



Research papers

Improving the prediction of daily reservoir releases over the CONUS using conditioned LSTM

Hoang Tran ^{a,*} , Tian Zhou ^a, Zeli Tan ^a , Yilin Fang ^b , L. Ruby Leung ^a ^a Atmospheric, Climate, and Earth Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA^b Earth System and Science Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA

ARTICLE INFO

This manuscript was handled by Dr Marco Borga, Editor-in-Chief, with the assistance of Georgia A. Papacharalampous, Associate Editor

Keywords:

Reservoirs
CONUS
Operational control
Machine learning
LSTM

ABSTRACT

Reservoirs play a vital role in regulating streamflow timing and variability for hydroelectricity, flood control, water supply, irrigation, and recreation. Despite their importance, many reservoirs lack comprehensive operational guidelines, making their management complex due to conflicting operational objectives. Hence traditional policy-based reservoir models often fail to capture real-world conditions accurately and they depend on perfect streamflow predictions, which are not always available. In contrast, data-driven models like Long Short-Term Memory (LSTM) networks offer a robust alternative. This study introduces an approach that integrates reservoir characteristics—such as main use, climate, and maximum capacity—into the LSTM model to enhance reservoir release predictions. Using data from nearly 200 reservoirs in the contiguous United States (CONUS), our conditioned LSTM model (*LSTM_{cond}*) was compared with both the vanilla *LSTM* and a traditional policy-based approach. Our results show that while both *LSTM_{cond}* and *LSTM* perform better than the policy-based approach, *LSTM_{cond}* consistently outperforms *LSTM* for hydroelectric, water supply, irrigation, and recreation reservoirs. The KGE median values for *LSTM_{cond}* for out-sample reservoirs are 0.764, 0.565, 0.821, and 0.779, respectively, for the aforementioned reservoir types, which are consistently higher than the corresponding KGE values of 0.737, 0.413, 0.775, and 0.713 of *LSTM*, demonstrating its advantages in improving generalizability. By integrating reservoirs' attributes into the LSTM model, our approach improves the performance of outflow simulation and provides insights into the contributing factors, making the predictions more explainable.

1. Introduction

Reservoirs are crucial infrastructures in human activities to regulate the streamflow timing and variability for purposes such as hydroelectricity, flood control, water supply, irrigation and recreation (Boulange et al., 2021; Ehsani et al., 2017; Forsberg et al., 2017; Lehner et al., 2011; Moran et al., 2018; Ortiz-Partida et al., 2016; Patterson and Doyle, 2018; Simonovic, 1992). In the contiguous US (CONUS), there are over 52,000 reservoirs which collectively hold 600,000 million cubic meters of water (Steyaert et al., 2022) and generate about 6.3 % of the total electricity (EIA, 2023). While some reservoirs are managed following operational guidelines (Klipsch and Hurst, 2007; Yates et al., 2005), the operational decisions for many reservoirs are complex due to trade-offs among conflicting operational targets. Consequently, representing reservoirs accurately in large-scale hydrology models is challenging (Liu et al., 2007; Turner et al., 2021b).

Reservoir modeling can be classified into two types: policy-based

models and data-driven models. Policy-based models have a long history of development and achieve satisfactory results (Giuliani et al., 2021; Hanasaki et al., 2006; Turner et al., 2021b; Voisin et al., 2013; Zhao et al., 2011). However, they have drawbacks, including the inability to capture real-world conditions (Giuliani et al., 2021), differences from actual reservoir characteristics or streamflow regimes, and reliance on perfect streamflow predictions (i.e. without forecast uncertainty), which may not be available or used in decision-making (Zhao et al., 2011). These drawbacks are summarized in Longyang and Zeng (2023).

Data-driven models offer a more robust alternative in reservoir modeling (Aboutalebi et al., 2015; Chen et al., 2022a; Chen et al., 2022b; Coerver et al., 2018; Dong et al., 2023; Fan et al., 2023; Gangrade et al., 2022; Hipni et al., 2013; Li et al., 2024; Lin et al., 2006; Longyang and Zeng, 2023; Mateo et al., 2014; Tran et al., 2024b; Turner et al., 2020a; Turner et al., 2020b; Wei and Hsu, 2008; Yang et al., 2017; Yassin et al., 2019; Zhang et al., 2018a; Zhao and Cai, 2020). As their

* Corresponding author.

E-mail address: hoang.tran@pnnl.gov (H. Tran).

name implied, data-driven models aim to extract the patterns of reservoir operation from records of related hydrometeorological variables to represent their complex nonlinearity. Data-driven models provide accurate results while maintaining low computational costs (Zhang et al., 2018b). Among various types of data-driven models, the Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) has gained significant attention in reservoir modeling because of its accurate and dependable outcomes (Longyang and Zeng, 2023; Tran et al., 2024b; Zhang et al., 2018b).

LSTM has been successfully applied to model reservoir systems at regional (Fan et al., 2023; Tran et al., 2024b) and continental scales (Longyang and Zeng, 2023). However, due to LSTM's black-box nature, it is challenging to interpret and understand the influence of different input predictors. This lack of transparency can lead to issues where models may perform well for certain reservoirs but fail to generalize across different contexts. For an application of LSTM to model "universal" streamflow behaviors, Kratzert et al. (2019) has suggested to incorporate catchment attributes when training LSTM to achieve the best results.

Therefore, we propose a new approach of integrating reservoir characteristics into LSTM to model reservoir releases in this study. By incorporating these characteristics, our approach aims to enhance the model's ability to capture the nuances of individual reservoirs, making the LSTM model more interpretable. This approach is compared with the vanilla LSTM model and a policy-based approach for nearly 200 large reservoirs across the CONUS. We also discuss improvements and potential drawbacks compared to the vanilla LSTM.

2. Methodology

2.1. ResOpsUS and GRanD dataset

The ResOpsUS dataset developed by Steyaert et al. (2022) offers daily data on reservoir storage, inflow, and outflow for 679 major reservoirs across the US, with the best coverage available for around 200 reservoirs between 1980 and 2020. Fig. 1 illustrates these reservoirs from the ResOpsUS dataset, color-coded based on different climate regions, main uses, and operating agencies. Temperate reservoirs are located on the west and east coasts of the CONUS, while boreal and arid reservoirs are mixed in the Midwest region (Fig. 1a). Flood control reservoirs are mostly temperate and large capacity (>1000 MCM-million cubic meter – 1×10^6 m 3) reservoirs (Fig. 1c and 1d), and irrigation reservoirs are predominantly operated by the Bureau of Reclamation (BOR) (Fig. 1b and 1c).

The ResOpsUS dataset provides various characteristics of the reservoirs, such as geographic locations, operating agencies, and the start and

end dates of the time series. Additionally, we utilized the Global Reservoirs and Dams (GRanD) dataset (Lehner et al., 2011; Wang et al., 2022) to supplement other reservoir characteristics, including main use and storage capacity. GRanD encompasses data on 6862 dams and their associated reservoirs, collectively holding a storage capacity of 6197 km 3 .

2.2. The inferred storage targets and release Functions (ISTARF), a traditional policy-based approach

The Inferred Storage Targets and Release Functions (ISTARF) was developed by Turner et al. (2021b) to represent the storage and release policies of the large reservoirs in the US. ISTARF uses harmonic functions to model the reservoir storage and releases. The ISTARF shows significant and reliable enhancement in both release and storage compared to the widely-used Hanasaki (Hanasaki et al., 2006) method.

ISTARF's harmonic functions involve 19 parameters, including the upper and lower bounds of the Normal Operating Range (NOR) (Turner et al., 2020a; Turner et al., 2021a, 2021b).

The upper bound storage is calculated as:

$$\widehat{S}_t^\uparrow = \begin{cases} \mu^\downarrow + \alpha^\downarrow \sin 2\pi\omega t + \beta^\downarrow \cos 2\pi\omega t, & \widehat{S}_{\min}^\uparrow \leq \mu^\downarrow + \alpha^\downarrow \sin 2\pi\omega t + \beta^\downarrow \cos 2\pi\omega t \leq \widehat{S}_{\max}^\uparrow \\ \widehat{S}_{\min}^\uparrow, & \mu^\downarrow + \alpha^\downarrow \sin 2\pi\omega t + \beta^\downarrow \cos 2\pi\omega t < \widehat{S}_{\min}^\uparrow \\ \widehat{S}_{\max}^\uparrow, & \mu^\downarrow + \alpha^\downarrow \sin 2\pi\omega t + \beta^\downarrow \cos 2\pi\omega t > \widehat{S}_{\max}^\uparrow \end{cases} \quad (1)$$

The lower bound storage is calculated as:

$$\widehat{S}_t^\downarrow = \begin{cases} \mu^\downarrow + \alpha^\downarrow \sin 2\pi\omega t + \beta^\downarrow \cos 2\pi\omega t, & \widehat{S}_{\min}^\downarrow \leq \mu^\downarrow + \alpha^\downarrow \sin 2\pi\omega t + \beta^\downarrow \cos 2\pi\omega t \leq \widehat{S}_{\max}^\downarrow \\ \widehat{S}_{\min}^\downarrow, & \mu^\downarrow + \alpha^\downarrow \sin 2\pi\omega t + \beta^\downarrow \cos 2\pi\omega t < \widehat{S}_{\min}^\downarrow \\ \widehat{S}_{\max}^\downarrow, & \mu^\downarrow + \alpha^\downarrow \sin 2\pi\omega t + \beta^\downarrow \cos 2\pi\omega t > \widehat{S}_{\max}^\downarrow \end{cases} \quad (2)$$

Where \widehat{S}_t^\uparrow and \widehat{S}_t^\downarrow are fitted bounds of the NOR for week t , μ , α , and β are baseline, amplitude and phase parameters, ω is the frequency. The weekly release \widehat{R}_t is calculated as:

$$\widehat{R}_t = \alpha_1 \sin 2\pi\omega t + \alpha_2 \sin 4\pi\omega t + \beta_1 \cos 2\pi\omega t + \beta_2 \cos 4\pi\omega t \quad (3)$$

Where α_1 , α_2 , β_1 , and β_2 are four release parameters.

Turner et al. (2021a) trained ISTARF parameters for all the data-rich reservoirs in the ResOpsUS dataset, resulting in a dataset of inferred policies. The weekly reservoir releases are predicted using a pre-determined structure based on reservoir states (e.g., storage), characteristics (e.g., capacity), and seasonal variability, with the

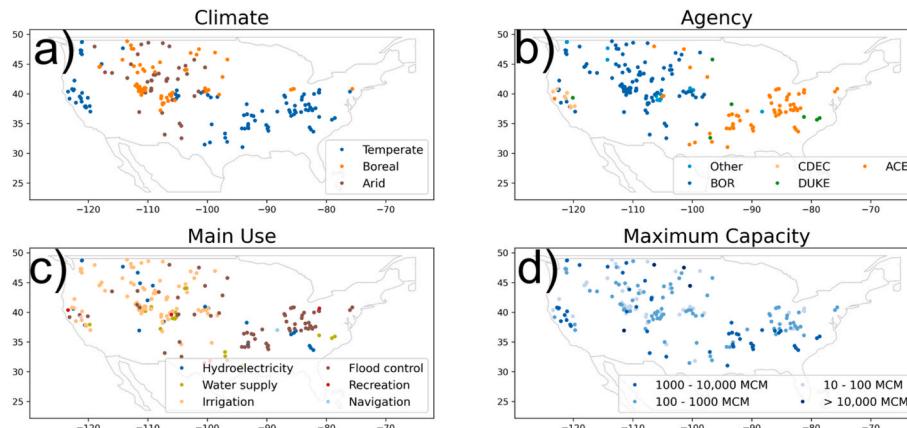


Fig. 1. Reservoirs in the ResOpsUS dataset in the US with color-coded reservoir's characteristics. (a) climate regions where reservoirs are located, (b) operating agency of the reservoirs, (c) main use of the reservoirs, and (d) Maximum capacity of the reservoirs.

aforementioned trained parameters. ISTARF is being used as a water management module in the Model for Scale Adaptive River Transport with Water Management (MOSART-WM (Thurber et al., 2021)).

2.3. Vanilla LSTM scheme for daily reservoir release

We used the Long Short-Term Memory (LSTM (Hochreiter and Schmidhuber, 1997)) to model daily reservoir releases. LSTM is a type of Recurrent Neural Network designed to learn temporal dependencies in input data. It addresses vanishing gradient and long-term dependency issues by introducing a memory cell. The memory cell is controlled by three gates: the input gate, the forget gate, and the output gate, which allow the LSTM to selectively learn and discard information, effectively capturing long-term dependencies. The LSTM maintains a hidden state, which is updated based on the current input, the previous hidden state, and the current memory cell state.

LSTM has been successfully applied to various applications in hydrology and water resources managements (Alizadeh et al., 2021; Bennett et al., 2024; Cheng et al., 2020; Cui et al., 2022; Fan et al., 2023; Fang et al., 2025; Hull et al., 2024; Kao et al., 2020; Kratzert et al., 2018; Kratzert et al., 2019; Le et al., 2019; Lees et al., 2021; Longyang and Zeng, 2023; Ni et al., 2020; Shen, 2018; Sit et al., 2020; Tran et al., 2021; Tran et al., 2024b; Yang et al., 2021; Zhang et al., 2018b). Similar to experiment 4 of Longyang and Zeng (2023), which applied LSTM for modeling daily reservoir releases across the US, this study trains an LSTM using predictors from the past seven days of inflow and storage to predict the next day's reservoir release. We used a single layer LSTM with a hidden size of 32 units which resulted in 4513 trainable parameters. Hereafter, this model will be referred to as LSTM. We trained the LSTM model using a batch size of 64 with shuffle mode enabled. The model was trained for 300 epochs, requiring approximately six hours of GPU time. The training process was conducted on the National Energy Research Scientific Computing Center's (NERSC) Perlmutter HPC.

2.4. Conditional LSTM

Motivated by the premise that given similar input timeseries, the LSTM model should generate different results for two different reservoirs that feature different reservoir attributes, we developed a new model that extends the LSTM model by adding reservoir attributes as predictors. We selected three basic reservoir characteristics to condition the LSTM model: main use, climate, and maximum capacity (Fig. 2). The reservoir's main use and maximum capacity were obtained from the GRanD dataset, while the climate classification was determined based on the reservoirs' locations relative to the Köppen-Geiger Climate Classification Map (Kottek et al., 2006).

To integrate the reservoir characteristics into the LSTM, we used the Conditional RNN for Keras tool from GitHub (https://github.com/phlipperemy/cond_rnn). This tool allows the transformation of condition variables into the correct shape to serve as the internal state of the LSTM. At the first timestep of the forward step, the vector containing the condition variables acts as the hidden state of the LSTM. This method

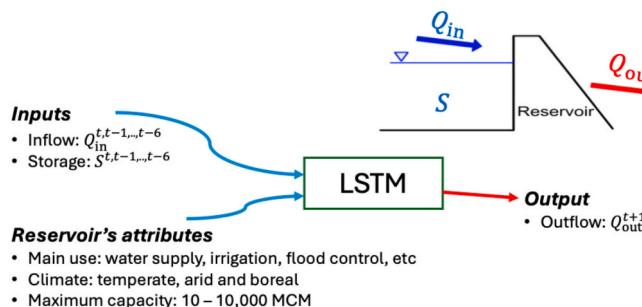


Fig. 2. Reservoir release prediction framework.

effectively conditions non-temporal inputs without mixing them with temporal inputs (Karpathy and Fei-Fei, 2015; Le et al., 2024; Vinyals et al., 2015). From this point forward in the manuscript, this model will be referred to as LSTM_cond. By adding an additional Dense layer to transform the condition variables for the LSTM model, the number of trainable parameters for LSTM_cond is increased to 4641 parameters. LSTM_cond was configured as similar to vanilla LSTM as possible so the effect of condition variables could be highlighted.

2.5. Evaluation metrics

To evaluate the predicted reservoir release performance, we used the Kling-Gupta Efficiency (KGE (Gupta et al., 2009; Kling et al., 2012)) and the Root Mean Squared Error (RMSE). While RMSE is used for bias quantification, KGE is used to evaluate the temporal agreement between observations and simulations. KGE has been proven to give a more balanced evaluation metric than the traditional Nash-Sutcliffe Efficiency (NSE (Nash and Sutcliffe, 1970)) (Mizukami et al., 2019; Santos et al., 2018). KGE is computed as follow:

$$\text{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (4)$$

Where r is the Pearson correlation coefficient, α represents the variability of prediction errors, β is a bias term. α and β are calculated as follow:

$$\alpha = \frac{\sigma_s}{\sigma_o} \quad (5)$$

$$\beta = \frac{\mu_s}{\mu_o} \quad (6)$$

Where σ_s and μ_s are variance and mean of the predicted timeseries, σ_o and μ_o are variance and mean of the observed timeseries.

2.6. Training, validation and test workflow

We selected reservoirs that have seven consecutive days of data for inflow and storage. Next, we filtered out reservoirs exhibiting outliers and unnatural behaviors (e.g. sudden, unexplained fluctuations in water levels that are not aligned with typical seasonal variations or operational rules) to ensure data reliability and consistency for model training and evaluation. Reservoirs were then further divided into in-sample and out-sample reservoirs.

In-sample reservoirs must represent all types of main uses, climates, and maximum capacities. The data for these reservoirs spans from January 1, 1980, to December 31, 2020, resulting in a total of 143 qualified in-sample reservoirs. Out-sample reservoirs were selected randomly but must be spatially heterogeneous and represent as many classes of each category as possible. The test period for these out-sample reservoirs includes all available data, totaling 20 reservoirs (Table 1).

Before feeding the data into the LSTM models, inflow, storage, and release values were normalized. Inflow and release were normalized by dividing them by the 80th quantile of historical inflow, while storage was normalized by dividing it by the corresponding maximum capacity. We chose the 80th quantile for normalizing inflow and release to reflect regular operation of the reservoirs as in Yassin et al. (2019). The training period for in-sample reservoirs spans from January 1980 to December

Table 1
Training, validation, testing periods for in-sample and out-sample reservoirs.

	In-sample reservoirs (n = 143)	Out-sample reservoirs (n = 20)
Training period	01-01-1980 – 12-31-2014	N/A
Validation period	01-01-2015 – 12-31-2015	N/A
Test period	01-01-2016 – 12-31-2020	01-01-1950 – 12-31-2020

2014. For the 143 reservoirs, there are a total of 2,295,729 training samples grouped into 35,871 batches, each containing sixty-four samples. Each training process includes 200 iterations, with validation occurring every 10 iterations.

Both LSTM and LSTM_cond were trained simultaneously on the National Energy Research Scientific Computing Center (NERSC)'s Perlmutter A100 GPUs, with one iteration taking 60 s and 70 s, respectively. The entire training process took 200 min for LSTM and 233 min for LSTM_cond, resulting in a Mean Absolute Error (MAE) of 0.006 and 0.0063, respectively, in the last iteration. Testing for in-sample reservoirs covers the period from January 2016 to December 2020, while testing for out-sample reservoirs spans from January 1950 to December 2020, depending on data availability for each reservoir.

3. Results

3.1. Test period (01-01-2016 – 12-31-2020) for in-sample reservoirs

For the test period, the median KGE for LSTM_cond, LSTM, and ISTARF is 0.708, 0.718, and 0.472, respectively. LSTM_cond achieves a KGE higher than 0.5 in more than 88 % of reservoirs, compared to 80 % for LSTM and 48 % for ISTARF (Fig. 3). It is worth noting that while both LSTM_cond and LSTM only have 6 out of 144 reservoirs with negative KGE, ISTARF has 15 reservoirs with negative KGE. Table 2 summarizes the KGE statistics for ISTARF, LSTM and LSTM_cond grouped by different classifications, namely, main use, climate, maximum capacity and operating agencies.

Regarding operating agencies, over 88 % of reservoirs are operated by either the Bureau of Reclamation (BOR) or the Army Corps of Engineers (ACE). ISTARF performs adequately for BOR-operated reservoirs, mostly located in the western half of the CONUS (Fig. 3), with a KGE mean of 0.515 and a median of 0.595 (Table 2). Specifically, ISTARF outperforms LSTM in a few reservoirs in the mountainous western regions, typically designed for irrigation (Fig. 4b). This is expected, as pointed out by Longyang and Zeng (2023), because with four trained parameters (α_1 , α_2 , β_1 , and β_2), ISTARF can approximate the long-term

mean inflow and release whereas daily scale LSTM models sometimes fail to capture the implicit long-term patterns. However, ISTARF performance declines for ACE-operated reservoirs in the eastern half of the CONUS which are mainly used for flood control (Fig. 3 and Fig. 4b). This could be due to ISTARF being developed for weekly reservoir release prediction. Although it can run on a daily scale, ISTARF often overestimates releases for flood control reservoirs. This behavior is further explained in section 3.2.

For in-sample reservoirs, LSTM_cond and LSTM show similar performance for the majority of reservoirs, both outperforming ISTARF. Similar to ISTARF, both LSTM and LSTM_cond are more skillful for ACE-operated than BOR-operated reservoirs, although the skill differences in terms of both the mean and median KGE between the two types of reservoirs are smaller compared to those of ISTARF. For climate clustering, LSTM performs better than LSTM_cond in the boreal climate zone, which is primarily in the Midwest of the CONUS. However, LSTM_cond also performs well, with a KGE median of 0.754 and a KGE mean of 0.702, compared to 0.785 and 0.716 for LSTM for boreal reservoirs. For temperate and arid regions, LSTM_cond produces releases more consistent with observations (KGE medians of 0.691 and 0.796 for temperate and arid regions, respectively) than those predicted by LSTM (KGE medians of 0.682 and 0.712 for temperate and arid regions, respectively) (Table 2).

Regarding reservoirs categorized by primary use, performances from LSTM and LSTM_cond show sharper contrasts than other classifications. Overall, LSTM_cond performs better for hydroelectric and irrigation reservoirs, while LSTM performs slightly better for water supply and flood control reservoirs (Fig. 4 and Table 2). Flood control reservoirs are more homogeneous in terms of their climate and maximum capacity, so adding reservoir characteristics introduces additional noise in the LSTM_cond predictions. This is shown in Fig. 4c for a representative flood control reservoir, Cecil M. Harden Lake Dam in Indiana where adding reservoir characteristics (LSTM_cond) does not significantly improve upon LSTM.

In contrast, for more diverse reservoir systems like those operated by BOR, LSTM_cond demonstrates its advantages (KGE median of 0.769

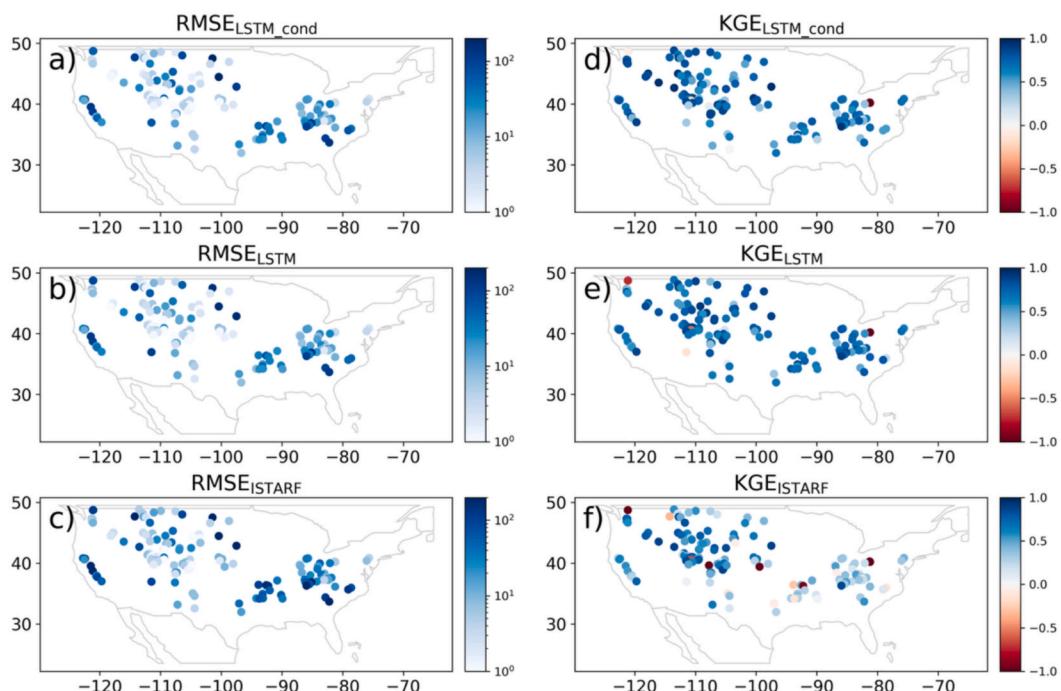


Fig. 3. Spatial distribution of Root Mean Squared Error (RMSE) and Kling-Gupta Efficiency (KGE) score for daily reservoir release using (a and d) conditioned LSTM (LSTM_cond), (b and e) LSTM, and (c and f) the Inferred Storage Targets and Release Functions (ISTARF). There are 143 in-sample reservoirs displayed in each subplot.

Table 2
Summary statistics of KGE (mean, median) for in-sample reservoirs classified by their main use, climate, operating agency, and maximum capacity. Bold numbers indicate the best value for LSTM_cond, LSTM and ISTARF.

Sub-class	No of res.	KGE median			KGE mean			RMSE median			RMSE mean		
		LSTM_cond		LSTM	LSTM_cond		LSTM	LSTM_cond		LSTM	LSTM_cond		LSTM
		KGE	median		KGE	median		KGE	median		KGE	median	
Main use	Hydroelectricity	18	0.775	0.726	0.477	0.651	0.587	0.383	50.39	44.91	63.93	53.26	57.90
	Water supply	12	0.587	0.789	0.541	0.712	0.458	4.025	2.702	4.872	17.02	13.11	27.78
	Irrigation	57	0.769	0.712	0.591	0.705	0.669	0.526	3.105	3.247	5.561	12.73	10.90
	Flood control	55	0.691	0.709	0.322	0.581	0.647	0.254	15.42	15.89	27.78	31.05	26.09
	Recreation	1	0.701	0.636	0.719	0.701	0.636	0.719	5.063	4.623	5.277	5.063	4.629
	Temperate	67	0.691	0.682	0.330	0.592	0.604	0.257	20.78	19.49	34.69	33.06	30.81
	Arid	29	0.769	0.712	0.591	0.678	0.666	0.529	5.840	6.273	9.355	25.42	21.61
	Boreal	47	0.754	0.785	0.607	0.702	0.716	0.522	3.046	2.623	5.061	13.54	11.88
	BOR	77	0.769	0.727	0.595	0.688	0.674	0.515	3.446	3.532	6.154	12.61	11.64
	ACE	50	0.691	0.711	0.319	0.585	0.639	0.246	20.92	19.90	36.05	40.96	35.40
Climate	CDEC	4	0.650	0.625	0.368	0.603	0.624	0.392	39.17	38.08	73.82	56.75	56.25
	Duke	3	0.528	0.767	0.138	0.592	0.599	0.138	39.42	46.64	75.29	39.74	35.49
	Other	9	0.831	0.768	0.650	0.668	0.586	0.347	5.063	6.962	11.00	26.31	64.45
	10–100 MCM	39	0.786	0.738	0.566	0.645	0.631	0.417	1.825	1.786	3.391	3.216	3.317
	100–1000 MCM	66	0.717	0.722	0.506	0.683	0.657	0.437	7.248	7.585	11.84	14.20	15.47
	1000–10,000 MCM	34	0.608	0.698	0.320	0.577	0.681	0.312	39.58	32.53	60.02	55.10	47.53
	>10,000 MCM	4	0.721	0.820	0.306	0.634	0.587	0.330	156.9	133.3	174.2	165.9	122.4

compared to 0.727 for LSTM) in adapting reservoir releases to each reservoir type (Table 2). For example, for Tulloch Reservoir in California, which is managed by storage-release rule curves for hydroelectricity purpose, the LSTM_cond prediction (green line in Fig. 4d) is much closer to the observed values. LSTM and ISTARF exhibit more generic release behaviors, leading to overpredictions of reservoir release during the summer months. This is because ISTARF and LSTM only consider inflow and storage as predictors, resulting in identical releases for reservoirs with similar time series of these predictors, regardless of their different climate regions and primary objectives. LSTM_cond, by integrating reservoir characteristics as its static predictors, achieves more accurate release predictions.

When characterizing reservoirs based on their maximum capacity, nearly half (66 over 143 reservoirs) have a maximum capacity between 100 and 1000 MCM. Smaller reservoirs (with capacities between 10–100 MCM) and larger reservoirs (with capacities greater than 1000 MCM) are represented by 39 and 38 reservoirs, respectively. Generally, LSTM_cond performed better than LSTM for reservoirs with capacities smaller than 1000 MCM, while LSTM outperformed LSTM_cond and ISTARF for larger ones (Table 2). This could be because larger reservoirs (> 1000MCM) often serve multiple purposes, such as irrigation, water supply, and flood control. When LSTM_cond is provided with only the main use of the reservoir, it may over- or under-predict when the reservoir serves other functions as well. Lastly, the KGE of LSTM appears to have narrower ranges than those of LSTM_cond, which might indicate overfitting to in-sample reservoirs—a topic to be discussed in the next section (Fig. S1).

3.2. All available data (01-01-1950 – 12-31-2020) for out-sample reservoirs

For out-of-sample reservoirs excluded from training for both LSTM and LSTM_cond, the KGE medians for LSTM_cond, LSTM, and ISTARF are 0.769, 0.746, and 0.520, respectively. Out of the 20 reservoirs, LSTM_cond has the highest KGE mean for 15 reservoirs and LSTM has the highest KGE mean for the remaining 5 reservoirs. ISTARF continues to perform well for the out-of-sample reservoirs in the western half of the CONUS (Fig. 5). In the Alcova reservoir, LSTM predictions (blue line) are closer to observations during the summer months but tend to under/overestimate during the spring and fall. LSTM_cond estimates (green line) for Alcova reservoir mitigate the under/overestimations observed in the LSTM predictions. For Echo reservoir, the ISTARF estimates are closer to observations than those from LSTM but not as close as those from LSTM_cond. For reservoirs operated by ACE, the KGE mean for ISTARF is 0.25, while the KGE means for LSTM and LSTM_cond are significantly higher at 0.759 and 0.690, respectively. One reason for ISTARF's low performance is its unusually high release peaks during flood seasons, such as for the Green River Lake Dam (Fig. 5f). Additionally, ISTARF estimates lower releases for reservoirs like Lake Mohawk Dam, whose primary use is recreation.

LSTM_cond consistently outperforms LSTM for hydroelectric, water supply, irrigation, and recreation reservoirs (Fig. 7). The KGE median values for LSTM_cond for these reservoir types are 0.764, 0.565, 0.821, and 0.779, respectively while the corresponding KGE values for LSTM are 0.737, 0.413, 0.775, and 0.713. LSTM_cond is better generalized for out-of-sample reservoirs, such as the Alcova and Echo irrigation reservoirs (Fig. 5d and 5e). Fig. 6 also shows similar RMSE scores between LSTM and LSTM_cond and a higher RMSE score in ISTARF.

For flood control reservoirs, LSTM remains the best model for predicting releases. However, the gap in KGE means between LSTM and LSTM_cond for out-of-sample reservoirs is 0.040, smaller than the 0.066 gap for in-sample reservoirs. This is because flood control reservoirs, like Cecil M. Harden Lake Dam (Fig. 4c) and Green River Lake Dam (Fig. 5f), rely primarily on inflow information for release due to the nature of their operation. For such reservoirs, adding information beyond inflow introduces additional noise in the prediction. Coincidentally, for out-of-

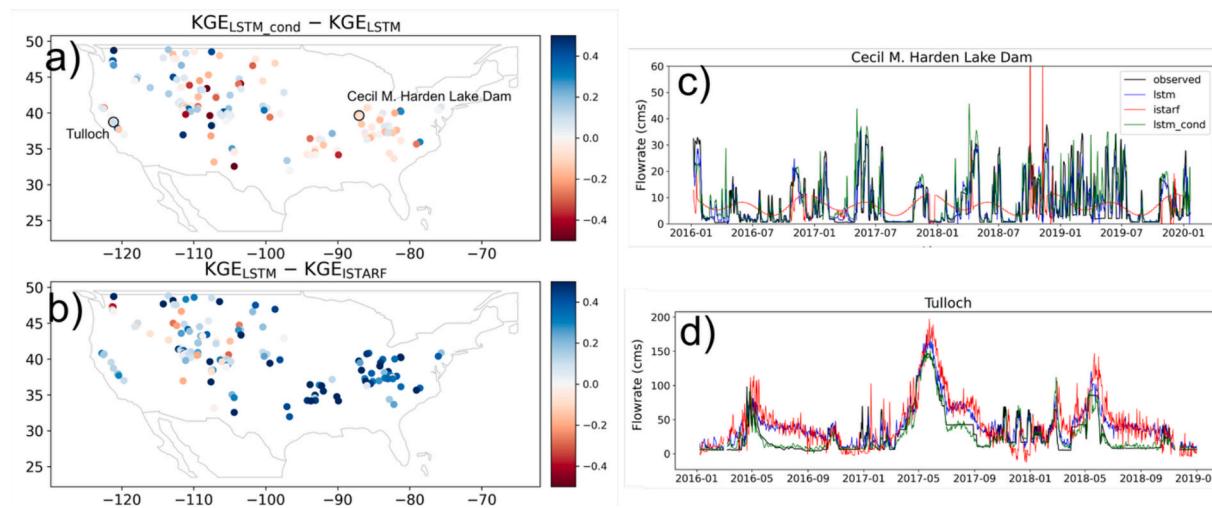


Fig. 4. Spatial distribution of difference in Kling-Gupta Efficiency (KGE) score for daily reservoir release between (a) conditioned LSTM vs LSTM and (b) LSTM vs ISTARF. There are 143 in-sample reservoirs displayed in each subplot. Subplots (c) and (d) show the observed and simulated reservoir releases from LSTM (blue line), ISTARF (red line), and LSTM_{cond} (green line) for the test period from January 2016 to December 2020. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

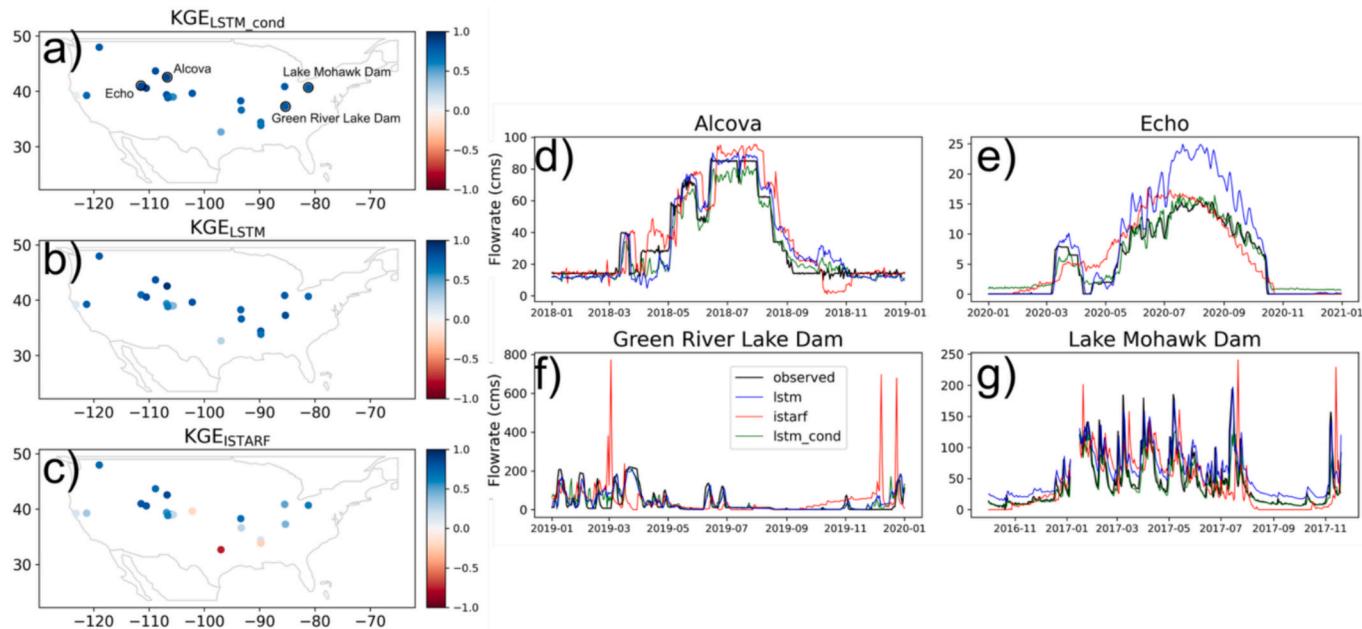


Fig. 5. Spatial distribution of Kling-Gupta Efficiency (KGE) score for daily reservoir release using (a) conditioned LSTM ($LSTM_{cond}$), (b) LSTM, and (c) the Inferred Storage Targets and Release Functions (ISTARF). There are 20 out-sample reservoirs displayed in each subplot. For illustration, subplots (d), (e), (f) and (g) show the observed and simulated reservoir releases from LSTM (blue line), ISTARF (red line), and $LSTM_{cond}$ (green line) for the last year of the test period. Locations and names of representative reservoirs are shown in subplot (a). Please note that for out-of-sample reservoirs, the test period spans the entire available data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sample reservoirs, most large capacity reservoirs (>1000 MCM) are flood control reservoirs. This explains why LSTM (blue box) performs better than $LSTM_{cond}$ (green box) for 1000–10,000 MCM reservoirs (Fig. 7d).

4. Discussions

4.1. Performances of ISTARF and LSTM models

Policy-based approaches for reservoir release have a long history of development and have proven effective for various applications such as addressing water scarcity and ensuring power grid reliability. They have

also been used in global hydrologic models to capture the influence of water management on water availability and floods (e.g., Hanasaki et al. 2006). One such approach, ISTARF, performs well for irrigation reservoirs in the western half of the CONUS. However, because ISTARF was developed for weekly reservoir release prediction, it often overestimates releases during flood seasons (Fig. 5f) and exhibits generic behavior for flood control reservoirs (Fig. 4c).

In contrast, data-driven approaches for reservoir release are relatively new but have demonstrated robustness. In this study, LSTM and $LSTM_{cond}$ outperform ISTARF for the majority of reservoirs in the CONUS (Fig. 8). Despite their success, data-driven models are not without limitations. One significant drawback is their black-box nature,

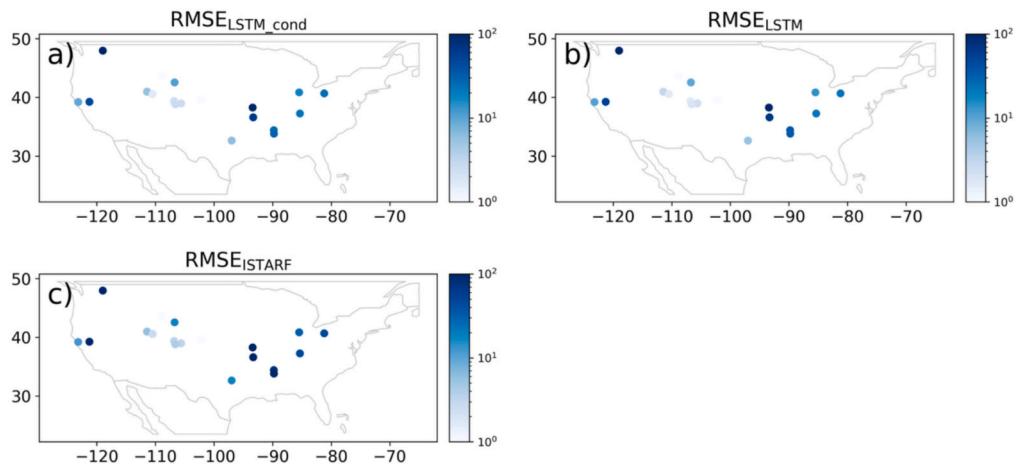


Fig. 6. Spatial distribution of RMSE score for daily reservoir release using (a) conditioned LSTM (LSTM_cond), (b) LSTM, and (c) the Inferred Storage Targets and Release Functions (ISTARF). There are 20 out-sample reservoirs displayed in each subplot. Please note that for out-of-sample reservoirs, the test period spans the entire available data.

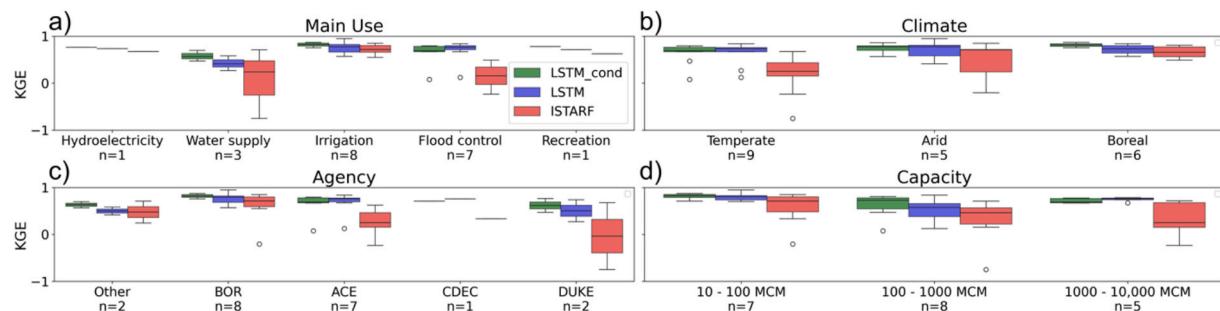


Fig. 7. Box plot of KGE scores for daily reservoir release using LSTM_cond (green boxes), LSTM (blue boxes), and ISTARF (red boxes). KGE scores are for out-sample reservoirs. Each subplot represents one attribute of the reservoirs. For illustration, subplots a, b, c, and d group reservoirs based on their main use, climate, operating agencies and maximum capacity, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

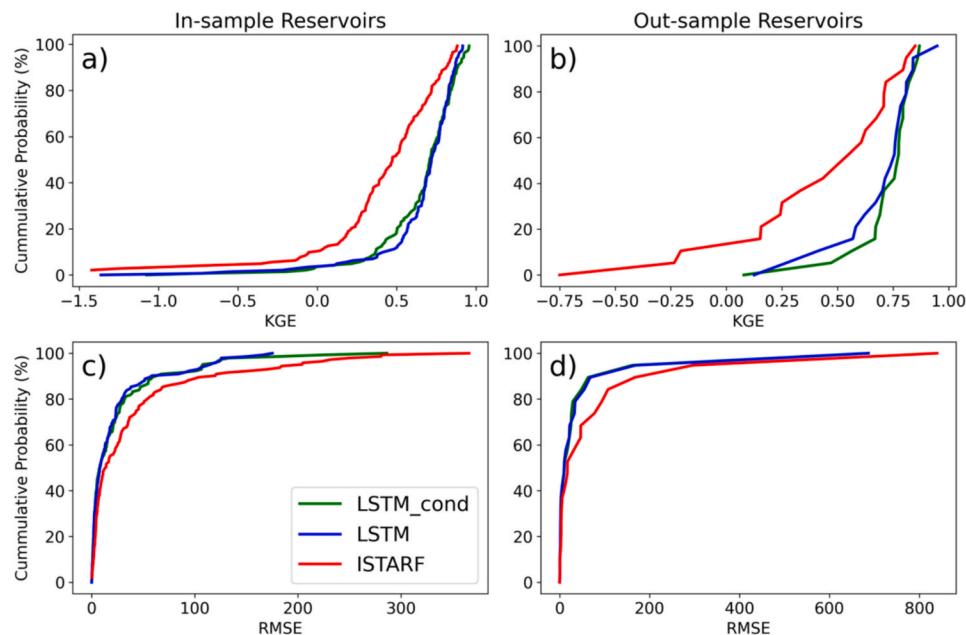


Fig. 8. Subplots (a) and (b) show the probability of exceedance of KGE, subplots (c) and (d) show the probability of exceedance of RMSE for in-sample and out-sample reservoirs, respectively. Green, blue and red lines are the probability of exceedance of KGE of LSTM_cond, LSTM, and ISTARF, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

making it unclear which predictors guarantee optimal predictions and how they function within the models. Efforts to represent the hierarchical nature of anthropogenic influences within these models have been made (Li et al., 2024; Longyang and Zeng, 2023), but challenges remain. Our study proposes an additional solution by integrating reservoir characteristics into the LSTM framework.

When comparing the performance of LSTM and LSTM_cond, both models show similar results for in-sample reservoirs. However, for out-sample reservoirs, LSTM_cond generally performs better across most categories, except for flood control reservoirs. Due to the nature of flood control operations, these reservoirs rely primarily on recent inflow and

storage information, making the addition of other data potentially noisy and counterproductive. A potential resolution could involve adding a consecutive LSTM layer to filter out these extraneous noises.

For irrigation reservoirs, the performance of LSTM drops significantly from in-sample to out-sample reservoirs. LSTM's KGE means are 0.712 for in-sample and 0.422 for out-sample reservoirs, suggesting that LSTM may overfit to the specific patterns and characteristics present in the in-sample data, thereby limiting its generalizability and robustness.

On the other hand, LSTM_cond demonstrates more consistent performance across both test sets, highlighting its superior generalization capability. For in-sample irrigation reservoirs, LSTM_cond maintains a

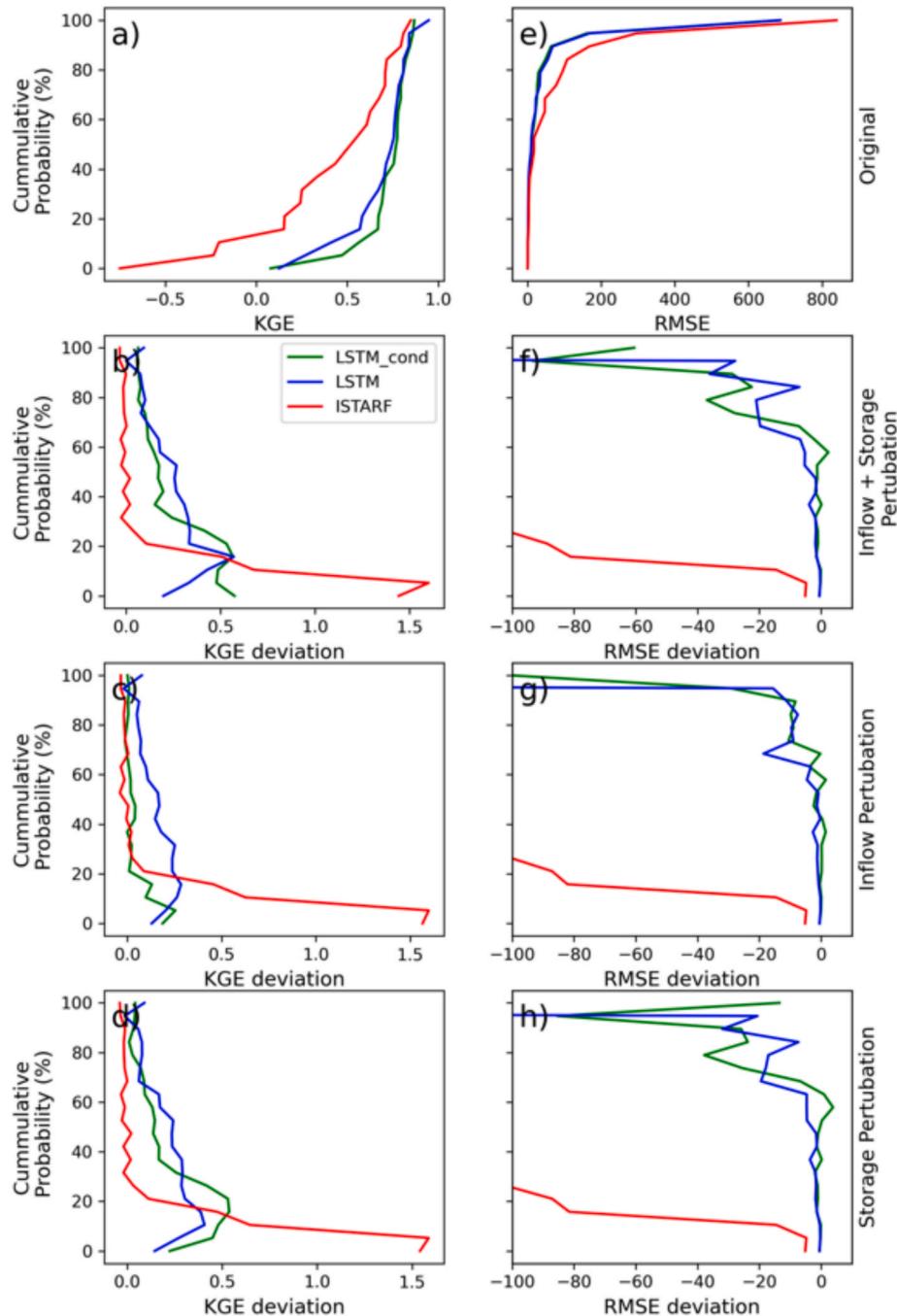


Fig. 9. Probability of exceedance of KGE (subplot a) and RMSE (subplot e) for perturbed LSTM_cond (green), LSTM (blue), ISTARF (red) for out-sample reservoirs. Subplots b,c,d show the deviation of KGE compared to the original inputs when perturbing both inflow and storage, inflow, and storage, respectively. Subplots f,g,h show the deviation of RMSE compared to the original inputs when perturbing both inflow and storage, inflow, and storage, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

high KGE mean, and this performance does not drop significantly when applied to out-sample reservoirs. This consistency is likely due to LSTM_cond's ability to incorporate additional reservoir characteristics as part of its initial state, allowing it to adapt better to various reservoir conditions and management practices.

4.2. Sensitivities of LSTM models to noises from predictors

The predictors used in this study are based on observations. However, if both LSTM and LSTM_cond are implemented as components in a large-scale hydrologic model, uncertainties in the simulated inflows and reservoir storages could contribute to uncertainty in modeling reservoir release. Therefore, we further explored the sensitivity of both LSTM and LSTM_cond to errors in their predictors, specifically inflow and storage. The following experiment is designed to test for hydrologic model integration.

We introduced Gaussian noise to the inflow and storage data for out-of-sample reservoirs and compared the results with those obtained using the “observed” predictors. For normalized inflow, we added noise with a mean of 0 and a standard deviation of 0.1, and for normalized storage, we added noise with a mean of 0 and a standard deviation of 0.05. Parameters for Gaussian noise of inflow were chosen from Chen et al. (2016). Since there is no existing reference regarding noise for reservoir storage, we decided to set its standard deviation to half that of the inflow.

As expected, the results from perturbed predictors are worse than those using observations as predictors. However, LSTM_cond is less sensitive to random errors in predictors compared to LSTM (Fig. 9). The results from perturbed storage closely resemble those obtained when both predictors are perturbed, indicating that both LSTM models are particularly sensitive to errors in reservoir storage. Lastly, the KGE for LSTM_cond with only perturbed inflow is only slightly lower than that without perturbation, which contrasts obviously with the more significant decrease in KGE observed in LSTM. This sensitivity analysis shows that by including reservoir attributes, LSTM_cond not only improves generalizability but also reduce sensitivity to the predictors compared to LSTM. It is worth noting that the ISTARF module is already integrated into a larger hydrologic and water management model, namely MOSART-WM (Thurber et al., 2021). The sensitivity test that includes ISTARF is provided at the appendix for readers interested in further details.

4.3. Mass balance conservation in the LSTM models

Data-driven models often face the challenge of lacking physical constraints. Unlike ISTARF, which constrains reservoir releases to not exceed the available water in storage plus inflow (Turner et al., 2021b), both LSTM and LSTM_cond do not include the mass balance constraint. To explore more, we counted the number of reservoirs with at least one day when releases exceeded the sum of inflow and storage (Fig. 10) using observations of reservoir storage (ResOpsUS data – blue bars) and releases generated by LSTM_cond (orange bars) and ISTARF (green bars).

It is worth noting that while the mass balance rule was conserved for most of the reservoirs during their operations, we noticed the mass balance conservation was violated in some flood control and irrigation reservoirs from the ResOpsUS dataset (blue bars – Fig. 10). This could be explained by additional water releases from leeways might not be recorded. These operational behaviors are reflected in the ResOpsUS data but are absent from ISTARF simulations due to the imposed mass balance constraint. Since LSTM_cond learns directly from ResOpsUS data, its outputs closely mirror the observed operational behaviors.

For LSTM_cond results, the number of reservoirs violating the mass balance is 2 (1.3 %) for in-sample reservoirs and 7 (35 %) for out-of-sample reservoirs. This suggests that most reservoirs simulated by LSTM_cond still follow the mass balance constraint. Future studies could incorporate more flexible physical constraints into deep learning models to ensure compliance with fundamental physical laws.

We further analyzed the mass deficit by calculating the long-term mean daily change in storage and the difference between inflow and outflow for each reservoir in the study. Under normal conditions, daily changes in storage should almost always be equal to the difference between inflow and outflow. Exceptions may occur in irrigation reservoirs, where additional water releases from leeways might not be recorded. Results from Fig. 11 show a strong correlation between these two variables in both the original dataset (ResOpsUS) and the reservoir releases modeled by LSTM, LSTM_cond, and ISTARF. ISTARF releases adhere to a strict mass balance constraint (i.e. most of the reservoirs are under the 1:1 line – Fig. 11d), while LSTM and LSTM_cond releases closely follow the observed dataset (Fig. 11a). LSTM_cond exhibits a higher correlation than LSTM for in-sample reservoirs but a lower correlation for out-of-sample reservoirs (Fig. 11b and 11c). These results suggest that the data-driven LSTM and LSTM_cond models conserve mass over the long term to a similar degree as observations (ResOpsUS) and a model

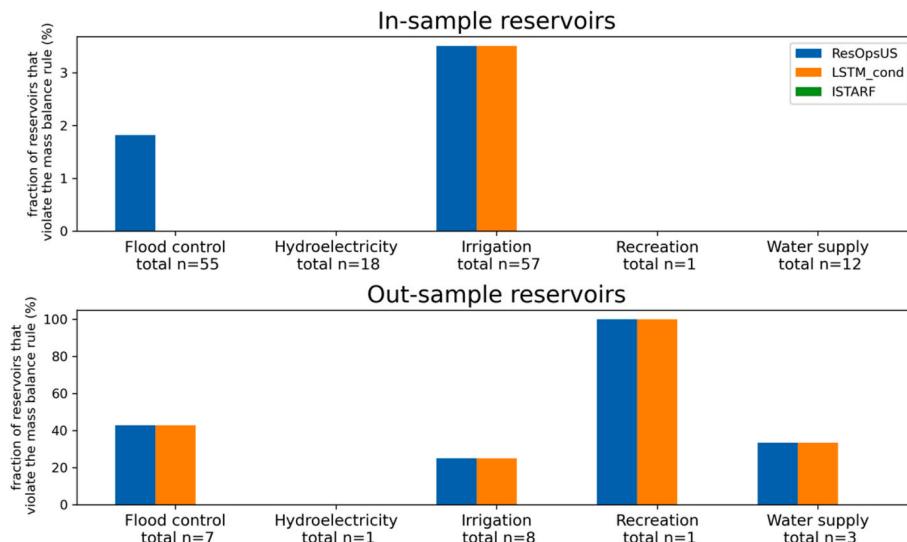


Fig. 10. Fraction of reservoirs that violate the simple mass balance constraint (i.e., releases must not exceed the sum of inflow and storage), categorized into five main use groups for both in-sample and out-of-sample reservoirs.

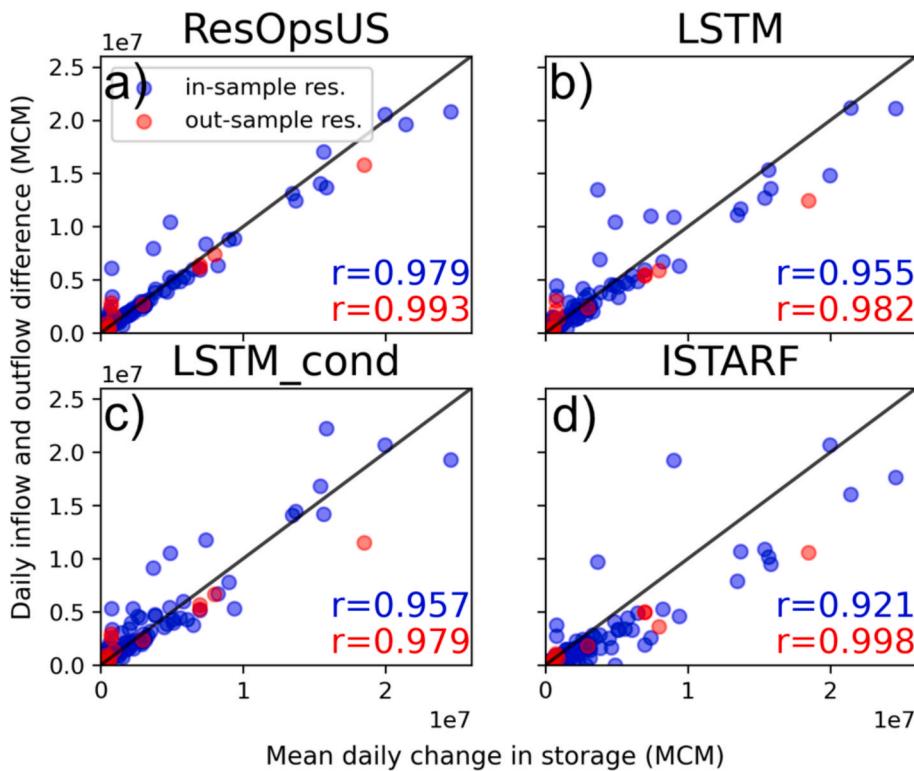


Fig. 11. Scatter plots showing the mean daily change in storage (MCM) versus the mean difference between daily inflow and outflow for each reservoir. Outflow values are taken-modeled from the ResOpsUS dataset (a), by LSTM (b), LSTM_cond (c), and ISTARF (d).

(ISTARF) that includes a mass balance constraint.

4.4. The effectiveness of LSTM in reservoir release modeling compared with a simpler ML model

Quilty et al. (2023) provided key insights into Bayesian Extreme Learning Machines (BELM), emphasizing their computational efficiency and ability to produce more accurate and reliable streamflow predictions compared to LSTM. They argued that simpler models utilizing Bayesian inference to estimate output layer weights could outperform LSTM in many cases (Quilty et al., 2023).

This motivates us to assess the performance of a widely used machine learning model, Random Forest (RF), for reservoir release predictions and compare its results with those from our previous analysis. Specifically, we focus on reservoirs where the vanilla LSTM performs comparably or better than LSTM_cond, namely irrigation and flood control reservoirs. We implement a simple RF model with 100 trees and use 18 input features, including the inflow and storage of the previous seven days, along with the four reservoir characteristics used in LSTM_cond.

In Fig. 12, RF performs comparably to LSTM models for irrigation reservoirs. However, for flood control reservoirs, while RF outperforms ISTARF, its results are consistently inferior to those of the LSTM models. This is expected as these two types of reservoirs have fundamentally different operational characteristics: Flood control reservoirs (for example: Cecil M. Harden Lake Dam (Fig. 4c) and Green River Lake Dam (Fig. 5f)) operate under stricter, event-driven protocols that are better captured by LSTM's temporal memory mechanisms, while irrigation reservoirs (for example: Alcova and Echo irrigation reservoirs (Fig. 5d and 5e)) exhibit more predictable, seasonal patterns that align well with RF's structure. We would expect that Bayesian inference could help to improve the performance of both the models.

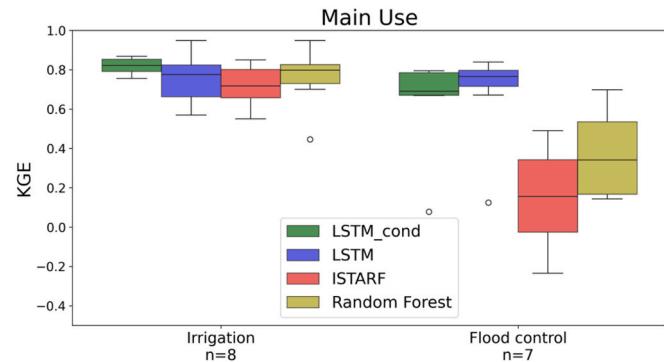


Fig. 12. Box plot of KGE scores for daily reservoir release using LSTM_cond (green boxes), LSTM (blue boxes), ISTARF (red boxes) and Random Forest (yellow boxes). KGE scores are for out-sample reservoirs. This figure groups reservoirs based on two of their main use, namely, irrigation and flood control, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Conclusions

In this study, we developed and evaluated both policy-based and data-driven approaches to model daily reservoir release across the contiguous United States. The policy-based approach, ISTARF, while effective in some contexts, often overestimate releases during flood seasons and exhibit generic behaviors for flood control reservoirs. On the other hand, data-driven approaches, specifically, LSTM and its conditional variant (LSTM_cond) have outperformed ISTARF in most scenarios, demonstrating their robustness and adaptability.

Our approach proposes a new method in global reservoir modeling with a goal of integrating the model into Earth system models. Building upon the current successful LSTM model, we extended its capability to

account for the reservoir's attributes and quantified the uncertainty in reservoir outflow predictions resulting from input uncertainties. Introducing Conditional LSTM to simulate reservoir outflow has three benefits: it improves the performance of outflow simulation, provides insights into the factors contributing to this improvement through the incorporation of the reservoir characteristics, and takes a step towards making LSTM more explainable.

We observed a notable drop in LSTM performance when transitioning from in-sample to out-sample reservoirs, with the KGE means decreasing from 0.712 to 0.422. This suggests a potential overfitting issue, highlighting the need for models that generalize better across diverse reservoirs. In contrast, LSTM_cond exhibited more consistent performance across both in-sample and out-sample test sets, likely due to its ability to incorporate reservoir-specific characteristics into its initial state, thus better capturing the variability in reservoir operations and conditions. LSTM_cond is also less sensitive to errors in inflow compared to LSTM. However, for a specific type of reservoirs, flood control or potentially, multi-purpose reservoirs, LSTM still has a better performance than LSTM_cond due to additional noise from adding reservoir characteristics that do not play an important role in flood control/multi-purpose reservoirs.

Overall, while advanced machine learning models have shown their potential in providing more accurate and reliable predictions of reservoir releases, there are challenges in the interpretability of these models and the need for further research to integrate hierarchical anthropogenic influences effectively. Our study contributes to this ongoing effort by demonstrating the benefits of conditioning LSTM models on reservoir-specific characteristics and suggesting potential improvements needed for handling noise in flood control reservoirs. Future work will focus on refining these models, exploring additional reservoir characteristics, land use/land cover change scenarios (Tran et al., 2024a), and extending the methodology to other regions and hydrological contexts (e.g., climate change (Leonarduzzi et al., 2022) or excessive evapotranspiration (Tran et al., 2023)).

While the method proposed in this study generates single deterministic forecasts, we recognize that probabilistic forecasts are often preferred in many streamflow forecasting applications (Quilty et al., 2023; Tran et al., 2024b). In future work, we will explore alternative approaches, such as using random initial perturbations or Bayesian inference. As noted by (Tran and Kim, 2022), quantifying uncertainty through initial condition perturbations offers computational advantages compared to addressing uncertainty from model structural differences and other sources. Lastly, we acknowledge that the only metric KGE presented so far is an aggregated metric from correlation, prediction residuals, and bias. Previous studies on data-driven methods for reservoir modeling have used Nash-Sutcliffe Efficiency (NSE) and KGE interchangeably to prove the model's effectiveness and accuracy as these metrics have been widely accepted. While no single metric is considered universal, to ensure our results are more comparable to other studies in the field, we also included an additional metric, NSE in the Appendix. The results for NSE are very similar to those obtained using the KGE metric.

CRediT authorship contribution statement

Hoang Tran: Writing – review & editing, Visualization, Software, Investigation, Conceptualization, Writing – original draft, Validation, Methodology, Formal analysis. **Tian Zhou:** Writing – review & editing, Methodology, Writing – original draft, Formal analysis. **Zeli Tan:** Writing – original draft, Writing – review & editing, Formal analysis. **Yilin Fang:** Writing – original draft, Conceptualization, Writing – review & editing, Formal analysis. **L. Ruby Leung:** Writing – original draft, Project administration, Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The ResOpsUS dataset is available at: <https://zenodo.org/records/5893641>. Reservoir characteristics dataset is available at: <https://www.globaldamwatch.org/grand>. The model code and compilation script are available at GitHub (https://github.com/philipperemy/cond_rnn).

References

- Aboutalebi, M., Bozorg Haddad, O., Loáiciga, H.A., 2015. Optimal monthly reservoir operation rules for hydropower generation derived with SVR-NSGAII. *J. Water Resour. Plan. Manag.* 141 (11), 04015029.
- Alizadeh, B., Ghaderi Bafti, A., Kamangir, H., Zhang, Y., Wright, D.B., Franz, K.J., 2021. A novel attention-based LSTM cell post-processor coupled with bayesian optimization for streamflow prediction. *J. Hydrol.* 601, 126526. <https://doi.org/10.1016/j.jhydrol.2021.126526>.
- Bennett, A., Tran, H., De la Fuente, L., Triplett, A., Ma, Y., Melchior, P., Maxwell, R.M., Condon, L.E., 2024. Spatio-temporal machine learning for regional to continental scale terrestrial hydrology. *J. Adv. Model. Earth Syst.* 16 (6) e2023MS004095.
- Boulange, J., Hanasaki, N., Yamazaki, D., Pokhrel, Y., 2021. Role of dams in reducing global flood exposure under climate change. *Nat. Commun.* 12 (1), 417.
- Chen, L., Singh, V.P., Lu, W., Zhang, J., Zhou, J., Guo, S., 2016. Streamflow forecast uncertainty evolution and its effect on real-time reservoir operation. *J. Hydrol.* 540, 712–726. <https://doi.org/10.1016/j.jhydrol.2016.06.015>.
- Chen, T., Song, C., Zhan, P., Yao, J., Li, Y., Zhu, J., 2022a. Remote sensing estimation of the flood storage capacity of basin-scale lakes and reservoirs at high spatial and temporal resolutions. *Sci. Total Environ.* 807, 150772. <https://doi.org/10.1016/j.scitotenv.2021.150772>.
- Chen, Y., Li, D., Zhao, Q., Cai, X., 2022b. Developing a generic data-driven reservoir operation model. *Adv. Water Resour.* 167, 104274. <https://doi.org/10.1016/j.advwatres.2022.104274>.
- Cheng, M., Fang, F., Kinouchi, T., Navon, I.M., Pain, C.C., 2020. Long lead-time daily and monthly streamflow forecasting using machine learning methods. *J. Hydrol.* 590, 125376. <https://doi.org/10.1016/j.jhydrol.2020.125376>.
- Coerver, H.M., Rutten, M.M., van de Giesen, N.C., 2018. Deduction of reservoir operating rules for application in global hydrological models. *Hydrol. Earth Syst. Sci.* 22 (1), 831–851. <https://doi.org/10.5194/hess-22-831-2018>.
- Cui, Z., Zhou, Y., Guo, S., Wang, J., Xu, C.-Y., 2022. Effective improvement of multi-step-ahead flood forecasting accuracy through encoder-decoder with an exogenous input structure. *J. Hydrol.* 609, 127764. <https://doi.org/10.1016/j.jhydrol.2022.127764>.
- Dong, N., Guan, W., Cao, J., Zou, Y., Yang, M., Wei, J., Chen, L., Wang, H., 2023. A hybrid hydrologic modelling framework with data-driven and conceptual reservoir operation schemes for reservoir impact assessment and predictions. *J. Hydrol.* 619, 129246. <https://doi.org/10.1016/j.jhydrol.2023.129246>.
- Ehsani, N., Vörösmarty, C.J., Fekete, B.M., Stakhiv, E.Z., 2017. Reservoir operations under climate change: storage capacity options to mitigate risk. *J. Hydrol.* 555, 435–446.
- EIA. (2023). Renewable generation surpassed coal and nuclear in the U.S. electric power sector in 2022. Retrieved August 1 from <https://www.eia.gov/todayinenergy/detail.php?id=55960>.
- Fan, M., Zhang, L., Liu, S., Yang, T., Lu, D., 2023. Investigation of hydrometeorological influences on reservoir releases using explainable machine learning methods [Original Research]. *Front. Water* 5. <https://doi.org/10.3389/frwa.2023.1112970>.
- Fang, Y., Tran, H.V., Leung, L.R., 2025. Subsurface hydrological controls on the short-term effects of hurricanes on nitrate-nitrogen runoff loading: a case study of Hurricane Ida using the Energy Exascale Earth System Model (E3SM) land Model (v2.1). *Geosci. Model Dev.* 18 (1), 19–32. <https://doi.org/10.5194/gmd-18-19-2025>.
- Forsberg, B.R., Melack, J.M., Dunne, T., Barthem, R.B., Goulding, M., Paiva, R.C., Sorribas, M.V., Silva Jr, U.L., Weisser, S., 2017. The potential impact of new andean dams on Amazon fluvial ecosystems. *PLoS One* 12 (8), e0182254.
- Gangrade, S., Lu, D., Kao, S.-C., Painter, S.L., 2022. Machine learning assisted reservoir operation model for long-term water management simulation. *JAWRA J. Am. Water Resour. Associat.* 58 (6), 1592–1603. <https://doi.org/10.1111/1752-1688.13060>.
- Giuliani, M., Lamontagne, J., Reed, P., Castelletti, A., 2021. A state-of-the-art review of optimal reservoir control for managing conflicting demands in a changing world. *Water Resour. Res.* 57 (12) e2021WR029927.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* 377 (1), 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Hanasaki, N., Kanae, S., Oki, T., 2006. A reservoir operation scheme for global river routing models. *J. Hydrol.* 327 (1), 22–41. <https://doi.org/10.1016/j.jhydrol.2005.11.011>.

- 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 Resour. Manag.* 27, 3803–3823.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Hull, R., Leonarduzzi, E., De La Fuente, L., Viet Tran, H., Bennett, A., Melchior, P., Maxwell, R.M., Condon, L.E., 2024. Simulation-based inference for parameter estimation of complex watershed simulators. *Hydrol. Earth Syst. Sci.* 28 (20), 4685–4713. <https://doi.org/10.5194/hess-28-4685-2024>.
- Kao, I.F., Zhou, Y., Chang, L.-C., Chang, F.-J., 2020. Exploring a long short-term memory based encoder-decoder framework for multi-step-ahead flood forecasting. *J. Hydrol.* 583, 124631. <https://doi.org/10.1016/j.jhydrol.2020.124631>.
- Karpathy, A., Fei-Fei, L., 2015. Deep visual-semantic alignments for generating image descriptions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.* 424–425, 264–277. <https://doi.org/10.1016/j.jhydrol.2012.01.011>.
- Klipsch, J. D., & Hurst, M. B. (2007). HEC-ResSim reservoir system simulation user's manual version 3.0. USACE, Davis, CA, 512.
- Kottke, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M., 2018. Rainfall-runoff modelling using long short-term memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* 22 (11), 6005–6022.
- Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G., 2019. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrol. Earth Syst. Sci.* 23 (12), 5089–5110.
- Le, P.V.V., Rathore, S.S., Coon, E.T., Ward, A., Haggerty, R., Painter, S.L., 2024. Hydrologic connectivity and dynamics of solute transport in a mountain stream: Insights from a long-term tracer test and multiscale transport modeling informed by machine learning. *J. Hydrol.* 639, 131562. <https://doi.org/10.1016/j.jhydrol.2024.131562>.
- Le, X.-H., Ho, H.V., Lee, G., Jung, S., 2019. Application of long short-term memory (LSTM) neural network for flood forecasting. *Water* 11 (7), 1387.
- Lees, T., Reece, S., Kratzert, F., Klotz, D., Gauch, M., De Brujin, J., Kumar Sahu, R., Greve, P., Slater, L., Dadson, S., 2021. Hydrological concept formation inside long short-term memory (LSTM) networks. *Hydrol. Earth Syst. Sci. Discuss.* 2021, 1–37.
- Lehner, B., Liermann, C.R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J.C., Rödel, R., Sindorf, N., Wisser, D., 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Front. Ecol. Environ.* 9 (9), 494–502. <https://doi.org/10.1890/100125>.
- Leonarduzzi, E., Tran, H., Bansal, V., Hull, R.B., De la Fuente, L., Bearup, L.A., Melchior, P., Condon, L.E., Maxwell, R.M., 2022. Training machine learning with physics-based simulations to predict 2D soil moisture fields in a changing climate [Original Research]. *Front. Water* 4. <https://doi.org/10.3389/frwa.2022.927113>.
- Li, D., Chen, Y., Lyu, L., Cai, X., 2024. Uncovering historical reservoir operation rules and patterns: insights from 452 large reservoirs in the contiguous United States. *e2023WR036686* *Water Resour. Res.* 60 (8). <https://doi.org/10.1029/2023WR036686>.
- Lin, J.-Y., Cheng, C.-T., Chau, K.-W., 2006. Using support vector machines for long-term discharge prediction. *Hydrol. Sci. J.* 51 (4), 599–612.
- Liu, J., Dietz, T., Carpenter, S.R., Folke, C., Alberti, M., Redman, C.L., Schneider, S.H., Ostrom, E., Pell, A.N., Lubchenco, J., 2007. Coupled human and natural systems. *AMBIO: A J. Human Environ.* 36 (8), 639–649.
- Longyang, Q., Zeng, R., 2023. A hierarchical temporal scale framework for data-driven reservoir release modeling. *e2022WR033922* *Water Resour. Res.* 59 (6). <https://doi.org/10.1029/2022WR033922>.
- Mateo, C.M., Hanasaki, N., Komori, D., Tanaka, K., Kiguchi, M., Champathong, A., Sukhapunphanaph, T., Yamazaki, D., Oki, T., 2014. Assessing the impacts of reservoir operation to floodplain inundation by combining hydrological, reservoir management, and hydrodynamic models. *Water Resour. Res.* 50 (9), 7245–7266. <https://doi.org/10.1002/2013WR014845>.
- Mizukami, N., Rakovec, O., Newman, A.J., Clark, M.P., Wood, A.W., Gupta, H.V., Kumar, R., 2019. On the choice of calibration metrics for “high-flow” estimation using hydrologic models. *Hydrol. Earth Syst. Sci.* 23 (6), 2601–2614. <https://doi.org/10.5194/hess-23-2601-2019>.
- Moran, E.F., Lopez, M.C., Moore, N., Müller, N., Hyndman, D.W., 2018. Sustainable hydropower in the 21st century. *Proc. Natl. Acad. Sci.* 115 (47), 11891–11898.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — a discussion of principles. *J. Hydrol.* 10 (3), 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- Ni, L., Wang, D., Singh, V.P., Wu, J., Wang, Y., Tao, Y., Zhang, J., 2020. Streamflow and rainfall forecasting by two long short-term memory-based models. *J. Hydrol.* 583, 124296. <https://doi.org/10.1016/j.jhydrol.2019.124296>.
- Ortiz-Partida, J.P., Lane, B., Sandoval-Solis, S., 2016. Economic effects of a reservoir operation policy in the Rio Grande/Bravo for integrated human and environmental water management. *J. Hydrol.: Reg. Stud.* 8, 130–144.
- Patterson, L.A., Doyle, M.W., 2018. A nationwide analysis of US Army Corps of Engineers reservoir performance in meeting operational targets. *JAWRA J. Am. Water Resour. Assoc.* 54 (2), 543–564.
- Quilty, J., Jahangir, M.S., You, J., Hughes, H., Hah, D., Tzoganakis, I., 2023. Bayesian extreme learning machines for hydrological prediction uncertainty. *J. Hydrol.* 626, 130138. <https://doi.org/10.1016/j.jhydrol.2023.130138>.
- Santos, L., Thirel, G., Perrin, C., 2018. Technical note: pitfalls in using log-transformed flows within the KGE criterion. *Hydrol. Earth Syst. Sci.* 22 (8), 4583–4591. <https://doi.org/10.5194/hess-22-4583-2018>.
- Shen, C., 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour. Res.* 54 (11), 8558–8593.
- Simonovic, S.P., 1992. Reservoir systems analysis: closing gap between theory and practice. *J. Water Resour. Plan. Manag.* 118 (3), 262–280.
- 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 Sci. Technol.* 82 (12), 2635–2670.
- Steyaert, J.C., Condon, L.E., Turner, W.D., Voisin, N., 2022. ResOpUS, a dataset of historical reservoir operations in the contiguous United States. *Sci. Data* 9 (1), 34. <https://doi.org/10.1038/s41597-022-0134-7>.
- Thurber, T., Vernon, C., Sun, N., Turner, S., Yoon, J., Voisin, N., 2021. mosartwmpy: a Python implementation of the MOSART-WM coupled hydrologic routing and water management model. *J. Open Source Software* 6 (PNNL-SA-161232).
- Tran, H., Fang, Y., Tan, Z., Zhou, T., Leung, L.R., 2024a. Quantifying the impacts of land-cover change on the hydrologic response to hurricane ida in the lower Mississippi River basin. *J. Hydrometeorol.* 25 (6), 899–914. <https://doi.org/10.1175/JHM-D-23-0094.1>.
- Tran, H., Leonarduzzi, E., De la Fuente, L., Hull, R.B., Bansal, V., Chennault, C., Gentile, P., Melchior, P., Condon, L.E., Maxwell, R.M., 2021. Development of a deep learning emulator for a distributed groundwater-surface water model: ParFlow-ML. *Water* 13 (23), 3393.
- Tran, H., Yang, C., Condon, L.E., Maxwell, R.M., 2023. The Budyko shape parameter as a descriptive index for streamflow loss [Brief Research Report]. *Front. Water* 5. <https://doi.org/10.3389/frwa.2023.1258367>.
- Tran, V.N., Ivanov, V.Y., Nguyen, G.T., Anh, T.N., Nguyen, P.H., Kim, D.-H., Kim, J., 2024b. A deep learning modeling framework with uncertainty quantification for inflow-outflow predictions for cascade reservoirs. *J. Hydrol.* 629, 130608.
- Tran, V.N., Kim, J., 2022. Robust and efficient uncertainty quantification for extreme events that deviate significantly from the training dataset using polynomial chaos-kriging. *J. Hydrol.* 609, 127716. <https://doi.org/10.1016/j.jhydrol.2022.127716>.
- Turner, S.W., Doering, K., Voisin, N., 2020a. Data-driven reservoir simulation in a large-scale hydrological and water resource model. *Water Resour. Res.* 56 (10) e2020WR027902.
- Turner, S.W., Xu, W., Voisin, N., 2020b. Inferred inflow forecast horizons guiding reservoir release decisions across the United States. *Hydrol. Earth Syst. Sci.* 24 (3), 1275–1291.
- Turner, S. W. D., Steyaert, J. C., Condon, L., & Voisin, N. (2021a). ISTARF-CONUS (0.0.1). <https://doi.org/10.5281/zenodo.4602277>.
- Turner, S.W.D., Steyaert, J.C., Condon, L., Voisin, N., 2021b. Water storage and release policies for all large reservoirs of conterminous United States. *J. Hydrol.* 603, 126843. <https://doi.org/10.1016/j.jhydrol.2021.126843>.
- Vinyals, O., Fortunato, M., Jaitly, N., 2015. Pointer networks. *Adv. Neural Informat. Process. Syst.* 28.
- Voisin, N., Li, H., Ward, D., Huang, M., Wigmosta, M., Leung, L.R., 2013. On an improved sub-regional water resources management representation for integration into earth system models. *Hydrol. Earth Syst. Sci.* 17 (9), 3605–3622. <https://doi.org/10.5194/hess-17-3605-2013>.
- Wang, J., Walter, B.A., Yao, F., Song, C., Ding, M., Maroof, A.S., Zhu, J., Fan, C., McAlister, J.M., Sikder, S., Sheng, Y., Allen, G.H., Crétaux, J.F., Wada, Y., 2022. GeoDAR: georeferenced global dams and reservoirs dataset for bridging attributes and geolocations. *Earth Syst. Sci. Data* 14 (4), 1869–1899. <https://doi.org/10.5194/essd-14-1869-2022>.
- 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 Resour. Res.* 44 (2).
- 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 Resour. Res.* 53 (4), 2786–2812.
- 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. *J. Hydrol.* 602, 126723.
- Yassin, F., Razavi, S., Elshamy, M., Davison, B., Saprizia-Azuri, G., Wheater, H., 2019. Representation and improved parameterization of reservoir operation in hydrological and land-surface models. *Hydrol. Earth Syst. Sci.* 23 (9), 3735–3764. <https://doi.org/10.5194/hess-23-3735-2019>.
- 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 Int.* 30 (4), 487–500.
- Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., Liu, X., Zhuang, J., 2018a. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *J. Hydrol.* 565, 720–736.
- Zhang, J., Zhu, Y., Zhang, X., Ye, M., Yang, J., 2018b. Developing a Long Short-Term memory (LSTM) based model for predicting water table depth in agricultural areas. *J. Hydrol.* 561, 918–929.
- Zhao, Q., Cai, X., 2020. Deriving representative reservoir operation rules using a hidden Markov-decision tree model. *Adv. Water Resour.* 146, 103753.
- Zhao, T., Cai, X., Yang, D., 2011. Effect of streamflow forecast uncertainty on real-time reservoir operation. *Adv. Water Resour.* 34 (4), 495–504.