamflow variability and human interferences across time scales.

1 Introduction

Anthropogenic activities, such as reservoir operation (Haddeland et al., 2006; Döll et al., 2009; Biemans et al., 2011; Zhao et al., 2021; Singh and Basu, 2022; Zeng and Ren, 2022), urbanization (Oudin et al., 2018; Li et al., 2020) and large-scale irrigation (Siebert et al., 2010; Ferguson et al., 2011; Condon et al., 2019; Wei et al., 2022), have become increasingly important or even dominant driving forces of hydrologic processes in many watersheds over the world. In these watersheds, the streamflow observed at gauging stations represents the interaction between hydrologic and anthropogenic driving forces, rather than the "natural" or "unregulated" flows simulated in hydrologic models (Clark et al., 2015; Blair and Buytaert, 2016). Reservoirs are one of the key water infrastructures that directly regulate the streamflow timing and variability to fulfill various purposes including flood control, water supply, hydroelectricity generation, navigation and fluvial ecosystem services (Simonovic et al., 1992; Lehner et al., 2011; Ehsani et al., 2017; Moran et al., 2018; Boulange et al., 2021; Forsberg et al., 2017; Ortiz-Partida, Lane, and Sandoval-Solis, 2016; Patterson and Doyle, 2018). In the US, the National Inventory of Dams reports that there are more than 90,000 reservoirs (defined as equal or exceed 25 feet in height and exceed 15 acre-feet in storage, or exceed 6 feet in height and equal or exceed 50 acre-feet storage) regulating the streamflow (DeNeale et al., 2019). These reservoirs altogether store freshwater resources equivalent to one year's average natural runoff (Graf, 1999), generates about 6.3% of total electricity and 31.3% of renewable energy production (EIA, 2022), and protect hundreds of millions of populations from flooding. Meanwhile, the current reservoir operation policies are challenged by shifting flow conditions under climate change (Boulange et al., 2021), elevated risks due to aging infrastructure (Lane, 2007), increasing demand for water supply reliability, and needs for aquatic habitat restoration (Tonkin et al., 2018; Palmer et al., 2019). Understanding how reservoirs are operated and their interaction with hydrologic cycle is vitally important for assessing reliability and risks of reservoir functioning (Brekke et al., 2009), designing adaption strategies for future climate (Ho et al., 2017), and mitigating the tradeoffs among conflicting operation targets (Suen et al., 2006; Chen et al., 2017; Giuliani et al., 2021) to achieve sustainable water resources management.

Reservoirs are decision hubs that integrate the complex feedbacks between hydrologic variability and operational targets under various constraints, such as reservoir inflow, water storage capacity, hydroelectricity generation requirement and competitions among different operation purposes. Challenges remain for modelling the reservoir release decisions, which often involve on complex and undocumented decision processes. Often, reservoir operation guidelines are based on predefined rule curves (Klipsch et al., 2007; Yates et al., 2005), which determine release decision based on water availability, which in turn, depends on inflow and storage (Chen et al., 2022). However, many reservoirs are actively managed, where the flow releases are determined by reservoir managers to account for the complex tradeoffs among different operation targets. This complicated decision-making process often cannot be described with simple operation rules. In addition, observations on reservoir operation (e.g., reservoir water level and release) are very limited due the complex ownership and regulations.

As a result, reservoirs, as coupled natural-human systems (Liu et al., 2007), are not adequately represented in current hydrologic or hydraulic models. Compared to natural hydrologic processes that can be expressed by physical relationships, it remains unclear how reservoirs are operated to regulate streamflow, as observations on reservoir operation (e.g., reservoir water level and release) are very limited due the complex ownership and regulations. For example, the National Water Model is able to predict streamflow for over two million reaches in US, while a limited number of reservoirs are simulated by a simple level pool routing scheme (Gochis et al., 2018; Khazaei et al., 2021) where reservoir releases are passively determined by reservoir water level and spillway characteristics based on hydraulic laws (e.g., weir flow equations). However, for the actively managed reservoirs, which are a key infrastructure that involves various stakeholders and has significant impact on downstream flow, the releases are regulated by gates and determined by reservoir managers depending on various real-world constraints and tradeoffs.

Traditionally, studies have used optimization to derive reservoir operation rules. Specifically, optimal releases are determined to achieve predefined operation objective(s) (e.g., minimize flood risk, maximize water supply reliability, increase hydroelectricity generation) under various constraints (e.g., reservoir storage capacity and allowable downstream release). However, actual reservoir release usually deviates from the optimized prescription due to several limitations. First, the theoretical optimal reservoir releases are obtained under a small set of predefined objectives and constraints, which often do not capture the full spectrum of real-world operation conditions (Giuliani et al., 2021). Second, reservoir characteristics (storage capacity vs water level relationship) or streamflow regime may be different from the conditions when optimal operation rule was derived. Third, optimization models assume that perfect streamflow predictions or a known streamflow prediction uncertainty, but it is not necessarily the case that streamflow prediction is available for operational purposes and whether reservoir managers utilize the streamflow prediction during the decision-making processes (Zhao et al., 2011). Therefore, with these deviations from assumptions, optimization model-derived reservoir operation rules may provide valuable normative solutions for the large-scale hydrologic and water resource model, but often fail to yield satisfactory results for predicting streamflow downstream of reservoirs.

Data-driven models (DDMs) offer a promising alternative to derive reservoir operation rules from historical records of hydrologic and reservoir data (Lin et al., 2006; Wei and Hsu, 2008; Hipni et al., 2013; Aboutalebi et al, 2015; Yang et al., 2017; Zhang et al. 2018; Zhao and Cai, 2020; Turner et al., 2020a, b; Chen et al. 2022). The rationale is straightforward: if a manager determines the reservoir releases based on some principles (either empirical or optimal) depending on hydroclimatic variation, data-driven techniques can recover the patterns of operation from the reservoir records and other hydroclimatic variables. In addition, compared to optimization models DDMs are computationally efficient and readily coupled with hydrologic and hydraulic models. Recent studies (Mateo et al. 2014; Coerver, Rutten, and Van De Giesen, 2018; Yassin et al. 2019) have demonstrated the capability of various machine learning techniques in capturing reservoirs release decision.

In this study, we hypothesize that reservoirs operation patterns vary across time scales, thus requiring a hierarchical temporal scale configuration of DDMs. First, reservoirs usually have multiple operation purposes that require decisions made at different time scales. For example, daily or hourly release decisions are made for hydroelectricity generation based on the demand from power grids, while the reservoir storage for agricultural water supply exhibits slow-varying seasonal pattern. Even for reservoirs with one primary operation purpose, hydroclimatic variabilities at different time scales may lead to different operation decisions. A reservoir designed for flood control may be actively operated only during wet seasons to mitigate floods, and the storage may remain relatively stable during dry seasons. Second, release decisions for different operational purposes are made based on different information that changes with time scales. For example, flood control decisions may depend on current reservoir water level and streamflow forecast with leading time up to several days, while water supply reservoirs may ignore the short-term streamflow variability and focus on hydrologic seasonal dynamics such as snowpack. Third, operation decisions made at different scales interact with each other. The flood control hourly operations during a high flow event may be constrained water level set by seasonal water supply targets; flood control operations, in return, determine initial water level for water supply release for next decision period. Based on these observations, capturing the reservoir operation decisions across time scales is essential to accurately represent the anthropogenic regulation on streamflow variability.

However, current reservoir operations derived from DDMs are typically based on a single time scale. Zhang et al., (2018) assessed the performances of various DDMs with different time resolution (e.g., hourly, daily, and monthly) for Gezhouba Dam, while neglecting the interactions of decision-making processes across time scales. Yang et al. (2021) provided a comprehensive comparison of different DDMs to simulate the daily reservoir outflow over the Upper Colorado Region using the daily inflow, storage, and calendar time as model inputs, which do not include decision variables at monthly scales. Turner et al., (2020b) built a daily scale DDM for reservoirs in the Columbia River basins with seasonally varying relations that specify water release as a function of prevailing storage levels and forecasted future inflow. However, this approach is based on pre-assumed linear piecewise relations to represent the seasonality, which still needs to be specified based on modeler's assumption. A more flexible generic framework is needed to capture the tradeoffs among multiple reservoir operation targets and interactions between hydroclimatic conditions and anthropogenic decisions using information across time scales.

To fill this gap, this study develops a hierarchical temporal scale framework to model reservoir operation decisions across various time scales. The framework has the flexibility to (1) use time scale-specific inputs for DDMs to learn reservoir operation behaviors pertinent to each time scale, and (2) enable decisions at different time scales to interact with each other. We demonstrate the framework with a two-layer configuration, at monthly/weekly and daily scales, respectively. The framework is validated using the daily operational records of 327 major reservoirs in the United States regulated by the United States Army Corps of Engineers (USACE) and the United States Bureau of Reclamation (USBR). These reservoirs cover a wide spectrum of hydroclimatic conditions, reservoir characteristics and operation purposes, therefore can examine the robustness of the proposed hierarchical temporal scale framework. The monthly- or weekly-scale data-driven model learns reservoir decisions not affected by short-term variability and provides constraints for the daily scale model which captures the event-scale operation rule that deviates from the monthly/weekly average. This framework is flexible to incorporate additional temporal layers (such as at hourly or seasonal scales). We further evaluate which variables are dominant for reservoir operations across various time scales and investigate the tradeoff between training variables and modeling temporal resolution in representing reservoir decisions.

2 Methods

2.1 Hierarchical temporal scale configuration of DDMs

The hierarchical temporal scale framework (shown in Figure 1) consists of multiple layers, where each layer has a DDM to learn the reservoir operation rules at the corresponding time scale (e.g., monthly, weekly, and daily). The configuration starts from the upper layer corresponding to a coarse time scale (i.e., monthly/weekly in this study) to capture the reservoir operation behaviors under slow-varying targets (e.g., storing water for growing season irrigation supply). Historical hydroclimate and reservoir records are aggregated to monthly/weekly time series to train a DDM. The lower layer refines the model to a fine time scale (i.e., daily scale in this study), and a second DDM is trained to simulate the "residual", defined as the difference between the fine scale release and release simulated by the coarse time scale DDM. The residual characterizes short-term deviations from release determined under long-term operation targets and may be caused by gaps between planned and actual situations and complicated tradeoffs between various purposes served in different periods.

The hierarchical configuration of the framework is flexible to add layers as needed to represent operation decisions at coarser (e.g., seasonal) or finer time scales (e.g., flood control release or hydroelectricity generation under power grid demand) if reservoir operation record is available. In addition, the hierarchical framework allows models at each time scale to take different training variables since difference operations decisions may depend on different information. For example, the operation for irrigation water supply may mainly depend on the crop water demand during the growing season, while operation for flood control may depend on current reservoir water level and upstream flow predictions for the next a few days. By learning the residuals between water release at fine time scale and the coarse time scale average, the DDM can capture the interactions of operation rule at different time scale and represent the tradeoffs between various operation targets. For example, the release for flood control may be depended on current reservoir water level, which is affected by storage target for water supply determined one month ago. The reservoir water level after flood control release may further affect water supply decision in future time steps. Therefore, the residual between two layers (i.e., two time scales) represents the tradeoffs between various operation targets.

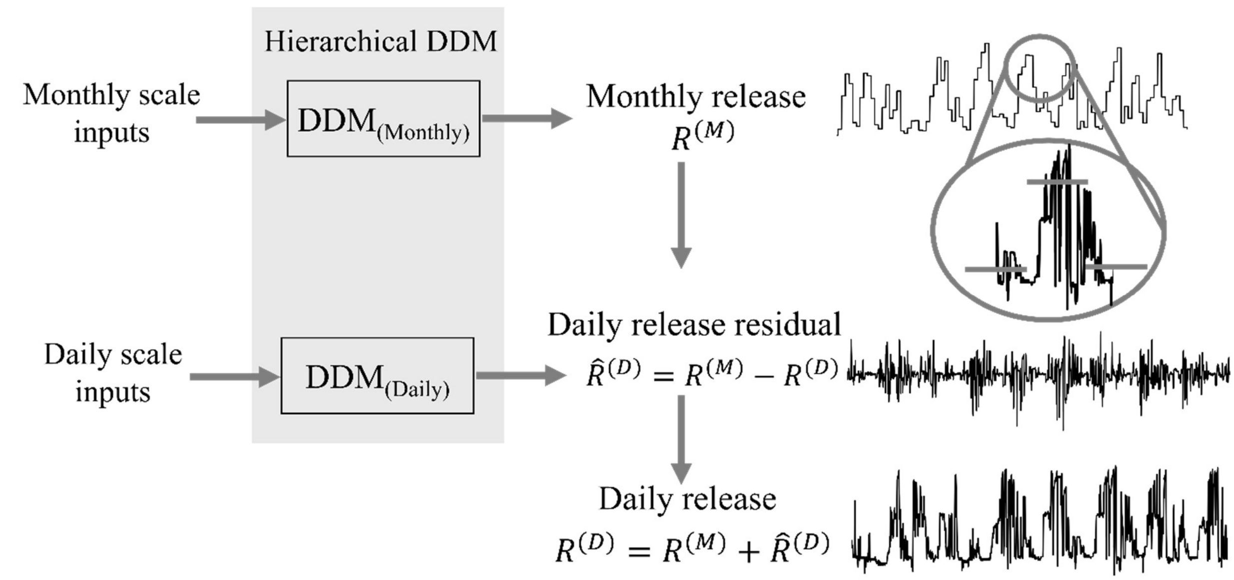

Figure 1. The hierarchical temporal scale framework with two layers shown for illustration. The top layer uses a monthly DDM to simulate monthly averaged release ($R^{(M)}$), and the subsequent bottom layer uses a daily DDM to simulate residual $\hat{R}^{(D)}$, or the difference between daily $R^{(D)}$ and monthly averaged $R^{(M)}$ releases.

2.2 Hydroclimatic and Reservoir Data

We apply the proposed framework to 248 reservoirs operated by the United States Army Corps of Engineers (USACE) and 79 reservoirs operated by the United States Bureau of Reclamation (USBR) across the Contiguous United States (CONUS). These reservoirs are generally actively managed reservoirs with multiple designed purposes. The standardized database for historical daily reservoir levels and operations of USACE reservoirs is developed by (Patterson and Doyle, 2018), while that of USBR reservoirs is accessed via Reclamation Information Sharing Environment (RISE). These observed records include daily reservoir water elevation (feet, ft), storage volume (acre-feet, af), inflow (cubic feet per second, cfs) and release (cubic feet per second, cfs) for each reservoir, with different record lengths and intermittent gaps in the middle of the record due to data collection issues. All reservoirs with continuous records are included in this study. For some reservoirs with missing data during only a short period of time (less than five days), the nearest neighbor interpolation method is applied to fill in these gaps to obtain a continuous record. Overall, the continuous records have the average length of 30 years.

The reservoir release data is used as target (response variable) to train and test the DDMs, and water storage volume, reservoir inflow records are used as inputs of the DDMs, along with hydroclimatic data. Specifically, the daily-scale meteorological forcing, including total precipitation rate (*P*, mm/day) and potential evapotranspiration (*PET*, mm/day) are obtained from the North American Land Data Assimilation System (NLDAS-2) forcing (Xia et al. 2012). The hydroclimatic data are aggregated over the catchment area upstream of the reservoir to encapsulate the local weather information relevant for reservoir operation. Specifically, the *PET* represents atmospheric demand for reservoir evaporative loss, which is substantial for reservoirs in the arid and semi-arid regions (Friedrich et al., 2018). The *P* may reflect the local runoff contribution to reservoir, while the reservoir inflow represents the runoff from the larger upstream contributing area. The difference between *P* and *PET* captures the crop irrigation water demand (Le Page et al., 2020), which may provide important information for reservoirs with irrigation water supply purpose. Depending on the specifics of a given reservoir, other information (e.g., hydroelectricity generation) can also be fed into DDMs as inputs.

2.3 Experimental Setup

Three groups of experiments are carried out to assess the performances of data-driven reservoir operation models with (1) under different time scale configurations and (2) different combinations of input variables (Table 1). The experimental setup is summarized in Table 1. The first group of experiments simulate reservoir release solely on a single daily scale (i.e., daily inputs are employed to model the daily release). This strategy is commonly implemented in existing machine-learning based reservoir models. The other two groups of experiments adopt a two-level hierarchical time scale framework. The second group of experiments receives weekly-average input variables in the first layer to generate weekly average release, and then use daily inputs to model the residual (difference between daily release and weekly average) in the second layer, herein referred to as “Weekly-Daily (WD)”. Similarly, the third group of experiments simulate monthly scale reservoir release in the first layer and refines reservoir release on daily scale in the second layer, referred to as “Monthly-Daily (MD)”. On the daily scale, we use the 7 days in the past and 7 days in the future of input variables to determine release on a given day. For the WD and MD models, the coarse-resolution input variables of the past 8 steps (weeks or months) and the future 4 steps are used to derive the release at the current time step, and the daily scale residuals are simulated with daily input variables of the past 7 days and the future 7 days. It has been proven that inflow forecasts could strongly influence the seasonal reservoir operations particularly for the high-elevation reservoirs fed by snowmelt in the western United States (Turner et al., 2020a). In this study, similar to Turner et al., (2020a), the observed records in the future time steps (i.e., perfect foresight) are used as a proxy for forecasted information are deployed to explore whether operators consider the streamflow forecasts during the decision-making processes, since it is difficult to acquire the actual forecasts available to operators at CONUS scale.

To explore the importance of each input variable for predicting reservoir operation at various time scales, the three experiment groups is further developed into six experiments with various combinations of input variables (Table 1). In Experiment 1, daily observed reservoir inflow (*I*), water storage (*S*), *P* and *PET* are all utilized to derive the release scheme. While other gain and loss terms in reservoir water budget (e.g., water diversion, seepage and evaporative loss) are unavailable for most reservoirs, the training variables may contain information related to these factors. For example, reservoir evaporative loss is related to *PET* and reservoir storage, which in turn correlates with water surface area. In Experiment 2, left out from the inputs to examine the importance of storage for estimating release. Similarly, in Experiment 3, reservoir inflow is not utilized. Meteorological information is hidden in Experiment 4, based on the assumption that the meteorological forcing may have no great impact on the reservoir release given storage and inflow. Experiment 5 derives the release scheme only from the observed inflow records. Experiment 6 explores whether the actual storage alone is able to capture reservoir release decisions. It is noted that based on the specified subset of inputs, DDMs will further infer the importance of these variables on predicting reservoir releases via the training process. Results of these experiments will be used to guide further sensitivity analysis based on models.

Table 1. Experiments using DDMs with different time scale configurations and subsets of input variables, including inflow (*I*), storage (*S*), precipitation (*P*) and potential evaporation (*PET*).

Time Scale	Experiment	Training variables
Daily (D)	D-1	*I, S, P, PET*
	D-2	*I, P, PET*
	D-3	*S, P, PET*
	D-4	*I, S*
	D-5	*I*
	D-6	*S*
Weekly-Daily (WD)	WD-1	*I, S, P, PET*
	WD-2	*I, P, PET*
	WD-3	*S, P, PET*
	WD-4	*I, S*
	WD-5	*I*
	WD-6	*S*
Monthly-Daily (MD)	MD-1	*I, S, P, PET*
	MD-2	*I, P, PET*
	MD-3	*S, P, PET*
	MD-4	*I, S*
	MD-5	*I*
	MD-6	*S*

In all the experiments, we use the Long Short-Term Memory (LSTM, Hochreiter and Schmidhuber, 1997), as the DDM in each layer. LSTM networks can learn temporal dependencies in both long and short terms and has a wide range of applications in hydrology and water resource management (Kratzert et al. 2018, 2019; Shen, 2018; Zhang et al. 2018; Feng et al., 2020; Sit et al., 2020; Xu and Liang, 2021; Yang et al. 2021). The configuration of the LSTM model in this study is summarized in Supplementary Material Text S1. For the single-layer models (D1, …, D6), the LSTM model is trained by minimizing the mean square error of daily release. For hierarchical time scale models (WD, MD), the two LSTMs are trained together by minimizing the mean square errors of reservoir release at both time scales,

$$\min_{\boldsymbol{\theta}} \sum\left(\boldsymbol{R}-\widehat{\boldsymbol{R}}\right)^2+\sum(\boldsymbol{r}-\widehat{\boldsymbol{r}})^2$$

where R and $\hat{R}$ are the observed and simulated release at the monthly/weekly scales, r and $\hat{r}$ are the observed and simulated release residuals at the daily scale, θ represents the neural network weiths. The Adam optimizer (Kingma et al., 2020) is applied for training. The Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) of daily reservoir release is used for assessing model performance in all experiments. To reduce random effects arising during training, we initialize and train the models for 5 times, each time using a different random seed, and calculate the average performance metrics across 5 trials. The number of training epochs, the optimal number of layers and number of hidden units are found through trial-and-error. 60% of time series data are used during the training process, 10% of them for validation, and the rest for testing. It is noted that the multi-layer configuration is flexible to use other machine learning algorithms.

3 Results

3.1 Performance of DDMs with various time scale configurations and input variable combinations

Results from the three groups of experiments revealed noticeable differences in reservoir release simulation accuracy when the models use various time scale configuration and combinations of input variables (Figure 2). For experiments using the same training variables, the two-layer hierarchical model (WD and MD) consistently yields higher accuracy than the daily model (D), as shown by the probability of exceedance of NSE for all reservoirs (Figure 2). For example, in Experiment 1 with most comprehensive input dataset, the mean NSE for all reservoirs is 0.949, 0.880 and 0.743 for WD, MD and daily configuration, respectively. The WD configuration achieves NSE higher than 0.9 in more than 92% reservoirs, compared to 54% and 18% for the MD and D configurations, respectively. In most experiments, the WD configuration yields slightly better performance than the MD configuration. For the same length of records, the weekly scale data is four times more than the monthly scale data, thus providing more training samples to the DDMs. In addition, the finer resolution of weekly scale may better capture the release decision than the coarse monthly scale.

For all time scale configurations, reservoir inflow and storage are two dominant variables for modelling release behavior in most reservoirs, as shown by the small performance gap between Experiments 1 and 4. With only reservoir inflow as input data in Experiment 5 (Figure 2e), the average NSE reaches 0.452, 0.561 and 0.535 for daily, WD and MD temporal configuration, respectively. The inflow provides most predictive power in reservoirs with relatively small storage and/or navigation purpose. Although the inflow-only models in Experiment 5 does not explicitly consider reservoir states, the LSTM architecture is able to use the cell "memory" to store accumulated inflow as a proxy for reservoir storage trend and use this information to simulate reservoir releases. However, due to other reservoir water budget terms such as water diversion, seepage and evaporative loss, the accumulated inflow cannot fully replace reservoir storage. Therefore, it is not ideal for a DDM to simulate the state of reservoir system without storage as an important constraint, especially for reservoirs in the west mountainous regions usually designed for water supply and hydropower generation. Because reservoir storage is closely related to the operational purposes, and its seasonal variations typically reflect the consequences of the human interventions on the natural system, storage volume (or water level) is strongly recommended as an independent variable input into the reservoir operation model.

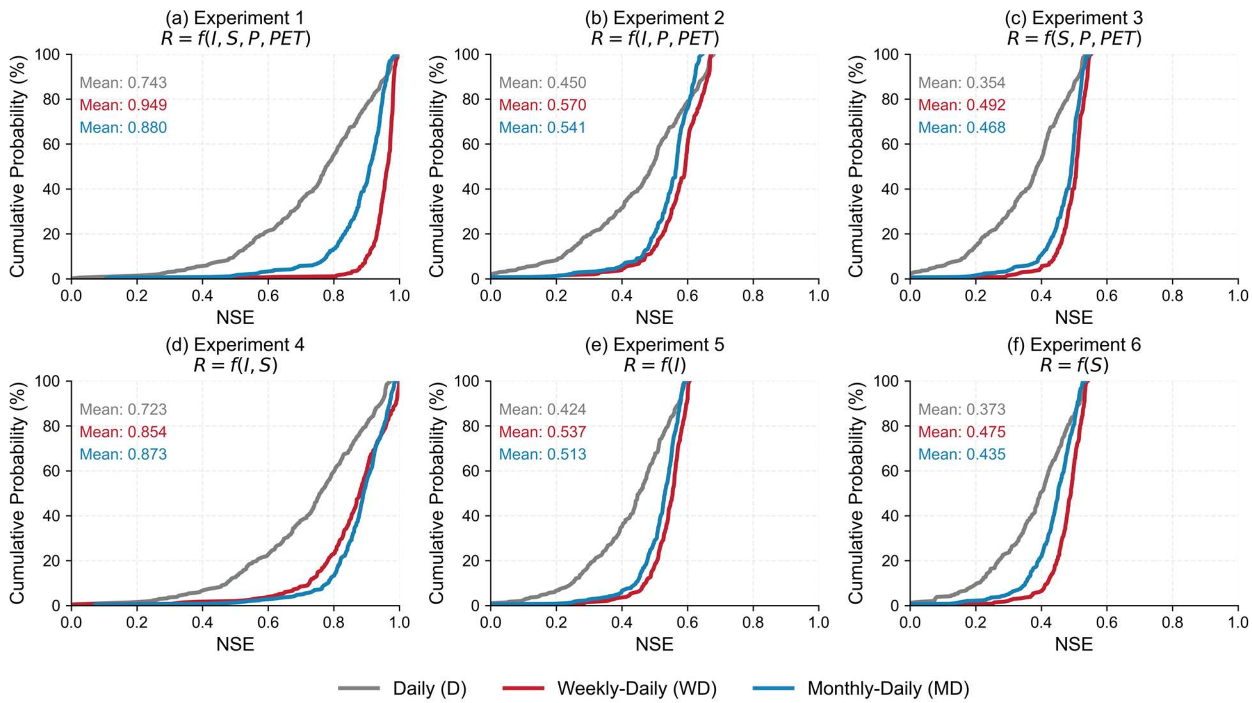

Figure 2. Probability of exceedance of NSE for all reservoirs resulting from single and hierarchical time scale models with different decision variables (Table 1)

The DDMs with storage alone as input in Experiment 6 have slightly lower predictive power compared to inflow-only models in Experiment 5 (Figure 2f) and produce average NSE of 0.373, 0.475 and 0.435 for Daily, WD and MD configuration, respectively. Using storage as the input captures operation of reservoirs with relatively large storage capacity and/or reservoir with water supply purpose where the release largely depends on the reservoir water level. In addition, reservoir storage serves as a proxy for reservoir water level and water surface area (both can be retrieved from the reservoir characteristic curve). The reservoir storage together with PET may implicitly contain information regarding reservoir evaporative loss, which is important in arid and semi-arid regions. Although storage-release rule curves are commonly used by reservoir operators (Yang et al. 2016), the seasonal patterns of reservoir operation and the interannual variability of inflow are missing in such curves. At monthly or seasonal scale, water control plans designed for specific purposes or hydroclimatic conditions that influence the upstream flow rate may exhibit low year to year variation within decades. At daily or sub-daily scale, however, reservoir inflow can vary a lot due to emergency events or weather fluctuations, especially for those reservoirs with complicated operational conflicts between multiple objectives or climate-sensitive reservoirs (such as reservoirs in the New England regions faced with potentially increasing flooding risks under the context of global warming). Although actual rule curves implemented by reservoir operators could provide substantial information to understand the decision-making process of water resource management, it does not adequately to represent the operation tradeoffs under various inflow conditions. Reservoir inflow should be considered as a paramount input while building data-driven operation models. Combining the inflow and storage in Experiment 4, the average NSE improves to 0.723, 0.854 and 0.873 for daily, WD and MD temporal configuration, respectively.

The performance improvement from including hydroclimatic variables (e.g., *P* and *PET*) is investigated by comparing accuracies of DDMs in Experiment 1 vs. 4, Experiment 2 vs. 5, and Experiment 3 vs. 6. For DDMs with only inflow (Experiment 2 vs. 5) or storage (Experiment 3 vs. 6), the improvement from additional hydroclimatic forcing is negligible (mean NSEs increase less than 0.04). For daily scale DDMs in Experiments 2 and 3, the overall performance even slightly downgrades when adding *P* and *PET*. When both inflow and storage are used (Experiment 1 vs. 4), adding *P* and *PET* enhances mean NSE from 0.723 to 0.743, from 0.873 to 0.880 and from 0.854 to 0.949 for daily, MD and WD configurations, respectively. It is noted that the NSE improvement is larger in the fine time scale WD configuration than the coarse time scale MD configuration, as the former can better represent the short-term variability in *P* and *PET*.

3.2 Effect of DDMs hierarchical temporal configuration on capturing reservoir operation behavior

After feeding the DDMs with dominant explanatory variables (e.g., inflow and storage), a better organization (i.e.., hierarchical temporal configuration) of the explanatory variables further enhances the performance. For example, in Experiment 4, re-arranging the training data in hierarchical configuration (WD and MD) improves the NSE by more than 20% compared to the single daily scale configuration, although the DDMs in this experiment contain the same amount of information. This highlights the benefits of incorporating the multi-temporal scale of reservoir behaviors into the configuration of DDM to capture the reservoir operation under various targets.

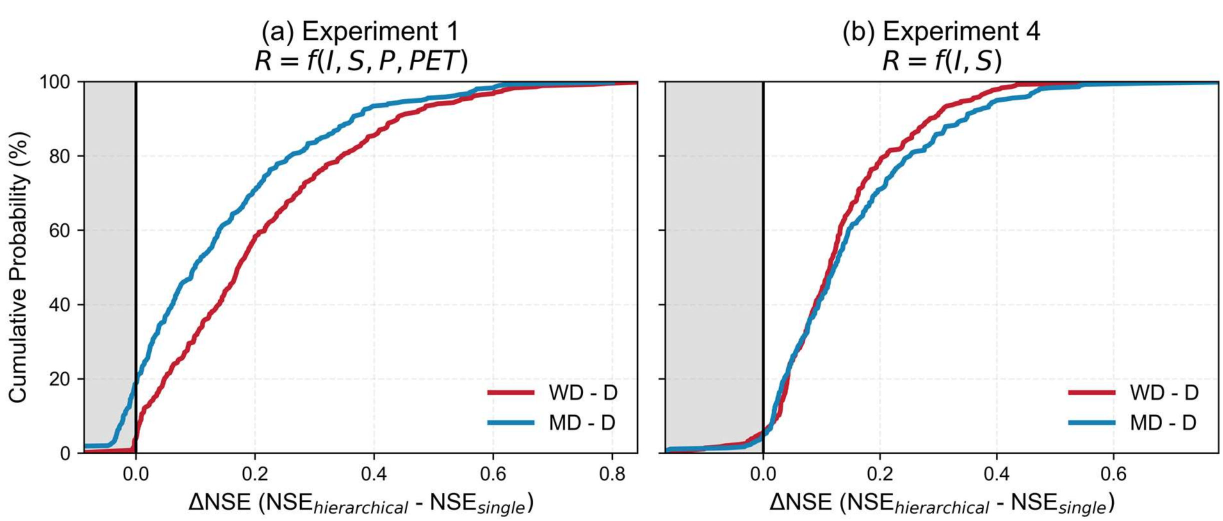

Figure 3. Improvement of NSE by hierarchical time scale framework ($NSE_{hierarchical}$ - NSE_{single}). $NSE_{hierarchical}$ represents the performances of hierarchical time scale models (WD, MD), while the NSE_{single} is the performance of a single time scale model (D). The difference

between Experiments 1 and 4 is, hydroclimatic forcing is employed in the former but is not in the latter.

Figure 3 further illustrates the improvement of the hierarchical framework for reservoir operation modeling and the nuances of such improvement with/without hydroclimatic information at different time scales. Most hierarchical temporal scale models under the same experiment settings perform better than the models constructed on the single time scale. Specifically, in Experiment 1 (Figure 3a) with the reservoir inflow, storage, *P* and *PET* as model inputs, performances of about 80% of reservoirs have been improved by hierarchical framework (MD), and it is more prominent for WD where the first layer simulates the reservoir release on the week scale. For 50% of reservoirs, MD with hydroclimatic forcing improves the NSEs by more than 0.1, and WD does by increasing more than 0.2 in model performances. In Experiment 4 (Figure 3b) without containing hydroclimatic forcing as model inputs, over 90% of the reservoir operation model gains a higher accuracy from the hierarchical architecture compared to the single daily scale model. In addition, there is negligible performance gain differences between hierarchical temporal configurations (e.g., WD and MD) if hydroclimatic forcing is not included (Figure 3b). While in Figure 3a with additional hydroclimatic forcing, the WD consistently improve NSEs by 0.1 for most reservoirs than the MD configuration. It indicates that hydroclimatic forcing (which shows significant short-term variability) contribute to the prediction for models with relatively fine temporal resolution (such as weekly).

3.3 Spatial pattern of DDM reservoir operation under various temporal configurations

Figure 4a, b and c show the spatial distribution of NSE for Daily, WD and MD configuration, respectively. The Daily scale model performance well (NSE higher than 0.95) for reservoirs along the Arkansas River (with navigation as primary purpose) and Columbia River (with hydroelectricity as primary purpose). These reservoirs are operated with single target based in the inflow (Figure 4a), which can be captured by a single daily scale model. Additional coarser weekly or monthly scale layer does not improvement the DDM performance (Figure 4d and e). The operation behavior in reservoirs in the Rocky Mountains cannot be captured by the Daily scale DDM, since these reservoirs are primarily operated for water supply and have slow-varying water storage. For other reservoirs, the Daily scale DDM produces NSE less than 0.7.

The MD scale DDM improves NSE over Daily scale in most reservoirs as shown in Figure 4e. The improvement is achieved as monthly scale release decision depends on different variables than the daily scale decision (Figure 4d). Reservoirs in the Rocky Mountains and California have the largest improvement, highlighting the signature of seasonal cycle of water supply operation in these reservoirs. Reservoirs on the High Plains (e.g., Texas, Oklahoma, Kansas) and in the Northeast do not have significant improvement with additional monthly layer. These reservoirs generally have both water supply and flood control purpose at the same time, suggesting that the monthly scale is too coarse to capture the tradeoff between flood control and water supply. For navigation reservoirs well represented by Daily scale DDM along the Arkansas River, adding a monthly scale (i.e., the MD model) even deteriorates the performance, as indicated by the negative NSE gain. This highlights the importance of identifying the appropriate modeling resolution to match the time scale at which reservoir release decisions are made.

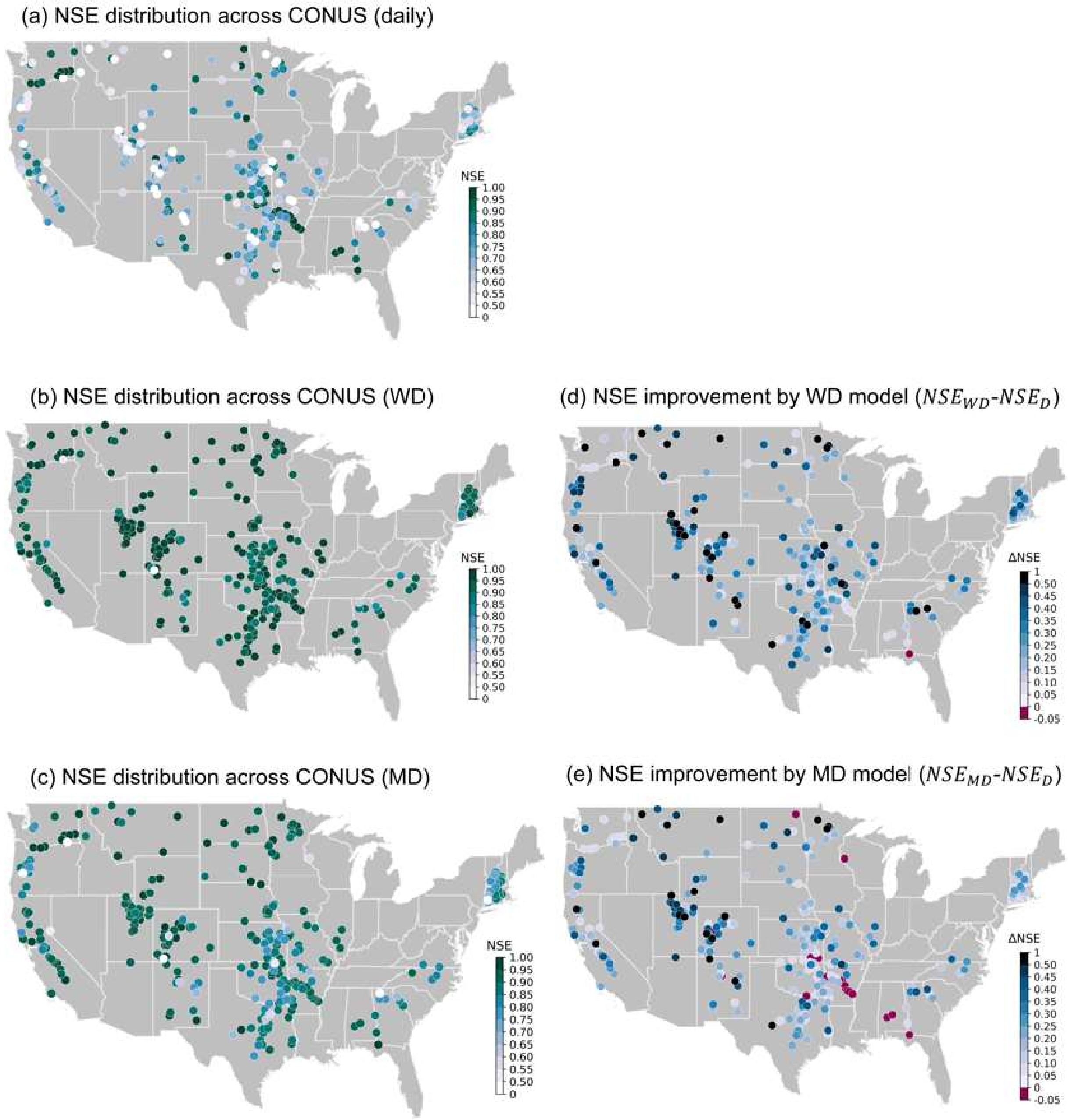

Figure 4. Spatial distribution of NSE of DDMs in Experiment 1 for a) Daily b) WD and c) MD configurations, and NSE improvement from Daily scale to d) WD and e) MD configuration.

In addition to similar improvement gained by MD scale DDM, the finer resolution WD scale configuration further improve the NSE for reservoirs on the High Plains (e.g., Texas, Oklahoma, Kansas) and in the Northeast (Figure 4d). The residual between daily release and weekly average release in the WD configuration is able to capture the coincided tradeoffs between water supply and flood control preferences.

3.4 Dominant variables of reservoir release across time scales

Although DDMs frequently achieve remarkable results in model performance, further sensitivity analysis would help to diagnose and interpret the empirical relations captured by the "black-box" DDMs. Different data-driven models have individual strengths and weaknesses in simulating the reservoir release, and few single models could consistently outperform others (Yang et al. 2021). Performances of different data-driven models can vary widely by the modeling schemes, by the ways of training data structure, as well as by the statistical measurement used. Model interpretability benefits further improvement in performance and providing insights on anthropogenic behaviors under hydroclimatic variabilities. The hierarchical configurations of DDMs allow us to explore whether reservoir operation depends on different variables and how the dominant variables change across time scales, thus revealing how tradeoffs among various reservoir operation targets are effectively captured by the DDMs.

We conducted sensitivity analysis based on the well-trained data-driven models to explore the impact of decision variables on reservoir release schemes across different time scales. For a certain variable, a small one-day perturbation (i.e., 5% increase) is imposed on the original dataset each time, and simulations are accordingly updated by the well-trained model with given the new inputs. Absolute fractional change in simulated release is averaged over time, then the variable that leads to the largest change is referred to as the most sensitive ones.

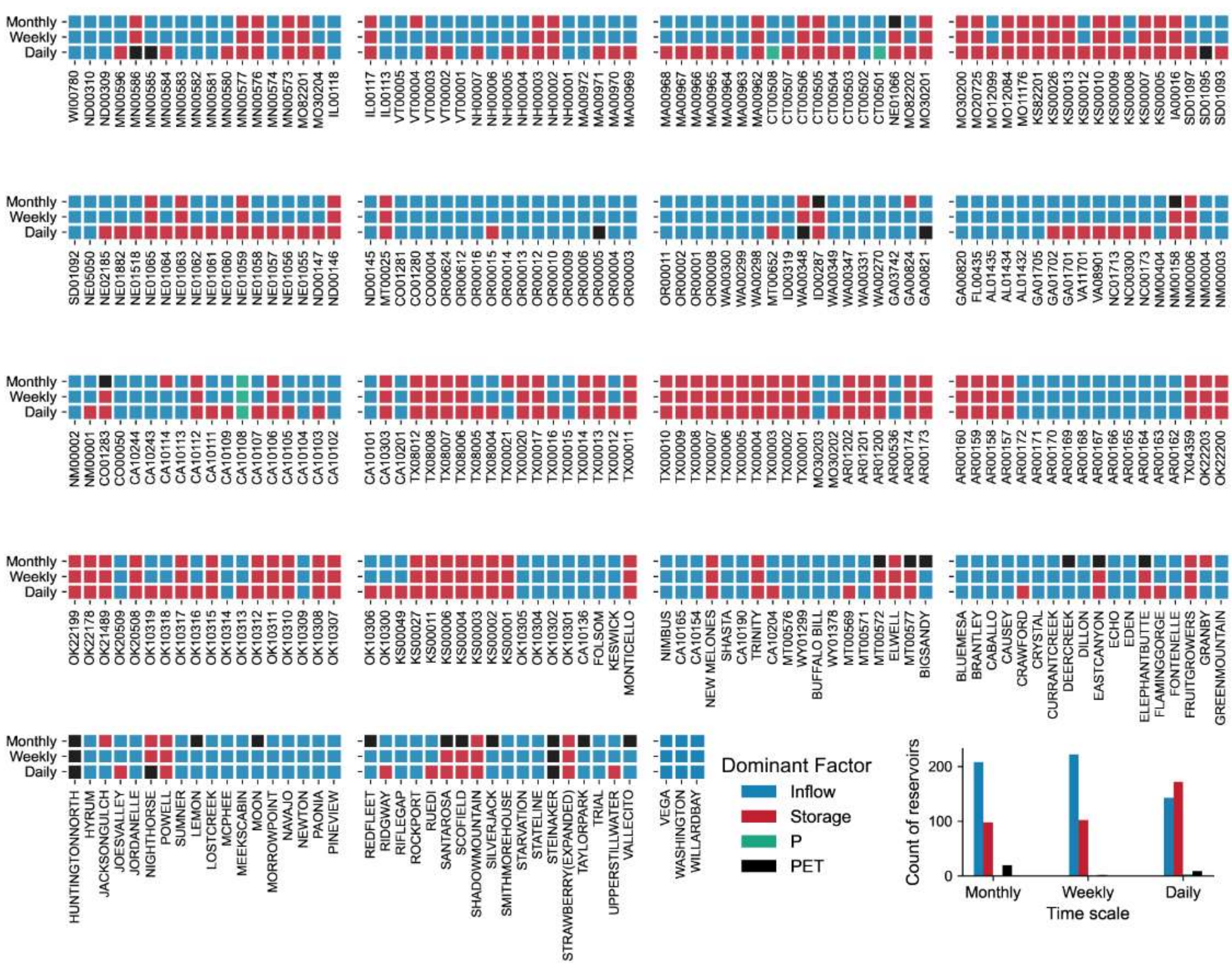

Figure 5. Dominant variable for reservoir operation at different time scales. The bar plot on the lower right corner shows the number of reservoirs with major reliance of reservoir release on the inflow, storage, *P* and *PET* across monthly, weekly and daily scales.

Figure 5 displays the most important variable for each reservoir across CONUS on the different time scales (daily, weekly and monthly). For most reservoirs (221 out of 327), the same variable has critical influences on the release on all time scales, likely implying the consistency of their operating strategies and trade-offs on various time scales, and there may be a primary purpose that dominates the operation process throughout the year. For 129 of these reservoirs, inflow play a decisive role in reservoir release at all time scales, while storage volume is the most instructive variable for 89 of these. It is worth mentioning that for two reservoirs located in Utah, the Huntington North Reservoir and the Steinaker Reservoir, *PET* has a major effect on reservoir release at the daily, weekly, and monthly scales, which could involve considerable reservoir evaporation and water use for agricultural irrigation in the arid, semi-arid western mountains. Only three reservoirs have *P* as the most predictive variable. These reservoirs are in headwater watershed with small contributing area. Therefore, the *P* is a good proxy for reservoir inflow. These results of model-based sensitivity analysis further validate the findings given by the comparison of Experiments 1 and 4. That is, reservoir inflow or storage volume has a paramount influence on the release decision rather than hydroclimatic forcing. Only for very few reservoirs, hydroclimatic forcing directly dominates the reservoir release.

It is interesting to notice that more than one third of (106 out of 327) reservoirs vary in their dependency on decision variables at different time scales, suggesting that reservoir operators consider different information at different time scales to fulfill multiple designed purposes. At the monthly scale, operations of 208 reservoirs primarily depend on the reservoir inflow, and 98 reservoirs rely more on storage volume. At the daily scale, the number of reservoirs with major dependency on inflow decreases to 143 and that of reservoirs relying more on storage volume increases to 172. From the coarse scale to the fine scale, more than 20% reservoirs (73 out of 327) shift their primary dependence from inflow to storage volume.

Figure 6 shows the spatial distribution of dominant factors across daily, weekly and monthly scale. Daily models with good performance (e.g., along the Arkansas River and Columbia River) generally identify inflow as the primary variable, as inflow exhibits high short-term variability and can effectively inform the daily release decision. The MD configuration captures the dependence of reservoirs (mostly located in the Rocky Mountains) monthly release decision on *PET* (Figure 4c), as reservoir are mainly operated for agricultural water supply.

The daily models identify storage as the primary variable in most reservoirs over the High Plains (Figure 4a) and capture the release's dependence on water level based on weir discharge equation. However, these reservoirs are also actively managed for flood control purpose, which is dependent on inflow condition. The failure to capture the flood control operation in the Daily model is corrected in the hierarchical temporal scale WD and MD, where some reservoirs' decision dependence changes from storage to inflow (Figure 4b, c).

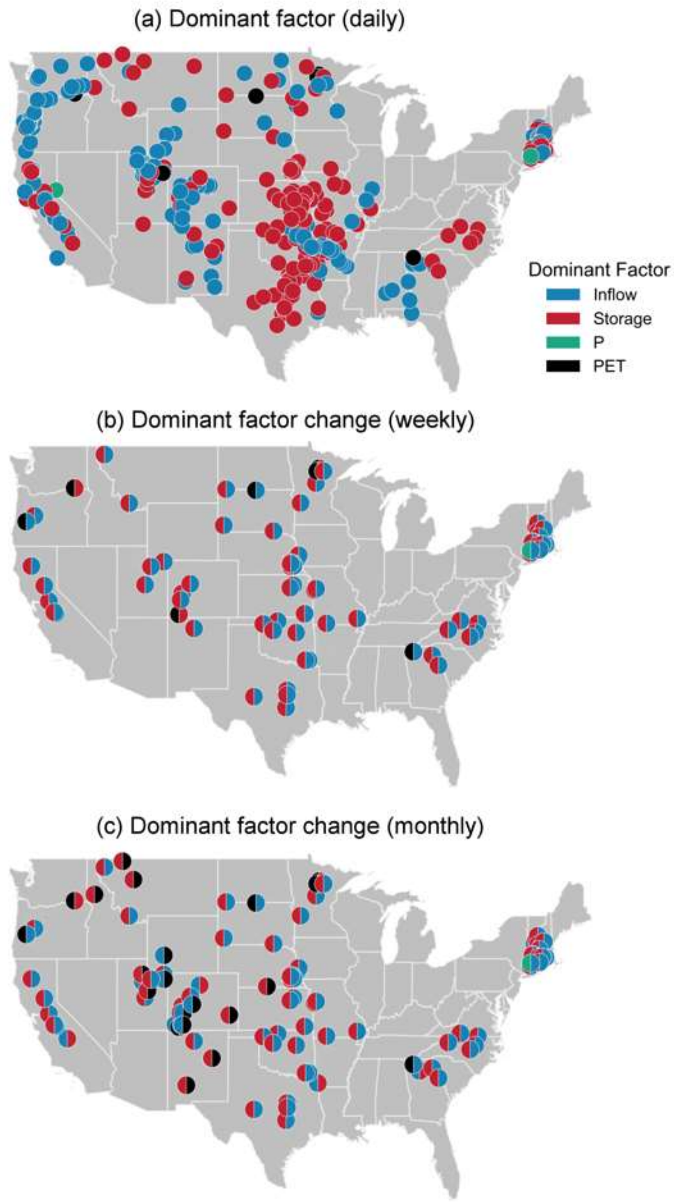

Figure 6. Spatial distribution of dominant factors across a) daily, b) weekly and c) monthly scale. The left half circle in b) and c) displays the major factor at the daily scale while the right half shows that at weekly or monthly scale, different from the daily dominated decision variable.

4 Discussion

4.1 Reservoir release behaviors across time scales

Compare to attempts to capture reservoir operation at a fixed time scale, the hierarchical temporal configuration in this study achieves better performance with the same input information. In addition, the sensitivity analysis suggests that operation in many reservoirs depends on different information at different time scales. In the following paragraphs, we picked the multi-purpose Belton Lake reservoir to elaborate how various operation targets manifest their signatures at different time scales, thus requiring hierarchical temporal configuration to fully capture the tradeoffs among multiple operation targets. The results of other reservoirs can be found at uploaded data.

The Belton Lake (TX00002) is located on Leon River in Texas with 536.8 million cubic meter (or 435,500 acre-feet) conservation capacity (Texas Water Development Board, 2015) and the maximum storage volume of around 1440 million cubic meters. The 192-feet high dam maintains the water level at elevation between the conservation pool elevation of 594 feet and the crest elevation of 631 feet, with flood control, water supply and irrigation as listed operation targets under the management of U.S. Army Corps of Engineers. The annual mean inflow volume is 641.5 million cubic meters. The Belton Lake provides an example with large storage capacity in humid subtropical climate. The DDM in Experiment 1 (with the most comprehensive training dataset) has NSE of 0.843, 0.971, 0.945 for Daily, WD and MD configuration, respectively. The DDM identifies reservoir storage as dominant variable on release at Daily, WD, and MD scales, respectively.

Figure 7 shows the scatter plots of release vs. inflow and storage vs. inflow at various time scales. At the annual time scale (Figure 7d), the outflow is highly correlated with inflow, suggesting the reservoir has seasonal flow regulating capacity. The slightly lower annual release than the inflow (Figure 7d) indicates water balance is roughly hold on annual time average. The randomness between monthly inflow and release (Figure 7c) shows a wide range during different seasons indicating the seasonal buffering capacity of the reservoir storage. The storage vs. release scatter plot shows reconcilable patterns starting from monthly scale. It is interesting to find several lines in the storage vs. release scatter plot at daily scale (Figure 7e), showing different number spillway gates are open during release events.

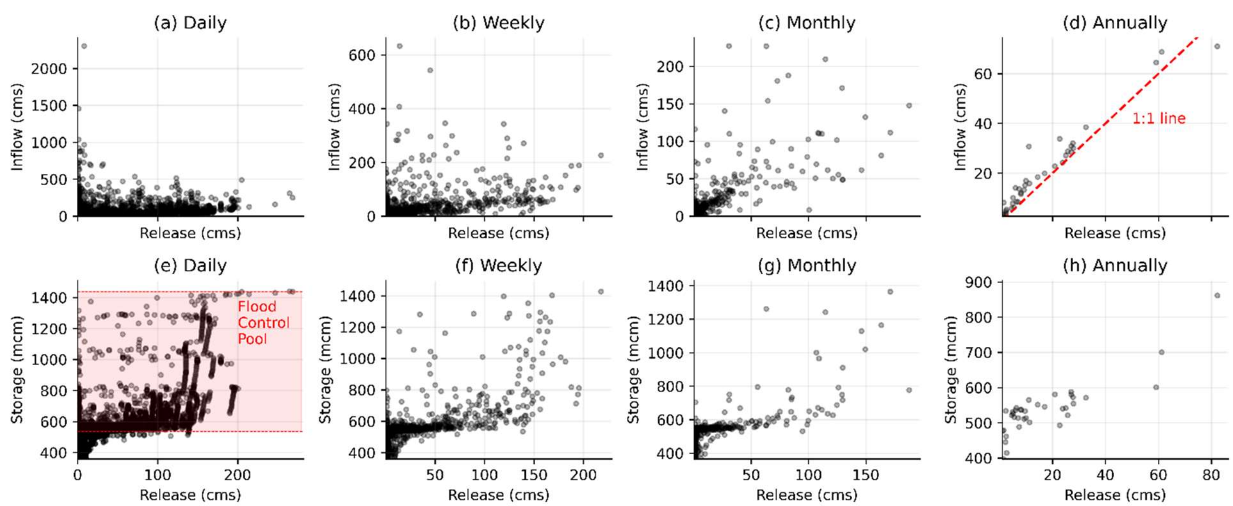

Figure 7. Relationship between inflow and release at a) daily, b) weekly, c) monthly, d) annual scale and; relationship between reservoir storage and release at e) daily, f) weekly, g) monthly, h) annual scale of Belton Lake (TX00002).

Figure 8a shows the flow duration curves of Belton Lake inflow and releases simulated by different DDM configurations. The Daily, WD and MD achieve similar predictability to capture the regulation during medium to high flow conditions (i.e., flow larger than 20% exceedance probability). The Daily scale DDM overestimates the low to medium flow range (i.e., flow less than 40% exceedance probability), and the MD scale DDM slightly overestimates the medium flow (i.e., flow between 25% and 45% exceedance) and underestimates the low flow range (i.e., flow less than 60% exceedance probability). The WD scale DDM reproduces the flow duration curve for almost all flow conditions although not perfectly.

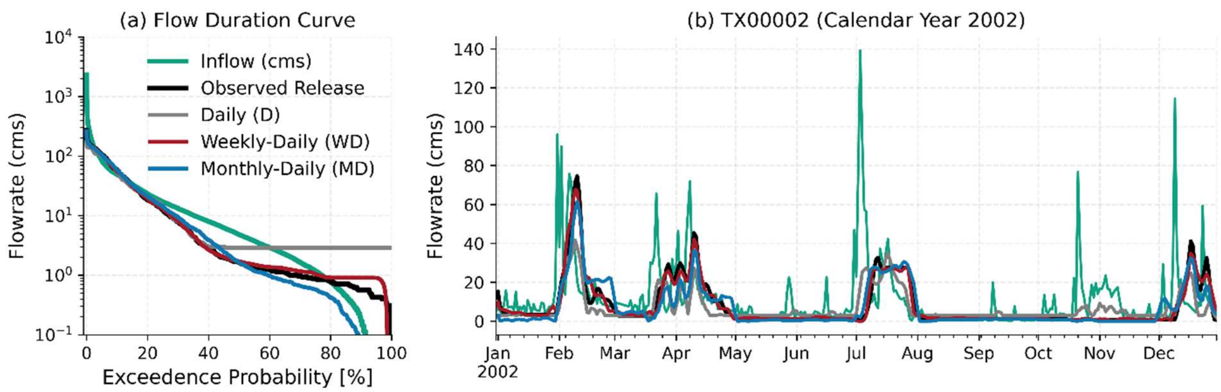

Figure 8. Inflow, observed release, release simulated by Daily (D), Weekly-Daily (WD) and Monthly-Daily (WD) models of Belton Lake (TX00002) shown in a) Flow Duration Curve (FDC) and b) hydrograph during the calendar year 2002.

The hydrograph of Year 2002 in Figure 8b shows the seasonal pattern and short-term variation produced by different DDM configurations. The MD DDM trends to have lower discharge and maintain the release longer after each flood event, as the monthly resolution in the upper layer is too large to capture the fast response under flood control purpose. The Daily scale DDM, on the other hand, trends to have faster decay of release after flood events, since the daily scale model is sensitive to the daily input and lacks the long-term information. The WD scale configuration works best to capture both seasonal water supply and flood control release at the Belton Lake.

These observations highlight the importance of appropriately organizing training data at various time scales in order to let machine learning techniques capture the underlaying relationships embedded at each time scale. We also used other machine learning techniques (e.g., random forest, support vector machine) to configure the hierarchical DDM and achieved satisfying results, suggesting the predictability is not limited by the choice of specific machine learning model. From the perspective of effectively training the machine learning models, hierarchical temporal configuration not only yields better predictability, but also provides more meaningful interpretation of the DDM.

4.2. Hierarchical nature of anthropogenic decisions

DDMs are generally not constrained by the complexity of training dataset and can achieve better prediction with more training variables. However, the experiments of hierarchical configuration comparison in Session 3.1 suggests that there exists a tradeoff between the number of training variables and time scales. Figure 3b shows that finer time scale (i.e., WD) configuration does not necessarily performance better than coarser configuration (i.e., MD), when only reservoir inflow and storage are used to train the DDM. After additional hydroclimatic variables are included in the training dataset, finer time scale configuration (WD) provides better predictability than the coarser configuration (i.e., MD) (Figure 3a). One possible reason is that the hydroclimatic variables contain short-term temporal variability that is necessary to improve the fine scale configuration. Furthermore, it indicates that reservoir operation decisions under different operation targets are associated with different time scales and require different information. Therefore, simply including more variables into the training datasets or increasing the hierarchical layers does not guarantee better predictability. This observation highlights the importance of providing appropriate information that matches the temporal resolution to capture reservoir release behavior under various targets.

Although the scaling issue in hydrologic processes has been well recognized by hydrologic community, there are few studies to investigate the scaling of decision making in water resources management. In representing anthropogenic components (by either simulation or optimization approach) in hydrologic models, the decision makings are generally based on one single time scale. For example, farmers' irrigation decision depends on soil moisture condition. The reservoir operation policy is optimized to balance the tradeoff between water supply benefits and flood risk based on daily streamflow. The hierarchical temporal scale configuration of DDM in this study explicitly shows that the single temporal scale model cannot fully capture the reservoir release under various operation targets. Different operation targets are associated with different temporal scale and require corresponding hydroclimatic information. For example, the reservoirs in the Colorado River Basin uses the seasonal snowpack condition to forecast the water supply (Xiao et al., 2018; Bureau of Reclamation, 2022), while the hydroelectric generation is based on hourly demands from power grids.

Beside the dependence on cross-scale information, anthropogenic decisions also interact at different scales. Short-term decisions (e.g., operation of water resources infrastructure) are constrained by long-term decisions (e.g., planning of water resources infrastructure), and the objectives of decisions at different scale may require tradeoff. For example, given the same amount of agricultural water supply, farmers can tradeoff between crop type and irrigated area (decisions made before growing season) and the actual irrigation intensity (decisions made during growing season), which results in different water release amount and frequency. The hierarchical temporal configuration of DDM in this study recognizes cross-scale interaction feature and handles this feature by simulating the daily release deviation from the weekly/monthly release. For traditional optimization formulation in water resources management, we believe the hierarchical optimization (Yeo et al., 2007; Karsanina et al., 2018) would be a promising configuration to represent interaction of decisions made across scales.

As hydrologic models and observations continue to improve and provide better prediction, the ultimate question is how hydrologic prediction (and what types of prediction) can be effectively utilized to improve the operation of reservoirs. Although we find very limited case where hydrologic forecast is used in operation in the 300 reservoirs, there are efforts to explore the reservoir operation using streamflow prediction (Delaney et al., 2020; Zarei et al., 2021). Hydrologic predictions at different time scales are based on different processes (e.g., seasonal projection based on snow water storage, short-term prediction based on weather forecast) and subject to various level of uncertainty. However, it remains challenging to have a consistent framework to integrate uncertainties from predictions across scales to inform decision makers on the tradeoffs among various reservoir operation targets.

5 Conclusions

In this study, we proposed a hierarchical temporal scale framework to improve the data-driven reservoir operation modeling. With observed inflow, storage, precipitation and potential evapotranspiration as inputs, more than 80% of reservoirs across the CONUS gain the improvement in model performances, while modeling of 90% of them can be more accurate by this framework if there is no hydroclimatic forcing.

This hierarchical framework is not model specific and therefore has broad applicability. By further adjusting the primary states simulated on the first coarse scale, which is partially similar to the operating process of reservoir managers in response to the daily inflow corresponding to the predefined water control plans, the hierarchical architecture is conducive to improve both the performances and the interpretability of data-driven models, and can be further adapted to be closely integrated with the decision-making of managers. It also demonstrates the similarity of a natural-human system and hydrologic processes across temporal scales. In future work, deep learning-based reservoir components can be embedded in physics-based models for more accurate hydrological process simulation.

Results of different experiment settings reveal that reservoir inflow and storage volume have a paramount influence on the release strategies. Model-based sensitivity analysis proves it, and further illustrates that variable importance can vary in time periods and across multiple time scales. For nearly 1/3 reservoirs across the CONUS, reservoir operations mainly depend on different decision variables at different time scales, and for several reservoirs, such as some in the Upper Colorado and Sacramento districts, hydroclimatic forcing still has major influence on the release, addressing more demands on the assessment and planning of reservoir status and accurate forecasting of hydroclimatic forcing.

Availability Statement

All data used in this research are publicly available. The meteorologic forcing (precipitation and potential evapotranspiration) is available at https://ldas.gsfc.nasa.gov/nldas/v2/forcing. The dataset of reservoir operations utilized in this study is available online (https://www.hydroshare.org/resource/79c262b627fc4ce293379b5e95457146/).

Acknowledgments

This research was made possible in part by the ASU ENGR-SEBE Startup grant.

References

Aboutalebi, M., Bozorg Haddad, O., & Loáiciga, H. A. (2015). Optimal monthly reservoir operation rules for hydropower generation derived with SVR-NSGAII. *Journal of Water Resources Planning and Management*, *141*(11), 04015029.

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

Blair, P., & Buytaert, W. (2016). Socio-hydrological modelling: a review asking" why, what and how?". *Hydrology and Earth System Sciences*, *20*(1), 443-478.

Boulange, J., Hanasaki, N., Yamazaki, D., & Pokhrel, Y. (2021). Role of dams in reducing global flood exposure under climate change. *Nature communications*, *12*(1), 1-7.

Brekke, L. D., Maurer, E. P., Anderson, J. D., Dettinger, M. D., Townsley, E. S., Harrison, A., & Pruitt, T. (2009). Assessing reservoir operations risk under climate change. *Water Resources Research*, *45*(4).

Bureau of Reclamation. (2022). *2022 Annual Operating Plan for Colorado River Reservoirs.*

Chen, W., & Olden, J. D. (2017). Designing flows to resolve human and environmental water needs in a dam-regulated river. *Nature communications*, *8*(1), 1-10.

Chen, Y., Li, D., Zhao, Q., & Cai, X. (2022). Developing a generic data-driven reservoir operation model. *Advances in Water Resources*, *167*, 104274.

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., ... & Zeng, X. (2015). Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research*, *51*(8), 5929-5956.

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.

Condon, L. E., & Maxwell, R. M. (2019). Simulating the sensitivity of evapotranspiration and streamflow to large-scale groundwater depletion. *Science Advances*, *5*(6), eaav4574.

Delaney, C. J., Hartman, R. K., Mendoza, J., Dettinger, M., Delle Monache, L., Jasperse, J., ... & Evett, S. (2020). Forecast informed reservoir operations using ensemble streamflow predictions for a multipurpose reservoir in Northern California. *Water Resources Research*, *56*(9), e2019WR026604.

Denaro, S., Anghileri, D., Giuliani, M., & Castelletti, A. (2017). Informing the operations of water reservoirs over multiple temporal scales by direct use of hydro-meteorological data. *Advances in water resources*, *103*, 51-63.

DeNeale, S. T., Baecher, G. B., Stewart, K. M., Smith, E. D., & Watson, D. B. (2019). *Current state-of-practice in dam safety risk assessment* (No. ORNL/TM-2019/1069). Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States).

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.

Ehsani, N., Vörösmarty, C. J., Fekete, B. M., & Stakhiv, E. Z. (2017). Reservoir operations under climate change: Storage capacity options to mitigate risk. *Journal of Hydrology*, *555*, 435-446.

Feng, D., Fang, K., & Shen, C. (2020). Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resources Research*, *56*(9), e2019WR026793.

Ferguson, I. M., & Maxwell, R. M. (2011). Hydrologic and land–energy feedbacks of agricultural water management practices. *Environmental Research Letters*, *6*(1), 014006.

Forsberg, B. R., Melack, J. M., Dunne, T., Barthem, R. B., Goulding, M., Paiva, R. C., ... & Weisser, S. (2017). The potential impact of new Andean dams on Amazon fluvial ecosystems. *PloS one*, *12*(8), e0182254.

Friedrich, K., Grossman, R. L., Huntington, J., Blanken, P. D., Lenters, J., Holman, K. D., ... & Kowalski, T. (2018). Reservoir evaporation in the Western United States: current science, challenges, and future needs. *Bulletin of the American Meteorological Society*, *99*(1), 167-187.

Graf, W. L. (1999). Dam nation: A geographic census of American dams and their large-scale hydrologic impacts. *Water resources research*, *35*(4), 1305-1311.

Giuliani, M., Lamontagne, J. R., Reed, P. M., & 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.

Gochis, D. J., Barlage, M., Dugger, A., FitzGerald, K., Karsten, L., McAllister, M., ... & Yu, W. (2018). The WRF-Hydro modeling system technical description,(Version 5.0). *NCAR Technical Note*, *107*.

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.

Hipni, A., El-shafie, A., Najah, A., Karim, O. A., Hussain, A., & Mukhlisin, M. (2013). Daily forecasting of dam water levels: comparing a support vector machine (SVM) model with adaptive neuro fuzzy inference system (ANFIS). *Water resources management*, *27*(10), 3803-3823.

Ho, M., Lall, U., Allaire, M., Devineni, N., Kwon, H. H., Pal, I., ... & Wegner, D. (2017). The future role of dams in the United States of America. *Water Resources Research*, *53*(2), 982-998.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Khazaei, B., Read, L. K., Casali, M., Sampson, K. M., & Yates, D. (2021). Improvement of Lake and Reservoir Parameterization in the NOAA National Water Model. In *World Environmental and Water Resources Congress 2021* (pp. 552-560).

Karsanina, M. V., & Gerke, K. M. (2018). Hierarchical optimization: Fast and robust multiscale stochastic reconstructions with rescaled correlation functions. *Physical review letters*, *121*(26), 265501.

Kiefer, J., & Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 462-466.

Kim, J., Read, L., Johnson, L. E., Gochis, D., Cifelli, R., & Han, H. (2020). An experiment on reservoir representation schemes to improve hydrologic prediction: Coupling the national water model with the HEC-ResSim. *Hydrological Sciences Journal*, *65*(10), 1652-1666.

Kingma Diederik, P., & Adam, J. B. (2014). A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Klipsch, J. D., and M. B. Hurst. (2007). HEC-ResSim reservoir system simulation user's manual version 3.0. *USACE, Davis, CA* 512.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005-6022.

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

Lane, N. (2007). *Aging infrastructure: Dam safety*. Congressional Research Service.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., ... & Wisser, D. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*(9), 494-502.

Le Page, M., Fakir, Y., Jarlan, L., Boone, A., Berjamy, B., Khabba, S., & Zribi, M. (2021). Projection of irrigation water demand based on the simulation of synthetic crop coefficients and climate change. *Hydrology and Earth System Sciences*, *25*(2), 637-651.

Li, C., Sun, G., Caldwell, P. V., Cohen, E., Fang, Y., Zhang, Y., ... & Meentemeyer, R. K. (2020). Impacts of urbanization on watershed water balances across the conterminous United States. *Water Resources Research*, *56*(7), e2019WR026574.

Li, X., Guo, S., Liu, P., & Chen, G. (2010). Dynamic control of flood limited water level for reservoir operation by considering inflow uncertainty. *Journal of hydrology*, *391*(1-2), 124-132.

Lin, J. Y., Cheng, C. T., & Chau, K. W. (2006). Using support vector machines for long-term discharge prediction. *Hydrological sciences journal*, *51*(4), 599-612.

Liu, J., Dietz, T., Carpenter, S. R., Folke, C., Alberti, M., Redman, C. L., ... & Provencher, W. (2007). Coupled human and natural systems. *AMBIO: a journal of the human environment*, *36*(8), 639-649.

Mateo, C. M., Hanasaki, N., Komori, D., Tanaka, K., Kiguchi, M., Champathong, A., ... & Oki, T. (2014). Assessing the impacts of reservoir operation to floodplain inundation by combining hydrological, reservoir management, and hydrodynamic models. *Water Resources Research*, *50*(9), 7245-7266.

Moran, E. F., Lopez, M. C., Moore, N., Müller, N., & Hyndman, D. W. (2018). Sustainable hydropower in the 21st century. *Proceedings of the National Academy of Sciences*, *115*(47), 11891-11898.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, *10*(3), 282-290.

Ortiz-Partida, J. P., Lane, B. A., & Sandoval-Solis, S. (2016). Economic effects of a reservoir re-operation policy in the Rio Grande/Bravo for integrated human and environmental water management. *Journal of Hydrology: Regional Studies*, *8*, 130-144.

Oudin, L., Salavati, B., Furusho-Percot, C., Ribstein, P., & Saadi, M. (2018). Hydrological impacts of urbanization at the catchment scale. *Journal of Hydrology*, *559*, 774-786.

Palmer, M., & Ruhi, A. (2019). Linkages between flow regime, biota, and ecosystem processes: Implications for river restoration. *Science*, *365*(6459), eaaw2087.

Patterson, L. A., & Doyle, M. W. (2018). A nationwide analysis of US Army Corps of Engineers reservoir performance in meeting operational targets. *JAWRA Journal of the American Water Resources Association*, *54*(2), 543-564.

Shen, Chaopeng. "A transdisciplinary review of deep learning research and its relevance for water resources scientists." *Water Resources Research* 54, no. 11 (2018): 8558-8593.

Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Döll, P., & Portmann, F. T. (2010). Groundwater use for irrigation–a global inventory. *Hydrology and earth system sciences*, *14*(10), 1863-1880.

Simonovic, S. P. (1992). Reservoir systems analysis: closing gap between theory and practice. *Journal of water resources planning and management*, *118*(3), 262-280.

Singh, N. K., & Basu, N. B. (2022). The human factor in seasonal streamflows across natural and managed watersheds of North America. *Nature Sustainability*, *5*(5), 397-405.

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, *82*(12), 2635-2670.

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), 1-8.

Suen, J. P., & Eheart, J. W. (2006). Reservoir management to balance ecosystem and human needs: Incorporating the paradigm of the ecological flow regime. *Water resources research*, *42*(3).

Texas Water Development Board. (2015). *Volumetric and Sedimentation Survey of Belton Lake*.

Tonkin, J. D., Merritt, D., Olden, J. D., Reynolds, L. V., & Lytle, D. A. (2018). Flow regime alteration degrades ecological networks in riparian ecosystems. *Nature ecology & evolution*, *2*(1), 86-93.

Turner, S. W., Xu, W., & Voisin, N. (2020). Inferred inflow forecast horizons guiding reservoir release decisions across the United States. *Hydrology and Earth System Sciences*, *24*(3), 1275-1291.

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.

Wei, C. C., & Hsu, N. S. (2008). Derived operating rules for a reservoir operation system: Comparison of decision trees, neural decision trees and fuzzy decision trees. *Water resources research*, *44*(2).

Wei, S., Xu, T., Niu, G. Y., & Zeng, R. (2022). Estimating Irrigation Water Consumption Using Machine Learning and Remote Sensing Data in Kansas High Plains. *Remote Sensing*, *14*(13), 3004.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., ... & Mocko, D. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, *117*(D3).

Xiao, M., Udall, B., & Lettenmaier, D. P. (2018). On the causes of declining Colorado River streamflows. *Water Resources Research*, *54*(9), 6739-6756.

Yang, T., Asanjan, A. A., Welles, E., Gao, X., Sorooshian, S., & Liu, X. (2017). Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. *Water Resources Research*, *53*(4), 2786-2812.

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, T., Zhang, L., Kim, T., Hong, Y., Zhang, D., & Peng, Q. (2021). A large-scale comparison of Artificial Intelligence and Data Mining (AI&DM) techniques in simulating reservoir releases over the Upper Colorado Region. *Journal of Hydrology*, *602*, 126723.

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.

Yates, D., Sieber, J., Purkey, D., & Huber-Lee, A. (2005). WEAP21—A demand-, priority-, and preference-driven water planning model: part 1: model characteristics. *Water international*, *30*(4), 487-500.

Yeo, I. Y., Guldmann, J. M., & Gordon, S. I. (2007). A hierarchical optimization approach to watershed land use planning. *Water resources research*, *43*(11).

Zarei, M., Bozorg-Haddad, O., Baghban, S., Delpasand, M., Goharian, E., & Loáiciga, H. A. (2021). Machine-learning algorithms for forecast-informed reservoir operation (FIRO) to reduce flood damages. *Scientific reports*, *11*(1), 1-21.

Zeng, R., Cai, X., Ringler, C., & Zhu, T. (2017). Hydropower versus irrigation—an analysis of global patterns. *Environmental Research Letters*, *12*(3), 034006.

Zeng, R., & Ren, W. (2022). The spatiotemporal trajectory of US agricultural irrigation withdrawal during 1981-2015. *Environmental Research Letters*.

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., ... & Zhuang, J. (2018). Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of Hydrology*, *565*, 720-736.

Zhao, Q., & Cai, X. (2020). Deriving representative reservoir operation rules using a hidden Markov-decision tree model. *Advances in Water Resources*, *146*, 103753.

Zhao, T., Cai, X., & Yang, D. (2011). Effect of streamflow forecast uncertainty on real-time reservoir operation. *Advances in water resources*, *34*(4), 495-504.

Zhao, Y., Liu, S., & Shi, H. (2021). Impacts of dams and reservoirs on local climate change: a global perspective. *Environmental Research Letters*, *16*(10), 104043.