A Hierarchical Temporal Scale Framework for Data-driven Reservoir Release 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 release modeling and validated across Contiguous United States.
- Reservoir release simulation would benefit from leveraging the availability of explanatory variables or comprehensive utilization of multiple temporal scales.
- The effects of decision variables on reservoir operations vary across time scales.

Abstract

As an important anthropogenic interference in 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 provide a promising approach to capture relationships embedded in historical records. Similar to hydrologic processes that vary across temporal scales, reservoir operations 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. Data-driven models were constructed to simulate reservoir releases 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 could compensate for the absence of key explanatory variables as model inputs, thereby efficiently capturing the release decisions of reservoirs situated in the west. 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 commonly face 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 rules, 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 need 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 modeling the reservoir release decisions, which often involve 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, the releases from actively managed reservoirs, which are crucial infrastructure involving multiple stakeholders and with significant downstream impacts, are regulated by gates and determined by reservoir managers based on a range of real-world constraints and trade-offs.

Traditionally, reservoir operation rules have been derived using optimization techniques. These models aim to determine optimal releases to achieve predefined objectives (such as minimizing flood risk or maximizing water supply reliability) under various constraints (such as 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). Recent studies have demonstrated the capability of various machine learning techniques in capturing reservoir release decisions (Mateo et al. 2014; Coerver, Rutten, and Van De Giesen, 2018; Yassin et al. 2019; Chen et al. 2022; Gangrade et al., 2022; Dong et al., 2023). 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 models 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. The primary motivation behind this study is to contribute to the development of simulation strategies that can enhance the representation of reservoirs in regional or national scale hydrological models, such as the National Water Model.

In this study, we hypothesize that reservoir 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 a 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 the 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.

Despite significant progress in data-driven reservoir modeling, current approaches typically rely on a single time scale for operations, with limited exploration of frameworks that account for multi-timescale interactions. For instance, 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 did not completely 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 the modeler's assumption. While single-scale models may adequately serve the needs of reservoir operators, investors, and decision makers for simpler reservoir systems, multi-objective reservoirs and multi-reservoir systems demand greater attention to the full range of timescales for improved reservoir operation modeling. The study conducted by Hejazi et al. (2008) using weekly/monthly datasets revealed that the importance of hydrologic indicators varies across seasons and purposes (i.e., flood control, water supply, hydropower, and irrigation) for reservoirs located in the California and Great Plains regions. It highlighted the interdependence between decision variables, purposes and time scales in reservoir operations. The time-varying sensitivity analysis at daily scale for a multi-reservoir system in the Red River Basin further illustrated that effective operating policies adapt the utilization of information over time while coordinating it across multiple reservoirs (Quinn et al., 2019). The challenges arise when simulating regulated flow downstream of such complex reservoirs. A general and flexible framework is needed, which can effectively simulate the reservoir release decisions and capture trade-offs among multiple reservoir operation objectives, as well as the interactions between hydroclimatic conditions and human decisions across various time scales. Furthermore, this framework is expected to be readily compatible with large-scale hydrologic and water resource management models.

This study develops a hierarchical temporal scale framework to model reservoir operation decisions across various time scales. The proposed framework exhibits generality in several aspects: (1) it does not require prior knowledge of reservoir operation objectives; (2) it supports the implementation of diverse data-driven modeling techniques; and (3) it utilizes commonly available datasets for training the machine learning models. 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-/weekly-scale data-driven model learns reservoir decisions unaffected 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

This study models reservoir release schemes at each temporal scale (e.g., daily, weekly, monthly) collectively under a set of hydroclimatic explanatory variables (e.g., streamflow, precipitation). We separate the raw daily time series into a coarse time-scale averages (i.e., monthly as illustrated in the example in Figure 1) and a fine-scale "deviation". The "deviation" $\hat{y}^{(D)}$ between daily scale release $y^{(D)}$ from the monthly scale release $y^{(M)}$ is defined as

$$\hat{y}^{(D)} = y^{(M)} - y^{(D)}$$

The "deviation" $\hat{y}^{(D)}$ includes (1) true signals (systematic bias or structured error) resulted from fine time scale reservoir release deviating from a coarse time scale operation (e.g., operation for daily release for flood control constrained by monthly water storage target for water supply) and (2) unstructured random error (e.g., Gaussian type random noise from measurement error). We hypothesize that the structured error between different time scales of observed release contains information that is not adequately represented at a single time scale, which can be effectively modeled using a hierarchical approach. For example, we found that temporal autocorrelation of the deviations of reservoir releases between daily and weekly/monthly scales exists in most of the reservoirs, probably indicating that relying solely on monthly/weekly averages may not fully capture the intricacies of reservoir release dynamics. This study utilizes coarse-scale averages as a source of long-term information to compensate for the limited forward-looking capacity of fine-scale limited time steps.

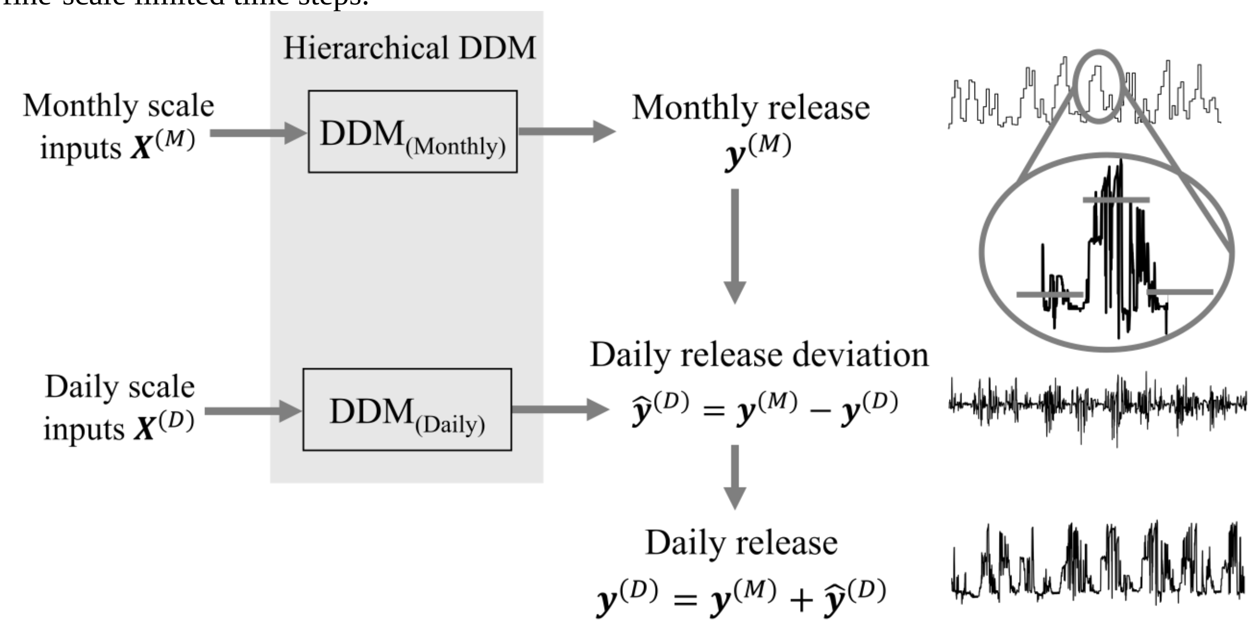

Figure 1. The hierarchical temporal scale framework with two layers shown for illustration. The top layer uses a monthly DDM to simulate monthly averaged release ($y^{(M)}$), and the subsequent bottom layer uses a daily DDM to simulate the daily deviation $\hat{y}^{(D)}$, or the difference between daily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

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 "deviation", defined as the difference between the fine scale release and release simulated by the coarse time scale DDM. The deviation 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. It is worth noting that the deviation $\hat{y}^{(D)}$ could be defined as the differences between observed releases at a coarse and a fine scale. The difference lies in the fact that the former, which defines the target deviation $\hat{y}^{(D)}$ as the difference between the observed daily release and simulated monthly/weekly release, to some extent, resembles the concept of a boosting algorithm, where the model is improved through the combination of multiple weak models to form a strong model, whereas the latter purely integrates multi-timescale information to generate the target fine-scale release. The effects of the two are considered equivalent when the model in the first layer is able to accurately predict the target release at the coarse scale (Figure S7 and S8 in Supplementary Materials).

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 different 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 few days. By learning the deviations 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 scales and represent the tradeoffs between various operation targets. For example, the release for flood control may be dependent on the current reservoir water level, which is affected by the storage target for water supply determined one month ago. The reservoir water level after flood control release may further affect water supply decisions in future time steps. Therefore, the deviation between two layers (i.e., two temporal scales) may represent the tradeoffs between various operation targets.

Two distinct strategies can be employed to train the DDM in each layer: “iterative” and “detached”. The iterative strategy enables concurrent updates to all temporal layers throughout the model training process. For neural network-type models such as Multi-Layer Perceptron (MLP) and Recurrent Neural Networks (RNN), a loss function that spans all temporal scales or multiple loss functions for each temporal scale can be defined, and weight updates are executed in each training epoch. The detached approach involves a simple arithmetic summation or weighted aggregation of the outputs from all layers to generate the final simulations. In this study, we use the iterative strategy to train the DDM.

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). We sourced some data from ResOpsUS, a comprehensive dataset on historical reservoir operations in the United States that was recently published by Steyaert et al. (2022). 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 and hydroclimatic data are used as inputs. The daily-scale meteorological forcing, including total precipitation rate (*P*, mm/day), potential evapotranspiration (*PET*, mm/day) and air temperature (*T*, °C) are obtained from the North American Land Data Assimilation System (NLDAS-2) forcing (Xia et al. 2012). The hydroclimatic data are averagely 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 the 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 purposes. The gridded snow depth (*SD*, mm) data retrieved from Broxton et al., (2019) is aggregated over the catchment area upstream of the reservoir to account for changes in snowmelt contributions over time. 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 configuration. 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 deviation (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 of input variables to determine release on a given day. For the WD and MD models, the coarse-resolution input variables of the past 4 steps (weeks or months) are used to derive the release at the current time step, and the daily scale deviations are simulated with daily input variables of the past 7 days. While inflow forecasts have been proven to strongly influence the seasonal reservoir operations, particularly for the high-elevation reservoirs fed by snowmelt in the western United States (Turner et al., 2020a), this study only uses the observed records in the past time steps, since it is difficult to acquire the actual streamflow forecasts for each reservoir in the historical period.

To explore the importance of each input variable for predicting reservoir operation at various time scales, we developed six experiments by varying the combinations of input variables in the three groups (Table 1). In Experiment 1, daily observed reservoir inflow (*I*), water storage (*S*), hydroclimatic information (*Met*, including *P*, *PET*, *SD* and *T*) 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 variables utilized in this study may contain information related to these factors. For example, reservoir evaporative loss is related to *PET* and water surface area, which in turn correlates with reservoir storage. Experiment 2 and Experiment 3 inputs exclude reservoir storage and inflow, respectively to evaluate the importance of reservoir information. Meteorological information is hidden in Experiment 4 to assess the impacts of meteorological forcing on reservoir release. 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*), potential evaporation (*PET*), snow depth (*SD*) and air temperature (*T*).

Time Scale	Experiment	Training variables
Daily (D)	D-1	*I, S, Met (P, PET, SD, T)*
	D-2	*I, Met (P, PET, SD, T)*
	D-3	*S, Met (P, PET, SD, T)*
	D-4	*I, S*
	D-5	*I*
	D-6	*S*
Weekly-Daily (WD)	WD-1	*I, S, Met (P, PET, SD, T)*
	WD-2	*I, Met (P, PET, SD, T)*
	WD-3	*S, Met (P, PET, SD, T)*
	WD-4	*I, S*
	WD-5	*I*
	WD-6	*S*
Monthly-Daily (MD)	MD-1	*I, S, Met (P, PET, SD, T)*
	MD-2	*I, Met (P, PET, SD, T)*
	MD-3	*S, Met (P, PET, SD, T)*
	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. As a powerful type of Recurrent Neural Networks (RNN), LSTM 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 internal calculation of the LSTM cell in this study is summarized in Appendix. 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), we utilize the iterative training strategy as mentioned in Section 2.1 to gain the optimal weights and bias. The two LSTMs are trained together by minimizing the mean square errors of reservoir release at both time scales, then the optimal parameters can be obtained by

$$\frac{1}{T}\sum_{t}^{T}\left(y_t^{(1)}-\widehat{y_t^{(1)}}\right)^2+\frac{1}{T}\sum_{t}^{T}\left(y_t^{(2)}-\widehat{y_t^{(2)}}\right)^2+\frac{1}{T}\sum_{t}^{T}(y_t-\widehat{y_t})^2$$

where $y_t^{(1)}$ and $\widehat{y_t^{(1)}}$ are the observed and simulated release at the monthly/weekly scales, $y_t^{(2)}$ and $\widehat{y_t^{(2)}}$ are the observed and simulated release deviations at the daily scale, y_t and $\widehat{y_t}$ are the observed and simulated release at the daily scale, θ represents the neural network parameters. The data at the coarse scale is remapped to the daily scale by resampling to ensure consistent lengths of data at both coarse and daily scales. 60% of time series data are used during the training process, 10% of them for validation, and the rest for testing. The Adam optimizer (Kingma and Adam, 2020) is applied for primary training and Stochastic Gradient Descent (SGD, Robbins and Monro, 1951) for finetuning. The number of training epochs and number of hidden units are found through trial-and-error. The learning rate during the pretraining process is 10^{-4} to 10^{-5} and the number of training epochs does not exceed 100, while the learning rate schedule is more complex during the finetuning process. Early stopping is implemented to decrease the probability of overfitting. To ensure the fairness of subsequent comparisons, the total number of parameters for both single-layer (D) and hierarchical time scale models (WD, MD) is constrained to be identical. Specifically, the hidden size in the single-layer model is almost equivalent to the sum of hidden size in all DDMs in the two-layer model. Concretely, we set the hidden size of daily single models for all reservoirs as 10, 15 or 20 to avoid excessively complex DDM models, ensuring that the maximum total number of parameters in single and hierarchical models does not exceed 2,000. The hidden size in the first layer of the hierarchical models is 5, 10 or 15, and that in the second layer is correspondingly adjusted. The Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) of daily reservoir release is used for assessing model performance in all experiments. To mitigate random effects arising during training, we initialize and train the models with different random seeds, calculating average performance metrics across the five trials. All the performances mentioned in the following sections are NSEs evaluated on the test sets. It is noted that the multi-layer configuration is flexible to use other data-driven 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). In experiments employing the same training variables, DDMs at the daily scale are capable of simulating the dynamics of reservoir release., and the two-layer hierarchical model (WD) exhibits consistent superiority over the daily model (D) in terms of accuracy, as evidenced by the probability of NSE exceedance across all reservoirs (Figure 2). MD configuration proves capable of outperforming the daily single scale model in select cases, notably for the majority of reservoirs examined in Experiments 2, 3, 5, and 6. In Experiment 1 with the most comprehensive input dataset, the median NSE for all reservoirs is 0.900, 0.831 and 0.872 for WD, MD and daily configuration, respectively. The WD configuration achieves NSE higher than 0.8 in more than 88% reservoirs, compared to 61% and 77% for the MD and D configurations, respectively. The WD configuration generally outperformed the MD configuration in most experiments. This may be attributed to the fact that weekly scale data provides four times more information than monthly scale data, thereby enabling the DDMs to be trained on more samples, even though both are resampled to the daily scale. Additionally, the finer resolution of the weekly scale may more accurately capture the variability of release decisions compared to the coarser monthly scale.

For all time scale configurations, reservoir inflow and storage are two key explanatory variables for modeling release behavior in most reservoirs, as indicated by the marginal performance gap between Experiments 1 and 4. With only reservoir inflow as model input in Experiment 5 (Figure 2e), the median NSE reaches 0.655, 0.826 and 0.762 for daily, WD and MD temporal configuration, respectively. The inflow provides most predictive power in reservoirs with relatively small storage and/or navigation purpose, particularly for run-of-river reservoirs located along the Columbia River or the Arkansas River, where there is a strong linear relationship between inflow and release at daily scale and the impact of storage can be negligible. Although the inflow-only models in Experiment 5 does not explicitly consider reservoir states, the LSTM architecture is expected to use the “hidden state” and “cell memory” to store accumulated inflow as a proxy for reservoir storage trend and use this information to simulate reservoir releases. However, due to the lack of 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 single time scale DDM to simulate the state of a 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.

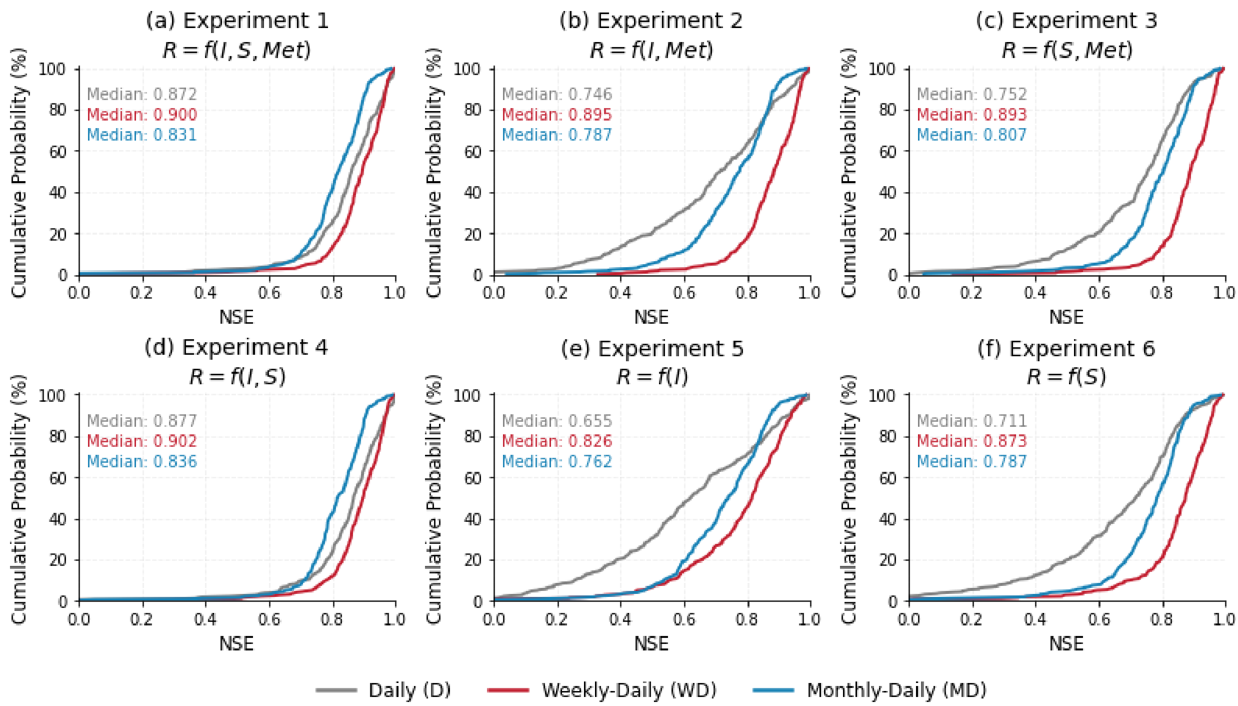

Figure 2. Probability of exceedance of NSE for all reservoirs resulting from single and hierarchical time scale models with different decision variables (Table 1).

The DDMs with storage alone as input in Experiment 6 have slightly higher predictive power compared to inflow-only models in Experiment 5 (Figure 2f) and produce median NSE of 0.711, 0.873 and 0.787 for Daily, WD and MD configuration, respectively. Using storage as the model 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. Storage-release rule curves are commonly used by reservoir operators (Yang et al. 2016), which covers the seasonal patterns of reservoir operation but the interannual variability of inflow are likely missing in such curves. At a 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 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 median NSE improves to 0.877, 0.902 and 0.836 for daily, WD and MD temporal configuration, respectively.

The performance improvement from including hydroclimatic variables (e.g., *P*, *PET*, *SD* and *T*) is investigated by comparing accuracies of DDMs in Experiment 1 vs. 4, Experiment 2 vs. 5, and Experiment 3 vs. 6. When both inflow and storage are used (Experiment 1 vs. 4), the improvement from additional hydroclimatic forcing is negligible (mean NSEs increase no more than 0.05). For DDMs with only inflow (Experiment 2 vs. 5) or storage (Experiment 3 vs. 6), adding hydroclimatic information slightly enhances the model performance, which is not unexpected as data-driven models typically benefit from more input information. Nevertheless, it may also underscore the potential of incorporating hydroclimatic conditions in reservoir release modeling (Denaro et al., 2017), particularly in regions where reservoir operation records are scarce.

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

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. Hierarchical temporal scale models work for some cases, although they do not always perform better than the models constructed on the single time scale under the same experiment settings. When one of the dominant explanatory variables (e.g., inflow or storage) is missing, a better organization (i.e., hierarchical temporal configuration) of the explanatory variables further enhances the performance. For example, in Experiment 2, 3, 5 and 6, more than 60% of reservoirs benefit from re-arranging the training data in hierarchical configuration (WD and MD) 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, in particular when hydrometeorological information or reservoir operational records are limited.

Regardless of the experimental settings, WD consistently outperforms another two-layer hierarchical model MD in simulating reservoir release decisions. Specifically, in Experiment 6 (Figure 3f) with the reservoir storage only as model inputs, performances of about 80% of reservoirs have been improved by hierarchical framework (WD), and it is more prominent than the MD where the first layer simulates the reservoir release on the monthly scale. It probably indicates that sub-monthly operational information and hydroclimatic forcing, which shows significant short-term variability, may provide a substantial portion of the information needed for accurate reservoir operation modeling. By incorporating information on moderate and fine time scales, WD DDM can well capture the complex dynamics of reservoir operations and yield highly accurate predictions, which may help inform the development of more effective and efficient reservoir management strategies in the face of increasing hydroclimatic variability.

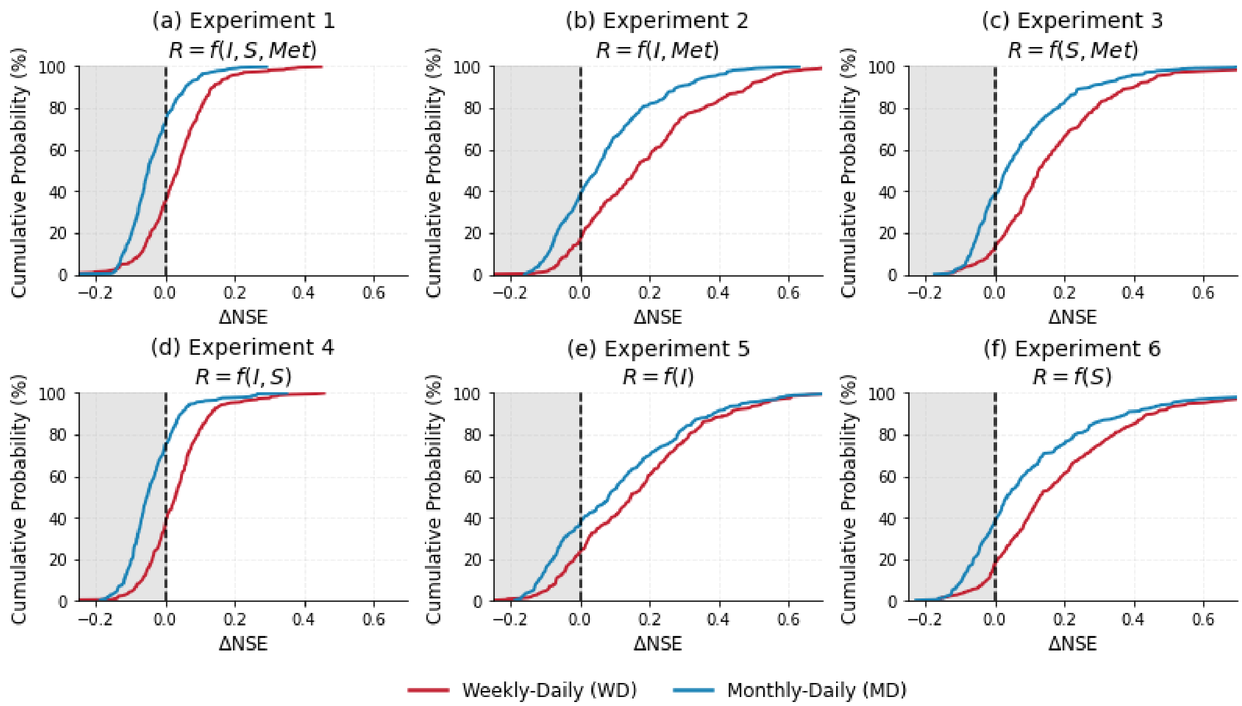

Figure 3. Improvement of NSE by hierarchical time scale framework ($\Delta NSE = NSE_{hierarchical} - NSE_{single}$) in a) Experiment 1 b) Experiment 2 c) Experiment 3 d) Experiment 4 e) Experiment 5 f) Experiment 6. $NSE_{hierarchical}$ represents the performances of hierarchical time scale models (WD, MD), while the NSE_{single} is the performance of a single time scale model (D).

3.3 Spatial pattern of DDM reservoir operation under various temporal configurations

Figure 4 shows the spatial distribution of average NSE improvement by WD and MD from Daily configuration for all six experiments, respectively. When the dominant explanatory variables (i.e., inflow and storage) are fed as model inputs (Experiments 1, 4), most reservoirs across the CONUS do not benefit significantly from the hierarchical temporal scale framework (Figure 4a, d). This can be attributed to the fact that the daily single model performs well for most reservoirs (Figure S1 in Supplementary Materials) with a median NSE higher than 0.85 (Figure 2a, d), which demonstrates the efficacy of data-driven models for reservoir release simulations. When the most relevant variables are sufficiently represented in the data, additional methods for regulated flow simulation refinement may not be necessary. Hierarchical models face challenges in improving the accuracy of models for reservoirs that primarily serve a single purpose or are predominantly operated at a single time scale. For instance, the hierarchical time scale model does not improve and even degrades release modeling of run-of-river reservoirs. In the New England district, where many reservoirs have limited storage capacity and are primarily used for flood control during flood seasons and recreation during non-flood seasons, hierarchical models are less effective across all experiments (Figure 4). 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 average NSE improvement ΔNSE from Daily scale to hierarchical time scale configuration of DDMs in a) Experiment 1 b) Experiment 2 c) Experiment 3 d) Experiment 4 e) Experiment 5 f) Experiment 6. The circles with black solid edges represent reservoirs labelled as "lock & dam".

Hierarchical models send positive signals for reservoirs in the Midwest. The hierarchical DDM improves NSE over Daily scale in many reservoirs in the western United States as shown in Figure 4, and the magnitude of improvement in model performance varies across different experiment setups. For experiment 1 and 4 that includes both inflow and storage as model inputs, the average NSE improvement ΔNSE is subtle (about 0.1~0.25) for some reservoirs in Montana, Utah, New Mexico and Texas (Figure 4a, d). It implies that the hierarchical model is effective in capturing reservoir release behavior in western regions, at least to a comparable degree as the daily single model. For experiment 2 and 5 that does not contain storage fed into models, release simulations are boosted by hierarchical temporal scale framework for reservoirs on the High Plains (e.g., Texas, Oklahoma, Kansas), highlighting the signature of seasonal cycle of water supply operation in these reservoirs. Reservoirs that provide water for agricultural irrigation or municipal/industrial use often base release decisions on the water level or storage status. In situations where operational records of storage are unavailable, comprehensive utilization of inflow data across various temporal scales may serve as a compensatory mechanism. Many water-supply reservoirs maintain nearly constant storage volume at the start and end of an operational year, resulting in a nearly balanced inflow and release volume at a certain temporal scale (monthly, seasonally, or annually). Thus, it becomes feasible to detect reservoir behavior when changes in inflow over the preceding months or weeks are known. It would facilitate accurate estimate of regulated flow regimes in the absence of readily available datasets on water level or storage under future scenarios. In the case of reservoirs located in the Rocky Mountains and the Colorado River basin, the hierarchical model consistently enhances the accuracy of release modeling, regardless of whether inflow or storage is excluded as explanatory variables (Experiment 2 and 5; Experiment 3 and 6). As stated in Section 3.1, inflow generally reflects short-term variability or the effects of fine-scale weather fluctuations, while storage represents the cumulative hydrologic response during past periods. The absence of either of these dominant factors results in a loss of vital information for accurate release modeling. Hence, the behavior of reservoirs in the west cannot be fully captured by DDMs at a single temporal scale (Figure S1c, d, e, f in Supplementary Materials). Among the observational records analyzed in this study, 216 reservoirs serve at least 3 purposes and 79 out of 327 serve at least 5 purposes. In spite of accounting for multiple time scales may not be imperative for simpler reservoirs that serve fewer purposes or operate under less complex conditions, it is crucial for effectively modeling multi-purpose reservoirs and multi-reservoir systems.

In summary, our analysis indicates that reservoir release modeling can be enhanced by leveraging the availability of adequate information, with particular emphasis on key explanatory variables. The inclusion of meteorological forcing data may also be beneficial for accurate simulation. In situations where the records of reservoir inflow or storage are inaccessible, the comprehensive utilization of multiple temporal scales can lead to improved modeling outcomes.

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 providing an interpretable avenue to enhance the understanding of reservoir behavior.

A prevalent method for enhancing interpretability is to analyze variable importance. Many approaches can be taken to assess feature importance of machine learning models. Wei et al. (2015) conducted a comprehensive review of various techniques for variable importance analysis in different disciplines and analyze their relative merits. Recently, Quinn et al. (2019) used time-varying sensitivity analysis to open the black box of multi-reservoir operation models. Additionally, Shapley Additive Explanations (SHAP) (Lundberg and Lee, 2017) and permutation feature importance (Breiman, 2001; Fisher, Rudin, and Dominici, 2018) have gained popularity in recent years. In this study, we used well-trained data-driven models to conduct a variable importance analysis that explores the impact of decision variables on reservoir release schemes across different time scales. We employed the permutation feature importance method to measure variable importance, which involves randomly permuting feature values in the input data and examining the effect on model performance, as measured by a specific metric (such as NSE in our study). The extent of the decrease in performance reflects the relative importance of the feature, with a greater decline in performance indicating a more influential feature in the model. Then the variable that leads to the largest change is referred to as the most important variable, or dominant factor.

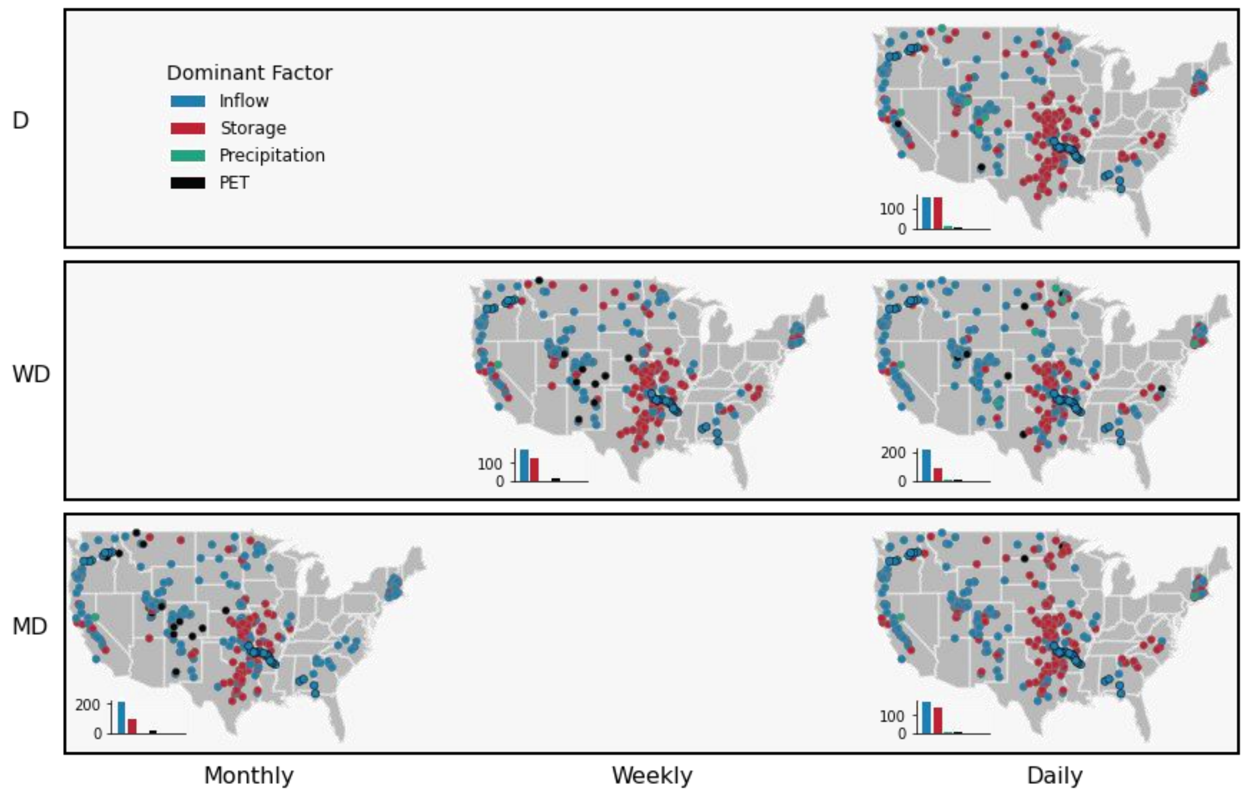

Figure 5. Spatial distribution of dominant factors across daily, weekly and monthly scales. The circles with black solid edges represent reservoirs labelled as "lock & dam". The inset at the bottom left corner depicts the number of reservoirs in which a certain variable (inflow, storage, precipitation or potential evapotranspiration) is identified as the dominant factor influencing release decisions.

Figure 5 displays the most important variable for each reservoir across CONUS on the different time scales (daily, weekly and monthly) of Daily, WD and MD configurations in Experiment 1 that contains all the variables (inflow, storage, precipitation, PET, snow depth and air temperature) as model inputs. For half of reservoirs (163 out of 327), the same variable has critical influences on the release on all time scales (daily, WD-weekly, MD-monthly, WD-daily, MD-daily), 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 120 of these reservoirs, inflow plays a decisive role in reservoir release at all time scales, while storage volume is the most instructive variable for 42 of these. Daily models with good performance (e.g., reservoirs labeled as “lock & dam” 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. Reservoirs located on the High Plains, where water level is a crucial factor in release operations, consistently show storage as the dominant factor influencing release decisions. The findings of variable importance in California and the High Plains differ slightly from those reported by Hejazi et al. (2008) (e.g., many reservoirs in these two regions reported by Hejazi et al., 2008 do not consider storage as the dominant factor), who investigated the dependency of operators’ release decisions using the method of information theory based on weekly/monthly operational records. It should be noted that Hejazi et al. (2008) included past release as a decision variable, while this study did not consider it as a model input. Furthermore, the operational dataset utilized in this study is updated to 2016, which may account for this discrepancy. Martis Creek Lake, located in the Sierra Nevada Mountains outside the town of Truckee, serves the dual purpose of flood control and recreation, with precipitation (P) being the most predictive variable for reservoir inflow at all timescales. The lake is situated in a headwater watershed with a small contributing area, which further supports the use of P as a reliable proxy for inflow prediction. It is worth mentioning that for two reservoirs, the Elephant Butte Reservoir in New Mexico and the Moon Lake in Utah, PET has a major effect on reservoir release at the daily, WD-weekly, and MD-monthly scales (maps along the diagonal shown in Figure 5), which could involve considerable reservoir evaporation and water use for agricultural irrigation in the arid, semi-arid western mountains. 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 (117 out of 327 for WD; 108 out of 327 for MD configuration) reservoirs vary in their dependency on decision variables at different time scales (shifted from weekly to daily in the WD; from monthly to daily in the MD configurations), suggesting that reservoir operators consider different information at different time scales to fulfill multiple designed purposes. For MD shown in Figure 5, at the monthly scale, operations of 214 reservoirs primarily depend on the reservoir inflow, and 91 reservoirs rely more on storage volume. At the daily scale, the number of reservoirs with major dependency on inflow decreases to 175 and that of reservoirs relying more on storage volume increases to 174. From the coarse scale to the fine scale, nearly 20% reservoirs (64 out of 327) shift their primary dependence from inflow to storage volume. As mentioned in Section 3.3, many reservoirs tend to maintain nearly constant water level or storage at the beginning and end of an operational year, which can result in a balance between the total volume of inflow and release at certain time scales (e.g., annually, seasonally). Consequently, it is not surprising that for almost two-thirds of the reservoirs studied, inflow exhibits the strongest relation with release at the monthly scale. At the daily scale, operators tend to give greater weight to current or recent storage status (or water levels) when making release decisions, since reservoir storage is a crucial factor in determining the availability of water for downstream users or for maintaining water levels within acceptable limits. Although neither snow depth nor temperature is detected as the dominant factor at any of the reservoirs, it would be premature to dismiss these two factors as unimportant. This is probably on the ground that the variable importance analysis used in this study is model-based rather than based on observational data, which sometimes might produce misleading results due to inadequate feature selection or inappropriate model configuration, particularly when snow depth or air temperature is tightly linked to other explanatory variables such as P or PET. It is important to exercise caution in interpreting the results. Additionally, we merely focus on the most important variables in this study, but snow depth and air temperature are likely to play a substantial role in snow-dominated, high-altitude mountain reservoirs.

3.5 Reservoir release behaviors across time scales

Compared to attempts to capture reservoir operation at a fixed time scale, the hierarchical temporal configuration in this study demonstrates improved model performance while utilizing the same input information, particularly when essential decision variables such as inflow or storage are inaccessible. 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 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 5 (with inflow only) has NSE of 0.848, 0.969, 0.920 for Daily, WD and MD configuration, respectively. The DDM identifies reservoir storage as the dominant variable on release at Daily, WD, and MD scales, respectively.

Figure 6 shows the scatter plots of release vs. inflow and storage vs. inflow at various time scales. At the annual time scale (Figure 6d), the outflow is highly correlated with inflow, suggesting the reservoir has seasonal flow regulating capacity. The slightly lower annual release than the inflow (Figure 6d) indicates water balance is roughly held on annual time average. Water supply withdrawals made through pumping or diversion have a limited impact on the mass balance. The randomness between monthly inflow and release (Figure 6c) 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.

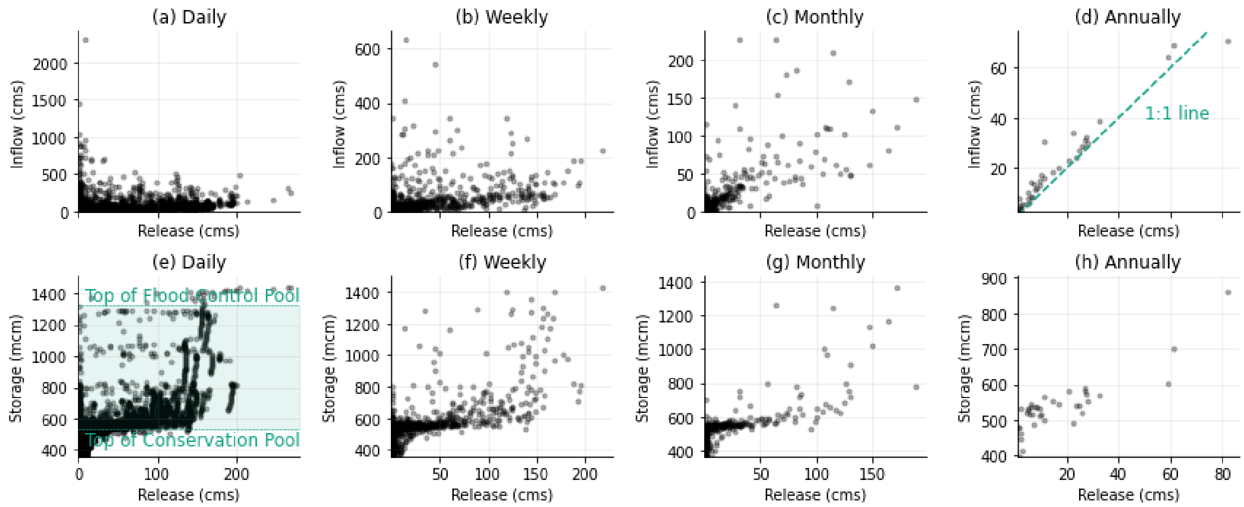

Figure 6. Relationship between inflow and release at a) daily, b) weekly, c) monthly, d) annual scale and; Relationship between reservoir storage and release at e) daily, f) weekly, g) monthly, h) annual scale of Belton Lake (TX00002).

Figure 7a 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) with given inflow only, 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.

The hydrograph of Year 2002 in Figure 7b shows the seasonal pattern and short-term variation produced by different DDM configurations. The Daily Scale DDM tends to exhibit a faster decay in release following flood events, since the daily scale model is sensitive to the daily input and lacks the long-term information. The WD scale configuration demonstrates superior performance in capturing both seasonal water supply and flood control release at the Belton Lake. As an illustration, in November 2002, when the daily model produces a false release response while the hierarchical models do not. It exposes the shortcomings of a single daily scale model for multi-purpose reservoirs that consider reservoir storage as an influential factor. Many large reservoirs in Texas adhere to a general strategy based on minimizing the risk and consequences of releases contributing to downstream flooding in the flood seasons, while ensuring the maximum design water surface is never exceeded (as shown in Figure 6e). Release decisions are contingent upon the flood control pool storage capacity. In the non-flood seasons, these reservoirs strive to maximize water levels within the conservation pool, without surpassing its upper limit (i.e., the top of conservation pool). When storage information is not explicitly provided as an input, it can be challenging for a daily single model to consistently and accurately respond to inflow information. Although data-driven models are expected to derive storage information from the physical constraints (e.g., water balance equation) and the accurately simulated release time series (Figure 7c and Figure S9 in Supplementary Materials), challenges remain due to the inaccurate simulation in release, error accumulations, missing water budget terms, etc. If only reservoir inflow is given, which typically represents short-term hydrologic variability, the long-term target may be overlooked by a daily single time scale model due to the absence of long-term hydrologic indicator. An example from Figure 7c illustrates that from September to November in 2002, the water level (storage status) fell below the target conservation capacity, resulting in the reservoir not releasing water in response to inflow events during this period. Storage is recognized as the dominant factor that determines the reservoir release decision at both daily and weekly scales during this period (Figure 7d, e). In the absence of the storage that can reflect long-term hydrologic variability, the Daily model fails to capture the implicit long-term patterns inherently embedded in the absent key variable. This "short sight" explains the erroneous response observed in Figure 7b (gray line), further emphasizing the importance of fully utilizing multiple temporal scales of information. In contrast, the hierarchical model with multiple time scales can better incorporate the complex dynamics of the reservoir system, which can lead to more reliable and robust simulation results.

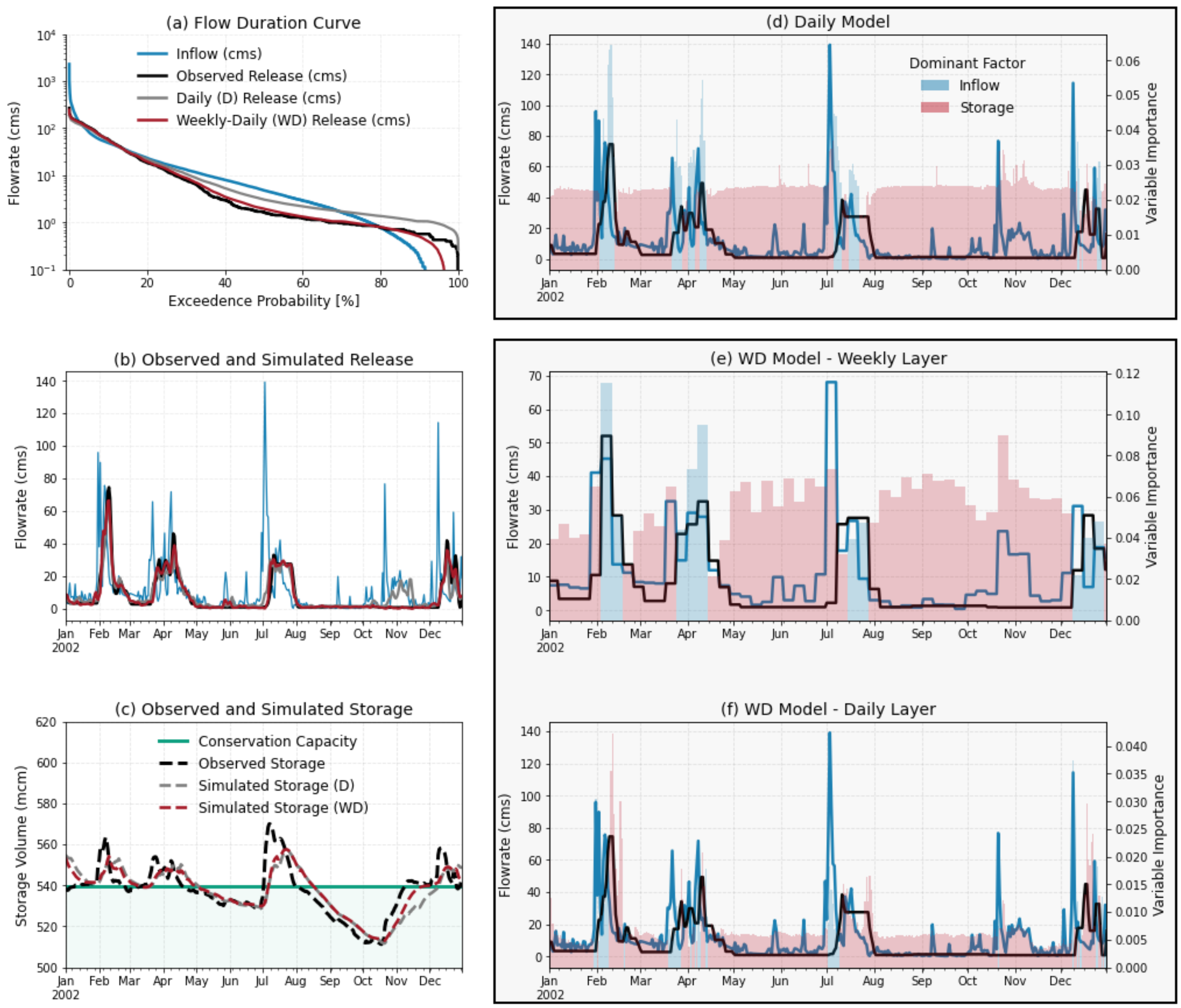

Figure 7. Experiments exploring the dominant factors and simulated outputs of Belton Lake (TX00002) in Texas. Comparisons of observed and simulated release in Experiment 5 (only inflow as inputs) shown in a) Flow Duration Curve (FDC) and b) hydrograph during the calendar

year 2002 (test period); c) Conservation capacity, observed storage, and simulated storage during the calendar year 2002, where the simulated storage is derived from the water balance equation, inflow, and simulated release; Time-varying dominant factors from Experiment 4 (both inflow and storage as model inputs) shown in d) daily model, e) weekly layer of WD model, and f) daily layer of WD model during the calendar year 2002. The sub-axis presents the computed absolute variable importance values obtained through the application of Shapley Additive Explanations (SHAP) method (Lundberg and Lee, 2017). The dominant factors, characterized by the highest absolute variable importance values, are denoted by different colors.

These observations highlight the importance of appropriately organizing training data at various time scales in order to enable machine learning techniques to capture the underlying relationships inherent at each time scale. We also used other machine learning techniques (e.g., Random Forest, see Figure S4, S5, S6 in Supplementary Materials) to configure the hierarchical DDM and achieved satisfactory 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 Discussion

4.1 Strategies and limitations of data-driven reservoir operation modeling

In this study, we employed LSTM networks to simulate reservoir release decisions, primarily due to their similarities to traditional hydrological models to some extent — for example, current hydrological fluxes are determined by current inputs and past states. The strength of LSTM networks lies in their ability to learn nonlinear patterns and long-term dependencies, making them ideal for simulating reservoirs where the hydrological behavior may change over time. LSTMs are expected to be suitable for modeling when the decision variables (or model inputs) exhibit temporal dependence. While LSTM networks have become widely used in the hydrology community, barriers may exist due to the requirement of a large amount of training data and careful finetuning processes to achieve accurate results. In addition, the measurement of feature importance in neural networks is not so straightforward and make it lack interpretability. It is essential to acknowledge that LSTM networks may not be the optimal choice for simulating reservoir operations all the time, especially in cases where actual operation rules are explicit. For instance, in some highly engineered watersheds in the western United States, which are equipped with a large number of dams, the reservoir release patterns can deviate considerably from the natural flow characteristics of the system. These deviations are a result of the complex interactions between the reservoir operations and the hydrological processes, which can be influenced by a range of factors such as climate change, water demand, and land use change. In these cases, other white-box models such as Classification and Regression-Tree (CART) or Random Forest (RF), which are more intuitive for decision-makers and excel in capturing patterns from various features, may be more appropriate (e.g., Yang et al., 2016). Moreover, a notable drawback of LSTM and other RNN-based models typically pertains to their dependence on data continuity, particularly when the lookback or look forward window is extensive. For instance, in the context of rainfall-runoff or models involving surface-groundwater interaction, such a window may span as much as 180, 270, or 365 days (e.g., Kratzert et al., 2019). While preprocessing techniques can handle missing data to create a continuous time series as inputs, the usefulness of models needing continuous data might be limited in situations where reservoir operation records are scarce.

Unlike many well-established data-driven models for reservoir operations, such as those developed by Turner et al. (2021), Chen et al. (2022), Dong et al. (2023), and Brunner et al. (2023), this study omits reservoir storage simulation, which is a frequently pursued research objective in the development of reservoir operation models. It is because this study aims at investigating the significance of multi-timescale information in data-driven reservoir operation modeling. Specifically, this study seeks to examine the impact of incorporating multiple temporal scales of decision variables in the construction of models for reservoir release and aspires to contribute to the ongoing effort to enhance the performance and robustness of data-driven regulated flow simulations. It is noteworthy that the interdependency between reservoir inflow, storage, and release across various time scales (as pointed out in Section 3.3 and the example shown in Figure 6) can be leveraged to extract informative features for input into white-box models (i.e., feature engineering considering multiple scale temporal information), with the potential to enhance the balance between model performance and interpretability. By exploiting the rich temporal dynamics of reservoir operations data, it can facilitate a more comprehensive and interpretable representation of the underlying processes.

The feasibility of data-driven reservoir simulations can be further boosted through the use of hybrid strategies that combine rule-based or conceptual operation schemes with machine learning techniques (Gangrade et al., 2022; Dong et al., 2023). By leveraging expert knowledge in the form of appropriate feature engineering (Yang et al., 2016, 2017), and by incorporating reservoir storage dynamics derived from a range of advanced sensing techniques (Eilander et al., 2014; Van Den Hoek et al., 2019; Chen et al., 2022; Sorkhabi et al., 2022), it is possible to use data-driven models to better reconstruct downstream flow in data-sparse regions, using meteorological forcing and inflow generated by hydrological models.

4.2 Hierarchical nature of anthropogenic decisions

DDMs are generally not constrained by the complexity of the training dataset and can achieve better prediction with more training variables. However, the results illustrated in Figure 4 suggest that in an identical experimental setup, employing congruent variables and model architecture while maintaining consistent model complexity (as indicated by an equivalent total parameter count), the hierarchical timescale model—which encompasses both coarse and fine scales and is anticipated to acquire an augmented amount of temporal data—does not invariably surpass the performance of a single-timescale model. 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 the 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 decisions depend on soil moisture conditions.

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 scales and require corresponding hydroclimatic information. For example, the reservoirs in the Colorado River Basin use 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 scales may require tradeoffs. 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 the 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. There are efforts to forecast informed reservoir operation (FIRO) (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. In addition, different forecast products have different lead-time (ranging from hours by short-term weather forecast to seasons by climate models), a better understanding of hydrometeorological factors at various time scales affecting reservoir operation would facilitate FIRO to select the forecast products suitable for a specific reservoir.

5 Conclusions

In this study, we proposed a hierarchical temporal scale framework to improve the data-driven reservoir release modeling. When the dominant explanatory variables observed inflow or storage are unavailable as inputs, more than 60% of reservoirs across the CONUS gain the improvement in model performances, while modeling of 80% of them can be more accurate by this framework if the first layer is constructed at weekly scale. The proposed framework accounts for the influence of multiple temporal-scale variability to accurately predict reservoir release behavior, which may have inspiring implications for data-driven reservoir release modeling in regions where operating records are incomplete or limited in availability.

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, data-driven reservoir components that comprehensive utilization of multi-timescale information could be incorporated into physics-based models to improve the accuracy of hydrological process simulations.

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, 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.

Appendix

The Long-Short Term Memory (LSTM) computations are expressed as

$$i_t = \sigma(W_{xi} \cdot x_t + W_{hi} \cdot h_{t-1} + b_i)$$
$$f_t = \sigma\left(W_{xf} \cdot x_t + W_{hf} \cdot h_{t-1} + b_f\right)$$
$$g_t = tanh\left(W_{xg} \cdot x_t + W_{hg} \cdot h_{t-1} + b_g\right)$$
$$o_t = \sigma(W_{xo} \cdot x_t + W_{ho} \cdot h_{t-1} + b_o)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$
$$h_t = o_t \odot tanh\,(c_t)$$

where W_{xi}, W_{xf}, W_{xg} and W_{xo} are learnable weights of inputs x_t, W_{hi}, W_{hf}, W_{hg} and W_{ho} are learnable weights of the previous hidden states h_t, and b_i, b_f, b_o and b_g are biases of the four gates, respectively. σ means sigmoid function, tanh is hyperbolic tangent function, and $\odot$ represents element-wise multiplication.

Availability Statement

All data used in this research are publicly available. The meteorological forcing (precipitation, potential evapotranspiration and air temperature) is available at https://ldas.gsfc.nasa.gov/nldas/v2/forcing. Snow depth data is retrieved from Daily 4 km Gridded SWE and Snow Depth from Assimilated In-Situ and Modeled Data over the Conterminous US, Version 1 (NSIDC-0719) (https://nsidc.org/data/nsidc-0719/versions/1). The dataset of reservoir operations utilized in this study is available online (https://www.hydroshare.org/resource/79c262b627fc4ce293379b5e95457146/), or directly from the United States Bureau of Reclamation (https://water.usbr.gov/api/web/app.php/api/) and the United States Army Corps of Engineers (collected via Duke University; https://nicholasinstitute.duke.edu/reservoir-data/, Patterson et al., 2018).

Acknowledgments

R. Zeng is partially supported by Ralph E. Powe Junior Faculty Enhancement Awards from Oak Ridge Associated Universities and NSF CAREER award 2239621. We appreciate the reviewers' comments that help to improve this study.

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).

Broxton, P., Zeng, X. and Dawson N. (2019). Daily 4 km gridded SWE and snow depth from assimilated in-situ and modeled data over the conterminous US, Version 1. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://nsidc.org/data/NSIDC-0719/versions/1. Date accessed 01-24-2023.

Brunner, M. I., & Naveau, P. (2023). Spatial variability in Alpine reservoir regulation: deriving reservoir operations from streamflow using generalized additive models. *Hydrology and Earth System Sciences*, *27*(3), 673-687.

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

Chen, T., Song, C., Zhan, P., Yao, J., Li, Y., & Zhu, J. (2022). Remote sensing estimation of the flood storage capacity of basin-scale lakes and reservoirs at high spatial and temporal resolutions. *Science of The Total Environment*, *807*, 150772.

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.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. *J. Off. Stat*, *6*(1), 3-73.

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.

Dong, N., Guan, W., Cao, J., Zou, Y., Yang, M., Wei, J., Chen, L. and Wang, H., (2023). A hybrid hydrologic modelling framework with data-driven and conceptual reservoir operation schemes for reservoir impact assessment and predictions. *Journal of Hydrology*, 129246.

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.

Eilander, D., Annor, F. O., Iannini, L., & Van de Giesen, N. (2014). Remotely sensed monitoring of small reservoir dynamics: A Bayesian approach. *Remote Sensing*, *6*(2), 1191-1210.

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.

Gangrade, S., Lu, D., Kao, S. C., & Painter, S. L. (2022). Machine Learning Assisted Reservoir Operation Model for Long-Term Water Management Simulation. *JAWRA Journal of the American Water Resources Association*.

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.

Grossmann, A., & Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis*, *15*(4), 723-736.

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.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C. and Liu, H.H., (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, *454*(1971), 903-995.

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.

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.

Liu, H., Tian, H. Q., Pan, D. F., & Li, Y. F. (2013). Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Networks. *Applied Energy*, *107*, 191-208.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

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.

Quinn, J. D., Reed, P. M., Giuliani, M., & Castelletti, A. (2019). What is controlling our control rules? Opening the black box of multireservoir operating policies using time-varying sensitivity analysis. *Water Resources Research*, *55*(7), 5962-5984.

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The annals of mathematical statistics*, 400-407.

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.

Sorkhabi, O. M., Shadmanfar, B., & Kiani, E. (2022). Monitoring of dam reservoir storage with multiple satellite sensors and artificial intelligence. *Results in Engineering*, *16*, 100542.

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.

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.

Van Den Hoek, J., Getirana, A., Jung, H. C., Okeowo, M. A., & Lee, H. (2019). Monitoring reservoir drought dynamics with landsat and radar/lidar altimetry time series in persistently cloudy eastern Brazil. *Remote Sensing*, *11*(7), 827.

Wang, J., Wang, Z., Li, J., & Wu, J. (2018). Multilevel wavelet decomposition network for interpretable time series analysis. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2437-2446).

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*.

Zeng, X., Broxton, P., & Dawson, N. (2018). Snowpack change from 1982 to 2016 over conterminous United States. *Geophysical Research Letters*, *45*(23), 12-940.

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., X. Liu & 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.