# Leveraging Contextual Cues from a Conceptual Model with Predictive Skills of Machine Learning for Improved Predictability and Interpretability in the Hydrological Processes

**Pravin Vasudev Bhasme [1], Udit Bhatia [1]**

[1]Civil Engineering Discipline, Indian Institute of Technology Gandhinagar, Palaj, Gandhinagar, 382055, Gujarat, India

**Key Points:**

- Model to synergize machine learning with process understanding of conceptual model for hydrological processes.
- Variants are developed for lumped and semi-distributed scales as well as for managed and unmanaged catchments.
- Proposed model outperforms conceptual model; annual water balance and runoff coefficient analysis reveals physical consistency of model.

Corresponding author: Udit Bhatia, `bhatia.u@iitgn.ac.in`

## Abstract

In recent years, Machine Learning (ML) techniques have gained the attention of the hydrological community for their better predictive skills. Specifically, ML models are widely applied for streamflow predictions. However, limited interpretability in the ML models indicates space for improvement. Leveraging domain knowledge from conceptual models can aid in overcoming interpretability issues in ML models. Here, we have developed the Physics Informed Machine Learning (PIML) model at daily timestep, which accounts for memory in the hydrological processes and provides an interpretable model structure. We demonstrated three model cases, including lumped model and semi-distributed model structures with and without reservoir. We evaluate the first two model structures on three catchments in India, and the applicability of the third model structure is shown on the two United States catchments. Also, we compared the result of the PIML model with the conceptual model (SIMHYD), which is used as the parent model to derive contextual cues. Our results show that the PIML model outperforms simple ML model in target variable (streamflow) prediction and SIMHYD model in predicting target variable and intermediate variables (for example, evapotranspiration, reservoir storage) while being mindful of physical constraints. The water balance and runoff coefficient analysis reveals that the PIML model provides physically consistent outputs. The PIML modeling approach can make a conceptual model more modular such that it can be applied irrespective of the region for which it is developed. The successful application of PIML in different climatic as well as geographical regions shows its generalizability.


## 1 Introduction

The reservoir operation, water resources planning and management, flood prevention, and risk evaluation can be handled better with reliable streamflow predictions (Z. Liu et al., 2015). Thus, accurate streamflow forecasting aids decision-makers in addressing issues related to water supplies, flood mitigation, and hydro-power generation (Yaseen et al., 2016). To meet these objectives, hydrologists often rely on a suite of hydrological models of varying complexities (e.g., lumped, distributed, and semi-distributed), scales (regional to global) and architectures (including data-driven, conceptual, empirical and physical) (Devia et al., 2015). Conceptual models are computationally efficient while representing various dominant catchment dynamics in a physically meaningful way with less number of parameters (Fenicia et al., 2011). Their potential is explored for hypothesis testing (Vaze et al., 2010; Fenicia et al., 2022), semi-distributed modeling (Aronica & Cannarozzo, 2000; Ajami et al., 2004; Das et al., 2008), and they have been used to support operational forecasting (Feng et al., 2020). Some of the popular conceptual models which are applied widely in the field of hydrology include GR4J (Perrin et al., 2003), Xinanjiang (Ren-Jun, 1992), Sacramento Soil Moisture Accounting Model (SAC-SMA), and SIMHYD (Chiew et al., 2002). However, these conceptual models are developed for a specific region. Thus, the purported "uniqueness of place" is the cost of the apparent "simplicity" of conceptual models (Fenicia et al., 2011), which calls for cautious application of these models outside the given specific region.

The emerging paradigm of data-driven approaches, specifically Deep Learning (DL) methods, has shown remarkable success in improving hydrological predictions, including streamflow modeling at multiple timescales (Gauch et al., 2021), streamflow predictions in ungauged basins (Kratzert et al., 2019) hinting towards the existence of inter-basin consistency which can further aid in developing a watershed-scale theory for the rainfall-runoff process (Nearing et al., 2021). Shen (2018) has provided a transdisciplinary review of DL applications and suggests that the DL has the potential to improve water science. However, the studies have applied a data-driven approach for the streamflow prediction with different inputs while ignoring the intermediate processes, and physical consistency checks (Parisouj et al., 2020; Thapa et al., 2020; Wu et al., 2022; Khosravi et al., 2022). Also, when constrained by the physics of processes, data-driven models often run into issues of equifinality and produce spurious insights (Bhasme et al., 2022; Reichstein et al., 2019). Thus, interpretability and physical consistency are the challenges associated with the application of purely data-driven models.

A recent perspective in Nature argued that synergistically combining physics with machine learning could be a promising way to address the limitations associated with the individual models (Reichstein et al., 2019). Thus, the aforementioned issues of interpretability, physical consistency, and generalizability can possibly be resolved by combining interpretability from the conceptual model and predictive skills of the data science approach using the Machine Learning (ML) model in a systematic way. Recently researchers have made numerous attempts at the synergistic application of ML and physics-based or conceptual models in hydrology. Karpatne et al. (2017) have discussed different approaches to combining domain knowledge with predictive skills of data-driven models under the umbrella of "Theory Guided Data Science." Willard et al. (2022) have classified the integration of physical principles with machine learning into four classes: physics-guided loss function; physics-guided initialization; physics-guided design of architecture; and hybrid modeling. One of the ways of hybrid modeling is to use the output of physics-based models as input for ML models. Zhou et al. (2022) has proposed an integrated model which combines the Xinanjiang conceptual model with the Monotone Composite Quantile Regression Neural Network (MCQRNN) for forecasting flood probability density where they fed the output of Xinanjiang model for forecasted steps, observed streamflow and rainfall at past steps to the MCQRNN model. Merely considering the streamflow in the forecasted inputs makes the model sensitive to the performance of the physics-based model. Also, their model has limited interpretability and ignores the physical consistency of various processes, as it doesn't account for intermediate processes. Parisouj et al. (2022) have developed a physics-informed data-driven model for 1-day ahead streamflow forecasting by applying ML with inputs as precipitation and observed streamflow at current and previous timestep with 1-day ahead forecasted streamflow from Hydrologic Engineering Center - Hydrologic Modeling System (HEC-HMS) model. However, ignoring intermediate processes in their study affects the interpretability of the model. Lu et al. (2021) has developed a physics-informed hybrid Long Short-Term Memory (LSTM) by using outputs of a physics-based model along with meteorological variables as inputs to the LSTM and improved the out-of-distribution (input data have very dry or very wet years for training period) streamflow predictions. However, their model structure does not consider any intermediate variable, which limits the interpretability of the model. K. Li et al. (2022) has demonstrated a physics-informed data-driven model for understanding the factors responsible for the baseflow, interflow, and overflow dynamics among the different variables such as precipitation, air temperature, and irrigation. However, their study excludes soil moisture which may have crucial information about baseflow processes. Jia et al. (2021) has developed a physics-guided recurrent graph model to predict the streamflow and temperature in the river network. They have used a pre-training technique that transfers the knowledge in the physics-based model to the ML model and also proposed a loss function that accounts for the river segments to balance the performance over it. However, their model does not account for physical consistency checks. B. Liu et al. (2022) has developed a hybrid physics-data methodology for streamflow and flood simulation by processing the output of a process-based model with meteorological forcings using LSTM. However, their study ignores intermediate processes, which limits the interpretability of the model.

One way to incorporate domain knowledge and include intermediate variables is to consider a conceptual or physics-based model structure with given inputs and intermediate variables, then employ ML algorithms to extract complex relationships between the variables involved in the processes (Willard et al., 2022). On a similar line, Khandelwal et al. (2020) have proposed a Physics Informed Machine Learning (PIML) for predicting daily streamflow, which follows a similar conceptual structure to the Soil and Water Assessment Tool (SWAT). However, their study ignores physical constraints required at various stages; for example, actual evapotranspiration should be less than or equal to potential evapotranspiration. For streamflow prediction, researchers (Bhasme et al., 2022) have developed a lumped PIML model for monthly streamflow predictions and demonstrated how PIML architectures result in significant performance gains in predicting target (streamflow) and intermediate (evapotranspiration) while ensuring physical consistency (mass balance) for basin scale hydrological processes. However, the coarse spatial scale and monthly temporal resolution limit the generalization of work to various water resource planning and management applications. We note that the scale issue in hydrology is identified as one of the 23 unsolved problems in hydrology (Blöschl et al., 2019) where authors discuss the scale variance of hydrologic laws at the catchment scale. Thus, translating a lumped model to a semi-distributed scale is a non-trivial task, given the processes' non-linearity.

To address these multifaceted challenges, we propose an approach of partitioning the conceptual model into different process components, then modeling each process separately using the ML models, and finally combining all the processes together while applying checks at various stages and ensuring the physical consistency in the overall model outputs. For example, in the case of a semi-distributed model, we partition the SIMHYD model into evapotranspiration and streamflow process components for each of the subcatchments within the catchment. Then we model evapotranspiration separately using the ML model for each subcatchment, and obtained output is fed to the streamflow modeling component. While with the predictive power of ML, both upstream and downstream parts streamflow is modeled together as the upstream part streamflow contributes downstream part streamflow. However, the past timesteps of inputs are informed by the DELAY parameter of the Muskingum routing method, as an understanding of temporal lag in the catchment response may help better predictability at higher temporal scales. Further, we combine these outputs and check for water balance. In this way, the PIML approach makes the conceptual model more generalizable while providing better predictive skills.

Synergizing the conceptual model with ML while ensuring the conservation of mass and physical consistency opens the way to better process representation. In this study, we used SIMHYD conceptual model structure to build PIML, and then its lumped and semi-distributed variants are applied in the three unmanaged catchments of peninsular India, while the semi-distributed variant with reservoir is applied in the two managed catchments (reservoirs in the catchments) of the United States. We modeled actual evapotranspiration (ET) and streamflow (Q) at daily timesteps for both upstream and downstream parts in a semi-distributed structure while considering spatial heterogeneity in the model inputs. In the case of managed catchments, we also modeled reservoir storage and release. Though our proposed PIML model provides the choice of ML models, we used LSTM as the ML model for this study. The rest of the paper is organized as follows: Section 2 gives details of the study area and data used in this study, followed by methods, including conceptual model cases and proposed PIML model cases. Section 4 briefs about different model setups based on the model case. The results are discussed in Section 5. Further, Section 6 gives a conclusion of this work.

## 2 Study area and data used

In this study, we have developed three PIML model structures: lumped, semi-distributed without reservoir, and semi-distributed with reservoir. We have assessed the applicability of the proposed lumped model to three catchments in peninsular India (Figure 1 (a)), where each catchment belongs to the Baitarni, Krishna, and Mahanadi river basins. The details of the study area with respective training and testing periods are given in Table 1. The required precipitation dataset is obtained from India Meteorological Department (IMD) (`https://www.imdpune.gov.in/`). Actual and potential evapotranspiration datasets are obtained from the latest version of (v3.6a) of Global Land Evaporation Amsterdam

Model (GLEAM) (`https://www.gleam.eu/`) datasets (Martens et al., 2017; Miralles et al., 2011). While using the GLEAM dataset, we ensured that the sum of average annual actual evapotranspiration and streamflow is less than the average annual precipitation for the SIMHYD model calibration and validation period. The precipitation, actual, and potential evapotranspiration datasets are obtained at daily timestep with a spatial resolution of $0.25^o \times 0.25^o$. The precipitation is aggregated with the Thiessen polygon method to lumped scale, while actual and potential evapotranspiration are aggregated through averaging. The streamflow datasets for Anandpur, Kantamal, and Keesara hydrological observation stations are obtained from India Water Resources Information System (India-WRIS; `https://indiawris.gov.in/wris/`) portal.

The semi-distributed model without a reservoir is also demonstrated on the three catchments used in the lumped modeling case. To divide the catchment into two parts, we considered hydrological observation stations in the upstream part of these catchments. Champua, Kesinga, and Madhira are the three upstream hydrological observation stations in the Anandpur, Kantamal, and Keesara catchments (Figure 1 (a)), respectively (Table 1). The streamflow data for these stations is obtained from India-WRIS. The actual and potential evapotranspiration are sourced from the GLEAM dataset, while precipitation data is obtained from IMD. Similar to the lumped case, we ascertained that for upstream part of the catchment has the sum of average annual actual evapotranspiration, and streamflow is less than the average annual precipitation for the SIMHYD model calibration and validation period. In the results and discussion section, these catchments are referred to based on the name of downstream hydrological observation station (for example, Anandpur catchment).

The application of semi-distributed model with reservoir is demonstrated on two catchments of the United States (Figure 1 (b)). These catchments have a single reservoir in its upstream. The selection of catchments is based on the percentage of snow water equivalent in the precipitation. Since the SIMHYD model does not consider snow in the model, we select catchments having less than two percent of snow water equivalent in the precipitation throughout the modeling period, including the warmup period. The two selected reservoirs, Brady Creek reservoir and Canyon lake, belong to Colorado and Guadalupe river basins (Figure 1 (b)), respectively (Table 1). The reservoir release data is obtained from United States Geological Survey (USGS) (`https://waterdata.usgs.gov/nwis`) for sites USGS 08145000 and USGS 08167800 for Brady Creek reservoir and Canyon lake, respectively and consideration of these stations for release data is consistent with ResOpsUS (Steyaert et al., 2022), a recently developed inventory of observed reservoir operations for conterminous United States (CONUS). Hereafter the catchments with the reservoir are referred to based on the name of the reservoir: Brady catchment and Canyon catchment. While downstream gauge stations selected are USGS 08146000 and USGS 08168500 for Brady and Canyon catchment, respectively. The reservoir storage data is obtained from Texas Water Development Board (`https://www.waterdatafortexas.org/reservoirs/statewide`). The actual and potential evapotranspiration is obtained from the GLEAM dataset. The daily precipitation data at 1 km resolution for US catchments is sourced from Daymet (Daily Surface Weather Data on a 1-km Grid for North America, Version 4 R1) (Thornton et al., 2022).

We used thirteen years of data for calibration and six years of data for validation of SIMHYD model, while additional three years of data is required as a warmup period (Table 1). Similarly, for ML and PIML models, training and testing period datasets are of thirteen and six years, respectively.

## 3 Methods

To demonstrate the proposed PIML model, we use state of the art conceptual model (SIMHYD in this case), ML (LSTM in this case) model and combination thereof. The

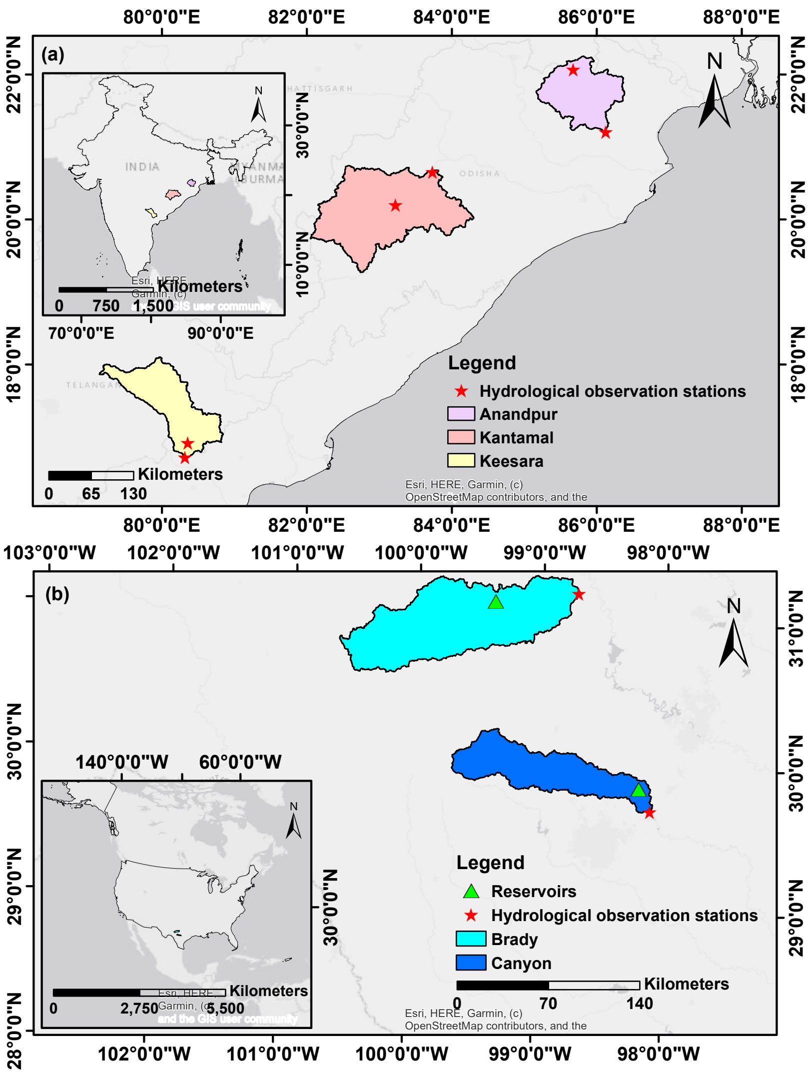


**Figure 1.** Location of study area. (a) Catchments used to demonstrate lumped and semi-distributed without reservoir modeling cases; (b) Catchments used to demonstrate the semi-distributed with reservoir modeling case.

SIMHYD model (Figure 2(a)) is lumped conceptual hydrological model that works at daily time-step (Chiew et al., 2002). It is widely applied for various hydrological studies, including hypothesis testing (Vaze et al., 2010), understanding impact of land-use change on catchment hydrology (Siriwardena et al., 2006), assessing climate change impact on runoff (Mpelasoka & Chiew, 2009; Chiew et al., 2010), runoff predictions in ungauged catchments (F. Li et al., 2014), and analyzing grid-based regionalization in data-sparse region (H. Li & Zhang, 2017). We applied the SIMHYD model at both lumped and semi-distributed scales. For the lumped modeling total nine parameters are used to calibrate the model against the observed ET and Q (See Text S1 in Supplementary Information (SI) for the SIMHYD model details and equations). While for the semi-distributed modeling, we made two cases: semi-distributed SIMHYD without reservoir, and semi-distributed SIMHYD with a reservoir which are discussed as follows:

**Table 1.** Study area details with respective training, testing periods, and DELAY parameter obtained in the SIMHYD model calibration. The model structures includes: (a) Lumped model; (b) Semi-distributed model without reservoir; (c) Semi-distributed model with reservoir.

| Catchment | Subcatchment | Area (sq.km) | training period* | testing period | DELAY (days) |
|---|---|---|---|---|---|
| (a) Lumped model | | | | | |
| Anandpur | - | 8671.27 | 1999 - 2011 | 2012 - 2017 | 1.24 |
| Kantamal | - | 20236.07 | 2000 - 2012 | 2013 - 2018 | 1.63 |
| Keesara | - | 10220.27 | 1998 - 2010 | 2011 - 2016 | 1.76 |
| (b) Semi-distributed model without reservoir | | | | | |
| Anandpur | Champua - Anandpur | 6849.89 | 1999 - 2011 | 2012 - 2017 | 0.87 |
| | Champua | 1821.38 | 1999 - 2011 | 2012 - 2017 | 0.50 |
| Kantamal | Kesinga - Kantamal | 8401.21 | 2000 - 2012 | 2013 - 2018 | 0.62 |
| | Kesinga | 11834.86 | 2000 - 2012 | 2013 - 2018 | 1.18 |
| Keesara | Madhira - Keesara | 8456.94 | 1998 - 2010 | 2011 - 2016 | 1 |
| | Madhira | 1763.33 | 1998 - 2010 | 2011 - 2016 | 0.54 |
| (c) Semi-distributed model with reservoir | | | | | |
| Brady | d/s of Brady reservoir | 6531.48 | 2003 - 2015 | 2016 - 2021 | 1.29 |
| | Brady reservoir | 1353.30 | 2003 - 2015 | 2016 - 2021 | 0.23 |
| Canyon | d/s of Canyon lake | 266.05 | 2003 - 2015 | 2016 - 2021 | 0.04 |
| | Canyon lake | 3713.27 | 2003 - 2015 | 2016 - 2021 | 1.43 |

*Additional three years of data is used as a warmup period for calibration of SIMHYD model cases.

### 3.1 Semi-distributed SIMHYD without reservoir

Researchers have tested conceptual models to the semi-distributed modeling (Aronica & Cannarozzo, 2000; Ajami et al., 2004; Das et al., 2008) with different calibration strategies, including lumped, semi-lumped and semi-distributed. In the case of lumped calibration strategy, the model inputs are provided in aggregated format with single time series for a given variable while keeping the same parameter for all the subcatchments. However, in the semi-lumped calibration strategy, the model inputs are provided separately for each subcatchment, while parameters are kept the same for all the subcatchments. The semi-distributed calibration strategy shows that inputs and parameters are spatially varied for all the subcatchments involved. Ajami et al. (2004) reported that the semi-lumped strategy outperformed other strategies in their study. F. Li et al. (2013) has calculated grid-wise runoff using the SIMHYD model. We have experimented with distributed parameters, and calculated average Nash Sutcliffe Efficiency (NSE) in the calibration period for the evapotranspiration and streamflow at both upstream and downstream parts of the catchment as 0.52 which is lesser than 0.63 for the model with the same parameters for all subcatchments. Thus, we used the same model parameters for the upstream and downstream parts of the catchment while having different inputs for the subcatchments. This model case requires two additional parameters for routing the runoff from upstream part of the catchment. However, the routing parameters are different for both subcatchments as they provide temporal lag in the catchment response, further assisting in the PIML model. The model is calibrated with target variables, including evapotranspiration at upstream ($ETu/s_t$) and downstream ($ETd/s_t$) part of the catchment, streamflow at upstream ($Qu/s_t$) and downstream ($Qd/s_t$) hydrological observation stations. The $ETu/s_t$, $ETd/s_t$, and $Qu/s_t$ are considered the target variables as these variables are later used to test the physical consistency in the PIML model.

### 3.2 Semi-distributed SIMHYD with reservoir

Reservoirs have significant effect on the flow regime characteristics and thus influences the ecological processes (Ekka et al., 2022). Hence, it is imperative to include reservoirs in modeling managed catchment. We considered two catchments with reservoirs to demonstrate the semi-distributed SIMHYD with a reservoir. Similar to the previous case of semi-distributed SIMHYD without a reservoir, the catchment is divided into two parts in which the upstream part is considered up to the reservoir location, and the downstream part is considered between the reservoir and downstream hydrological observation station. Recently, Turner et al. (2021) has developed weekly reservoir operation policies for all large reservoirs of CONUS and suggested that these policies may be applied to the daily time step. However, converting weekly reservoir release values to daily values may not be able to capture the variations observed at the daily time step. Since employing the best reservoir operation technique is outside the scope of this study, we used a generic reservoir routing model for release estimation. Gutenson et al. (2020) has compared two reservoir routing methods, including the method by Hanasaki et al. (2006) and Döll et al. (2003) applied on United States Army Corps of Engineers (USACE) operated 60 reservoirs for daily timesteps and found that later one is outperforming former. Thus, we used the empirical equation (Eq. 1) given by Döll et al. (2003) for the estimation of release. The semi-distributed SIMHYD with reservoir requires one additional parameter than without reservoir case attributed to reservoir release. While using the empirical release equation, mass conservation is also ensured by Eq. 2. Since reservoir inflow data is not available for both of the reservoirs, the model is calibrated with target variables including $ETu/s_t$ and $ETd/s_t$, $Qd/s_t$, reservoir live storage ($S_t$) and release ($R_t$).

$$R_t = k_r * S_t * \left(\frac{S_t}{S_{max}}\right)^{1.5} \tag{1}$$

where, the $k_r$ is outflow coefficient and $S_{max}$ is the maximum live storage capacity.

$$S_t + R_t = S_{t-1} + Qin_t \tag{2}$$

where, $Qin_t$ is the reservoir inflow.

### 3.3 Physics informed machine learning model

The PIML takes advantage of the contextual cues from the SIMHYD model. The choice of predictors and predictands are based on governing equations of the SIMHYD model. In the PIML (Physics Informed Machine Learning) model, the "physics informed" is attributed to model structure, imposing physical constraints wherever required and possible, choice of predictors and predictands, while "machine learning" is for extracting complex relationships between the predictors and predictands. The complexity of temporal dynamics in the catchment response increases with the temporal resolution of the model. The hydrological processes aggregated at lower temporal resolution may not capture the variations in various fluxes at the higher resolution important for a flood. However, understanding temporal lag in the catchment response may help in better predictability at a higher temporal scale. In this study, we have considered a delay in the catchment response with the help of a routing mechanism through the application of the Muskingum routing method. The DELAY parameter in the Muskingum routing method shows the time taken by flow in traveling river reach (O'Sullivan et al., 2012) (Refer Text S1 for Muskingum method equations and details). We have demonstrated three versions of PIML based on spatial scale and the mode of operation in the catchment. The different spatial scale includes lumped and semi-distributed scales, while the mode of operation considers managed and unmanaged catchment based on the reservoir availability in the upstream part of the catchment.

The proposed PIML model is flexible for choice of ML models, however in this study we used Long Short Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997), a recurrent neural network based architecture known for its ability to learn long-term information. It has been applied in various hydrological studies, including post-processing of physics-based model outputs (Frame et al., 2021), prediction of extreme events (Frame et al., 2022), leverage synergy when multiple datasets are used for given variable (Kratzert et al., 2021), flood forecasting (Nevo et al., 2022; Feng et al., 2020), improvement in the streamflow predictions of ungauged basins (Kratzert et al., 2019), streamflow prediction for multiple timescales (Gauch et al., 2021). Refer Text S2 and Figure S1 (in SI) for the LSTM model details and equations. We briefly discuss the PIML versions as follows:

### 3.3.1 Lumped PIML

The proposed PIML version of lumped scale (Figure 2(b)) combines process understanding from the conceptual model with the ability of ML models to extract the complex relationship between predictors and predictands. Here we used actual evapotranspiration ($ET_t$) as an intermediate variable to introduce interpretability in the model. However, to incorporate physical constraint, we predict a ratio of $ET_t$ with potential evapotranspiration ($PET_t$) as this ratio will not exceed one, and it is easy to apply this constraint using sigmoid activation function in the LSTM model structure. The output of the sigmoid activation function has a range of [0, 1]. The ratio of $ET_t$ with $PET_t$ is the function of precipitation ($P_t$), $PET_t$ and soil moisture at previous timestep ($SMS_{t-1}$) (Eq. 3). The streamflow ($Q_t$) is the function of $ET_t$, precipitation, soil moisture, groundwater storage (Eq. 4). The exact form of $a$ (Eq. 3) and $b$ (Eq. 4) is determined by ML model. However, a number of past timesteps (of predictors) which we referred as memory in the hydrological processes, are decided based on the DELAY parameter in the Muskingum routing. This DELAY parameter is evaluated in the SIMHYD model since we used the Muskingum routing method for streamflow routing. As the PIML model is developed for daily timestep, we approximated DELAY to the greater integer in case of a float value. For example, when DELAY (Table 1) is 1.24, then it is approximated as 2 ($j$ in Eq. 4). This approximation is useful since our model works at daily timestep, essentially integer. The proposed PIML model consists of two layers of LSTM models. The first layer output is multiplied with respective $PET_t$ to get $ET_t$ which is later fed to the second layer LSTM model along with other predictors to predict $Q_t$.

$$\frac{ET_t}{PET_t} = a(P_t, PET_t, SMS_{t-1}) \tag{3}$$

$$Q_t = b(P_t, ET_t, SMS_t, GW_t, ..., P_{t-j}, ET_{t-j}, SMS_{t-j-1}, GW_{t-j-1}) \tag{4}$$

### 3.3.2 Semi-distributed PIML without reservoir

The semi-distributed PIML without reservoir (Figure 2(c)) is the extended version of the lumped PIML while considering the spatial heterogeneity in the model inputs and intermediate processes such as evapotranspiration. Here we considered a simple case for semi-distributed modeling by distributing the catchment into two different subcatchments based on the location of the hydrological observation stations. The required input of spatial soil moisture and groundwater storage is obtained from the semi-distributed SIMHYD model. Similar to lumped PIML, we are predicting a ratio of $ET_t$ with $PET_t$ for both the upstream and downstream parts of catchments, which is further used as one of the inputs for the streamflow generation. The ratio of $ET_t$ with $PET_t$ is the function of $P_t$, $PET_t$ and $SMS_{t-1}$ in the respective upstream (Eq. 5) and downstream (Eq. 6) part of the catchment. Later, streamflow at the outlet of both upstream ($Qu/s_t$) and downstream ($Qd/s_t$) part of the catchment are predicted together by introducing physical loss. This physical loss (Eq. 8) is based on the physical constraint over the annual contribution of the upstream part streamflow at the downstream outlet, which should be always less than or equal to the annual downstream streamflow. The deployment of the loss function is such that whenever the annual streamflow contribution constraint is violated, the penalty

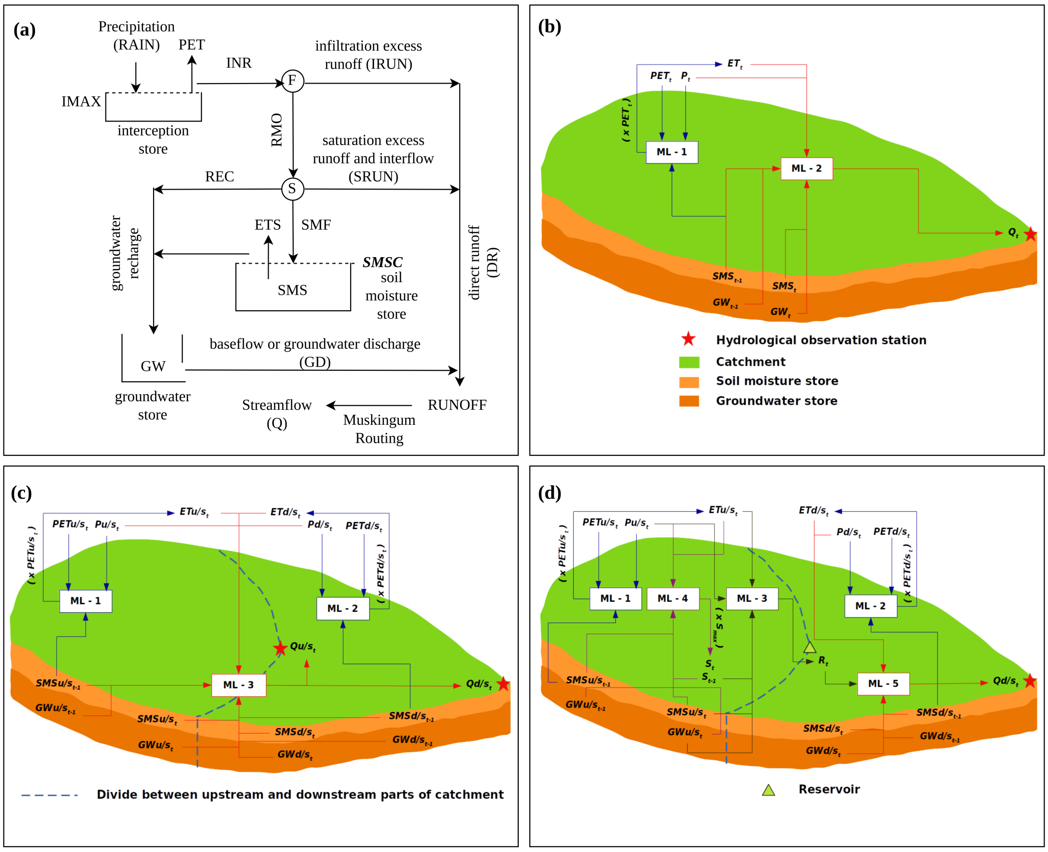


**Figure 2.** Different model architectures used in this study: (a) SIMHYD model structure. The IMAX, PET, INR, RMO, REC, ETS, SMF, and SMSC are the maximum interception, potential evapotranspiration, runoff after an interception, remaining moisture, recharge to groundwater store, soil evapotranspiration, part of RMO going into soil moisture store, and soil moisture store capacity, respectively; (b) Lumped PIML structure for no delay in catchment response (DELAY = 0). Blue arrows show evapotranspiration ($ET_t$) prediction using Machine Learning algorithm - 1 (ML - 1), while red arrows display streamflow ($Q_t$) prediction with the help of ML - 2; (c) Structure of semi-distributed PIML without reservoir model for 0 delays (DELAY = 0) in both subcatchments. The blue arrows show evapotranspiration predictions in both subcatchments using ML - 1 and ML - 2 for upstream ($ETu/s_t$) and downstream ($ETd/s_t$) parts of the catchment, respectively. The red arrows depict the combined prediction of streamflow at both upstream ($Qu/s_t$) and downstream ($Qd/s_t$) hydrological observation stations with the help of ML-3; (d) Structure of semi-distributed PIML with reservoir model for 0 delays (DELAY = 0) in both subcatchments. Similar to semi-distributed PIML without a reservoir model, blue arrows show evapotranspiration predictions in both subcatchments using ML - 1 and ML - 2 for upstream ($ETu/s_t$) and downstream ($ETd/s_t$) parts of the catchment respectively. The dark green arrows exhibit the prediction of reservoir release ($R_t$) with ML - 3, while the purple arrow conveys the reservoir storage ($S_t$) predictions using ML - 4. The red arrows show the streamflow prediction at the downstream hydrological observation station ($Qd/s_t$) with the help of ML - 5.

is applied in the loss function. The $Qu/s_t$ and $Qd/s_t$ are the function of $ET_t$, precipitation, soil moisture, groundwater storage at upstream and downstream parts (Eq. 7) with past timesteps informed by Muskingum DELAY parameter ($l$ and $m$ in Eq. 7)).

The exact functional forms of $c$, $d$, and $e$ are determined by ML model. The semi-distributed PIML without reservoir consists of three layers of LSTM (Figure 2(c)), of which two layers will provide $ET_t$ on multiplication of its outputs with respective $PET_t$ values for an upstream and downstream part in each of the layers. Later, this obtained $ET_t$ would be fed to the third layer of LSTM with other variables such as precipitation, soil moisture, and groundwater storages.

$$\frac{ETu/s_t}{PETu/s_t} = c(Pu/s_t, PETu/s_t, SMSu/s_{t-1}) \tag{5}$$

$$\frac{ETd/s_t}{PETd/s_t} = d(Pd/s_t, PETd/s_t, SMSd/s_{t-1}) \tag{6}$$

$$\begin{split} Qu/s_t, Qd/s_t = e(Pu/s_t, ETu/s_t, SMSu/s_t, GWu/s_t, ..., Pu/s_{t-l}, ETu/s_{t-l}, \\ SMSu/s_{t-l-1}, GWu/s_{t-l-1}, Pd/s_t, ETd/s_t, SMSd/s_t, GWd/s_t, ..., Pd/s_{t-m}, ETd/s_{t-m}, \\ SMSd/s_{t-m-1}, GWd/s_{t-m-1}) \end{split} \tag{7}$$

$$loss = \begin{cases} \lambda * \left( \frac{Qu/s_{pred} * A_{ratio}}{Qd/s_{pred}} - 1 \right) + MSE(Q_{pred}, Q_{act}) & \text{if } Qu/s_{pred} * A_{ratio} > Qd/s_{pred} \\ MSE(Q_{pred}, Q_{act}) & \text{otherwise} \end{cases} \tag{8}$$

Where $\lambda$ is the penalty and $A_{ratio}$ is the area ratio of upstream subcatchment and total catchment.

#### *3.3.3 Semi-distributed PIML with reservoir*

We demonstrated semi-distributed PIML with a reservoir (Figure 2(d)) using a simple case where the catchment is divided into parts based on the location of the reservoir and hydrological observation station. The model includes predictions of ratio of $ET_t$ with $PET_t$ at upstream and downstream parts, $Qd/s_t$, reservoir storage ($S_t$) and release ($R_t$). Similar to the earlier case of semi-distributed PIML without reservoir, the ratio of $ET_t$ with $PET_t$ is the function of $P_t$, $PET_t$ and $SMS_{t-1}$ in the respective upstream (Eq. 5) and downstream (Eq. 6) part of the catchment. In the absence of reservoir water demand data, the $S_t$ and $R_t$ are dependent on reservoir inflow, reservoir storage at the previous time step ($S_{t-1}$) based on the continuity equation for the reservoir (Eq. 2). Since the observed inflow is not available at both reservoirs, we used a similar approach as of lumped PIML. Thus the reservoir inflow can be presented in the form of its predictors (For example, $Pu/s_t$, $ETu/s_t$, $SMSu/s_t$, $SMSu/s_{t-1}$, $GWu/s_t$, $GWu/s_{t-1}$ are the predictors of reservoir inflow for 0 DELAY). In the case of $S_t$ prediction, physical constraint similar to $ET_t$ is used. We predict a ratio of $S_t$ with maximum live reservoir storage capacity ($S_{max}$) as this ratio will always be less than or equal to one. The $R_t$ and ratio of $S_t$ with $S_{max}$ (Eq. 9) are the function of $ET_t$, precipitation, soil moisture, groundwater storage at upstream part, and $S_{t-1}$. Further the obtained $ETd/s_t$ and $R_t$ along with precipitation, soil moisture, groundwater storage at downstream part are used predict $Qd/s_t$ (Eq. 10). The semi-distributed PIML with reservoir consists of five layers of LSTM, of which two layers will provides $ET_t$ for an upstream and downstream part in each of the layers on processing with respective $PET_t$. Later the $ETu/s_t$ will be fed to the third and fourth layer of LSTM with other variables such as precipitation, soil moisture, groundwater storages at current and past time steps based on the Muskingum DELAY parameter obtained in the SIMHYD model ($p$ in Eq. 9) and reservoir storage at previous timestep to predict $R_t$ and ratio of $S_t$ with $S_{max}$ in the respective layers. The final $S_t$ values are obtained by multiplying output of the fourth layer in PIML model with $S_{max}$. Further, the predicted $R_t$ is fed the fifth LSTM layer with other variables such as precipitation, $ETd/s_t$ (from the second layer), soil moisture, and groundwater storages at current and past time steps based on the Muskingum DELAY parameter obtained in the SIMHYD model for downstream part ($q$ in Eq. 10). The exact functional form of $f$ (Eq. 9) and $g$ (Eq. 10) can be identified by ML model (LSTM for this study).

$$R_t, \frac{S_t}{S_{max}} = f(Pu/s_t, ETu/s_t, SMSu/s_t, GWu/s_t, ..., Pu/s_{t-p}, ETu/s_{t-p}, SMSu/s_{t-p-1}, GWu/s_{t-p-1}, S_{t-1}) \tag{9}$$

$$Qd/s_t = g(Pd/s_t, ETd/s_t, SMSd/s_t, GWd/s_t, ..., Pd/s_{t-q}, ETd/s_{t-q}, SMSd/s_{t-q-1}, GWd/s_{t-q-1}, R_t) \tag{10}$$

The model performance is evaluated with Nash-Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970), Root Mean Square Error (RMSE), and Percent Bias (PBIAS) widely applied in the field of hydrology (Swain & Patra, 2017; Paul et al., 2019; Wagena et al., 2020). The details of these metrics can be referred from Text S3 in SI.

## 4 Model setups

### 4.1 Conceptual model setup

We used Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al., 2002) applied for multi-objective optimization in the various hydrological studies (Shin et al., 2015; Fowler et al., 2016; Mostafaie et al., 2018) for calibration of different cases in the SIMHYD model. This study applies NSGA-II with population size and maximum generation numbers of 100 and 500, respectively for all SIMHYD model cases. Same objective function is used for calibration of all cases in SIMHYD model and it is given by Eq. (11). The best parameters are selected based on the average SIMHYD model's performance in predicting intermediate variables and streamflow. The lumped SIMHYD model is calibrated against observed $ET_t$ and $Q_t$. It involves nine parameters such as INSC (interception store capacity), COEFF (maximum infiltration loss), SQ (infiltration loss exponent), SMSC (soil moisture store capacity), SUB (constant of proportionality in interflow equation), CRAK (constant of proportionality in groundwater recharge equation), K (baseflow linear recession parameter), DELAY (delay parameter in Muskingum routing (days)), x (storage weight parameter in Muskingum routing) having range of [0.5, 5], [50, 400], [0, 6], [50, 500], [0, 1], [0, 1], [0.003, 0.3], [0.5, 10] and, [0, 0.5], respectively. The best parameters obtained in the calibration process of lumped SIMHYD model are listed in the Table S1 in SI.

The semi-distributed without reservoir SIMHYD model is calibrated against observed $ETu/s_t$, $Qu/s_t$, $ETd/s_t$, $Qd/s_t$. We used the same model parameters for both the upstream and downstream parts, similar to the lumped model. However, additional routing parameters are employed for the upstream part, which results in a total of 11 parameters for without a reservoir case. Table S2 shows the best parameters obtained in the calibration process for the semi-distributed without reservoir SIMHYD model. In the case of the semi-distributed with reservoir SIMHYD model, due to the absence of reservoir inflow data, we used $ETu/s_t$, $R_t$, $S_t$, $ETd/s_t$, and $Qd/s_t$ for model calibration. In addition to the without reservoir case, a model with reservoir requires one more parameter ($k_r$) attributed to the empirical formula (Eq. 1) used for reservoir release, and its value ranges from [0.01, 0.9]. Thus, the semi-distributed with reservoir SIMHYD model has a total of 12 parameters. Since catchment size is comparatively small in the case of upstream and downstream parts of Brady and Canyon catchments, respectively, we restricted DELAY parameter for calibration process in upstream and downstream parts of Brady catchment as [0, 0.5] and [1, 1.5], respectively based on the travel time mentioned in David et al. (2011) while for the Canyon catchment it is restricted to [1, 1.5] and [0, 0.25] for upstream and downstream parts respectively. The best parameters obtained in the calibration process for without reservoir case are provided in the Table S3. The SIMHYD model output for the training and testing periods are generated using the best parameters obtained in the calibration process as the soil moisture store, and groundwater storage variables are further used as inputs in the PIML model.

$$Objective = 1 - NSE \tag{11}$$

## 4.2 PIML model setup

The proposed PIML models have the capability to use different ML models in the model structure. In this study, we have demonstrated it with LSTM as an ML model, applied using Tensorflow (Abadi et al., 2015). The lumped PIML model constitutes two layers of LSTM models (Figure 2(b)). In this case, both the LSTM models are trained and tested separately and sequentially. Both LSTMs have a single dense layer. We used the 'mean square error' loss function and 'Adam' optimizer for both models. The first layer predicts the ratio of $ET_t$ with $PET_t$, which uses the sigmoid activation function to avoid violation of known physical constraint over the ratio of $ET_t$ with $PET_t$. We have preprocessed input data with MinMax Scaler while the target variable lies between 0 to 1, due to which the target variable is not preprocessed. This selective preprocessing will help in executing the physical constraint. This same approach is applied in all PIML cases for evapotranspiration prediction. However, in the case of streamflow prediction, we have not preprocessed input as it is observed that preprocessing of data is not improving the model predictions. The second layer of lumped PIML is fed with the processed output ($ET_t$) of the first layer, precipitation, soil moisture store, and groundwater store (both obtained from the SIMHYD model) at current and past time steps based on the Muskingum DELAY parameter. The ReLU activation function is employed to have meaningful (non-negative) streamflow predictions. The LSTM model is tuned by applying different sets of hyperparameters, including dropout rate (0.1, 0.2, 0.3, 0.4), units (10, 20, 30, 40, 50, 60, 70, 80, 90, 100) and, epochs (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000). The different batch sizes (32, 64, 128, 256, 360) are also tried. Table S4 shows hyperparameters applied in the lumped PIML model. Similar sets of hyperparameters are also applied for LSTM as a simple ML model for the prediction of streamflow using precipitation and potential evapotranspiration, which are also inputs for the SIMHYD model. The final hyperparameters used in the ML modeling are listed in Table S5.

The semi-distributed PIML without a reservoir includes three layers of the LSTM model (Figure 2(c)). The first layer predicts ratio of $ETu/s_t$ with $PETu/s_t$ while the second layer predicts ratio of $ETd/s_t$ with $PETd/s_t$ using respective precipitation and potential evapotranspiration at the current timestep and soil moisture store at the previous timestep obtained from SIMHYD model output. The later processed output ($ETu/s_t$ and $ETd/s_t$) of these two layers are supplied to the third LSTM model with respective precipitation, soil moisture store, and groundwater store at the current and previous timesteps based on the DELAY parameter of Muskingum routing. The third LSTM model is used to predict both upstream ($Qu/s_t$) and downstream ($Qd/s_t$) streamflow while having two dense layers. We incorporated physical constraint through a custom loss function. This loss function ensures that annual contribution of the upstream part streamflow at the downstream outlet is always less than or equal to the annual downstream streamflow and it can be achieved using a batch size of 360 (close to 365 days in a year) which means that 360 samples are processed before model updation. Similar to streamflow prediction in the lumped PIML, the same sets of dropout rate, units, and model settings, such as optimizer and activation function, are used for both upstream ($Qu/s_t$) and downstream ($Qd/s_t$) streamflow prediction in the semi-distributed PIML without reservoir. The final hyperparameters used for semi-distributed PIML without a reservoir model are listed in the Table S6.

The third case of PIML is the semi-distributed PIML with a reservoir. It involves five layers of the LSTM models (Figure 2(d)). Similar to semi-distributed PIML without a reservoir model, these model predicts ratio of $ET_t$ with $PET_t$ in respective subcatchments in the first two layers. Later, $ETu/s_t$ is fed to the third and the fourth layer of LSTM with other inputs such as upstream part precipitation, soil moisture store, and groundwater store at current timestep and previous timesteps based on the DELAY parameter of Muskingum routing in the upstream part of the catchment, and reservoir storage at previous timestep. The third layer predicts reservoir release ($R_t$) using similar model settings (activation function, optimizer, loss function), hyperparameter sets (dropout rate, units, batch size) to the $Q_t$ prediction from lumped PIML model. The fourth layer predicts ratio reservoir storage ($S_t$) with maximum storage capacity ($S_{max}$) using the same inputs required for the prediction of $R_t$. However, the inputs are not preprocessed for the third and fourth layers. To impose physical constraint over reservoir storage, we follow a similar approach for predicting the ratio of $ET_t$ with $PET_t$. The predicted $R_t$ is then fed to the fifth layer of the LSTM model with downstream part precipitation, $ETd/s_t$ (from the second layer), soil moisture store, and groundwater store at the current timestep and previous timesteps based on the DELAY parameter of Muskingum routing in the downstream part of the catchment to predict $Q_t$. For the fifth layer, we kept similar model settings (activation function, optimizer, loss function) and hyperparameter sets (dropout rate, units, batch size) as of in $Q_t$ prediction from lumped PIML model. Since each of the LSTM models in the semi-distributed PIML with reservoir predicts a single variable for a given timestep, all of them are operated using a single dense layer. Table S7 shows the final hyperparameters used in the semi-distributed PIML with a reservoir model.

# 5 Results and discussions

## 5.1 Performance evaluation of lumped model

We compared the performances of SIMHYD and PIML models in the predictions of evapotranspiration and streamflow. While the results of the ML model are also compared for streamflow. We used the performance metrics including NSE, RMSE and PBIAS to evaluate the models. Figure 3 shows NSE, RMSE, and PBIAS in the subplots (a), (b), and (c), respectively for the model predictions in the testing period. In the actual evapotranspiration (ET) predictions, the PIML model shows higher NSE (Figure 3(a)) and lower RMSE (Figure 3(b)) than the SIMHYD model while the PIML model shows PBIAS (Figure 3(c)) near to zero as compared to SIMHYD model in the all catchments. Thus it shows that the PIML model outperforms the conceptual model in predicting the intermediate variable (actual evapotranspiration in this case) while ensuring the physical constraint over its ratio with $PET$. For streamflow (Q) predictions, the PIML model displays higher NSE (Figure 3(a)), lower RMSE (Figure 3(b)), and lesser PBIAS (in magnitude) (Figure 3(c)) than SIMHYD model while ML model performs well in terms of RMSE and PBIAS than SIMHYD. The Kantamal and Keesara catchments shows lesser PBIAS (in magnitude) for ML models than PIML and SIMHYD models, however its poor NSE and higher RMSE values indicates that PIML model performs better than SIMHYD and ML model in all the catchments. Thus, PIML shows robustness in the predictions of intermediate (ET) and target (Q) variables.

## 5.2 Performance evaluation of semi-distributed without reservoir model

Here we compare the performance of SIMHYD and PIML models in the evapotranspiration and streamflow predictions in both upstream and downstream parts of the catchment. Figure 4 shows model performance in terms of NSE, RMSE, and PBIAS in the subplots (a), (b), and (c), respectively. In $ETu/s$ prediction, the PIML model shows higher NSE (Figure 4(a)), lower RMSE (Figure 4(b)) and lesser PBIAS (in magnitude) (Figure 4(c)) than SIMHYD model. Similar performance is shown for the prediction of $ETd/s$.

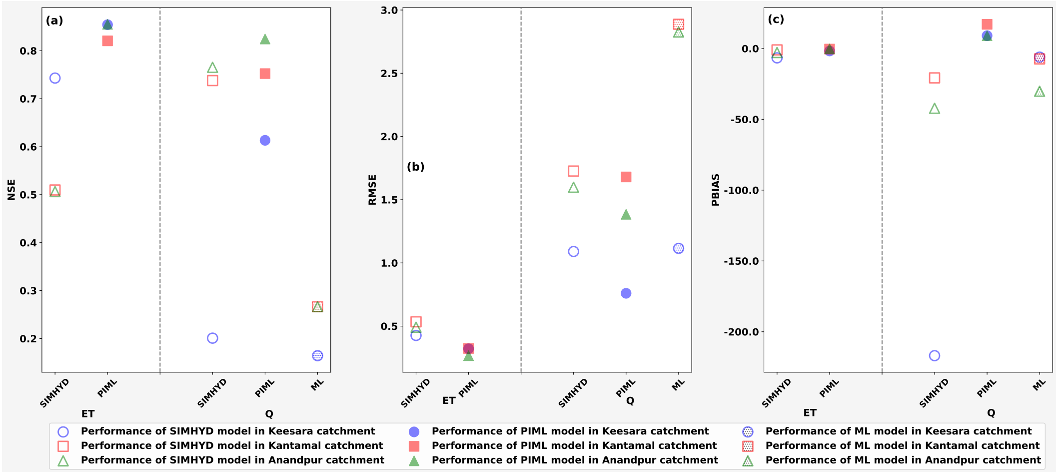


**Figure 3.** Performance assessment of lumped SIMHYD, PIML, and ML models in testing period. (a) The NSE is plotted for the prediction of ET, and Q. Hollow, filled and filled with hatching markers, shows the performance of the SIMHYD, PIML, and ML models, respectively. (b) Similar to NSE, the RMSE is plotted for the prediction of ET and Q. (c) The PBIAS is plotted. The positive and negative PBIAS value shows underestimation and overestimation in the model output.

Thus, the PIML model outperforms the SIMHYD model in predicting $ET$, an important intermediate variable in the rainfall-runoff process. We note that all the daily $ET$ values predicted by PIML follow its physical constraint with $PET$, which is achieved through proper choice of the activation function (sigmoid) and predictand (ratio of $ET$ with $PET$). In the case of upstream streamflow predictions, Anandpur and Keesara catchments show a higher NSE (Figure 4(a)) in the PIML model than SIMHYD model, while for the Kantamal catchment, both models show comparable NSE values. The PIML model shows lesser RMSE than the SIMHYD model in all three catchments (Figure 4(b)). In the Keesara and Kantamal catchments, the PIML model shows lesser PIBAS (in magnitude) than the SIMHYD model while conversely for the Anandpur catchment. However, overall PIML model performs better in predicting streamflow at the outlet of the upstream part of the catchment. The PIML model outperforms the SIMHYD model while the former shows higher NSE (Figure 4(a)), lower RMSE (Figure 4(b)) and lesser PBIAS (in magnitude) (Figure 4(c)) in comparison with later in downstream streamflow prediction. While getting better predictions, we ensured physical constraint in the contribution of upstream part streamflow at the outlet of the downstream part by employing a custom loss function (Eq. 8). Thus the semi-distributed without reservoir PIML model follows physical constraints and has better predictability than the SIMHYD model.

### 5.3 Performance evaluation of semi-distributed with reservoir model

Across the globe, around 77 % of the rivers are influenced by reservoir operation (Grill et al., 2019). Thus it is imperative to consider the reservoir in developing a hydrological model to study managed catchments. The applicability of the proposed semi-distributed PIML model with reservoir is demonstrated on two US catchments. In both catchments, $ETu/s$ and $ETd/s$ predictions of the PIML model show higher NSE (Figure 5(a)), lower RMSE (Figure 5(b)) and lesser PBIAS (in magnitude) (Figure 5(c)) than SIMHYD model while following physical constraint with $PET$. In the case of $R_t$ predictions, the PIML model displays higher NSE (Figure 5(a)), lower RMSE (Figure 5(b)) in comparison with SIMHYD model while in the PBIAS case it shows higher and lower

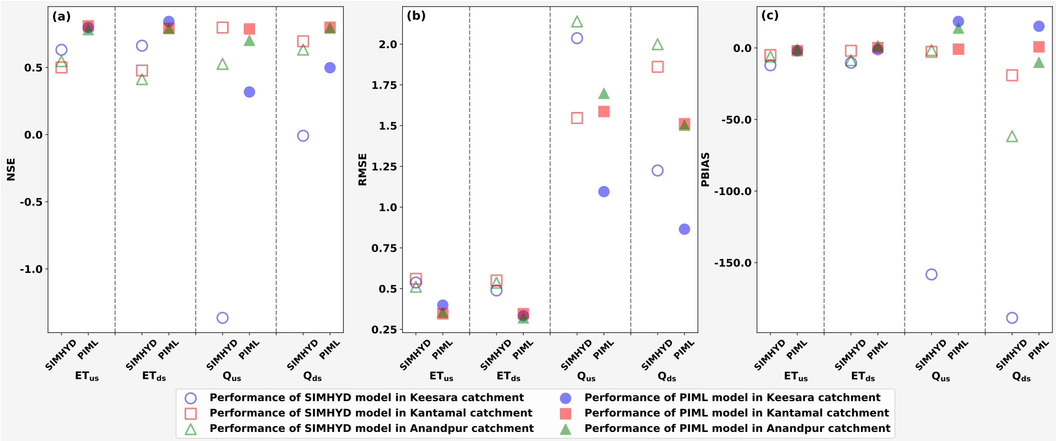


**Figure 4.** Performance assessment of semi-distributed without reservoir SIMHYD, and PIML models in testing period. (a) The NSE is plotted for the prediction of ET, and Q for both upstream and downstream parts of catchment. Hollow, and filled shows the performance of the SIMHYD, and PIML models, respectively. (b) Similarly, the RMSE is plotted. (c) The PBIAS is plotted for upstream and downstream part ET and Q.

value (in magnitude) for Brady and Canyon catchments respectively than the SIMHYD model. The PIML model shows higher NSE (Figure 5(a)), lower RMSE (Figure 5(b)) and lesser PBIAS (in magnitude) (Figure 5(c)) for $S_t$ predictions than SIMHYD model. We ensured that the PIML model gives a meaningful prediction of $S_t$ while imposing physical constraint with the help of the sigmoid activation function and proper choice of predictand (ratio of $S_t$ with $S_{max}$) to consistent with the output of activation function. Though the SIMHYD model shows negative NSE for $S_t$ predictions in both catchments, it shows a high correlation (0.76 for Brady catchment and 0.80 for Canyon catchment) with observed reservoir storages. The PIML model gives a robust performance in predicting streamflow at the outlet of the downstream part of the catchment with higher NSE (Figure 5(a)), lower RMSE (Figure 5(b)) and lesser PBIAS (in magnitude) (Figure 5(c)) than SIMHYD model. Thus semi-distributed PIML with a reservoir model outperforms the SIMHYD model while ensuring physical consistency at various stages.

### 5.4 Water balance and runoff coefficient analysis

We evaluate the physical consistency of the SIMHYD and PIML models using water balance. As precipitation data is the same for both models, we calculate deviation in the average annual sum of $ET$ and $Q$ with the average annual sum of observed data for respective variables in the testing period. For Keesara, Kantamal, and Anandpur catchments, we considered three cases, including an upstream part in the semi-distributed model, the total catchment in the semi-distributed model, and the total catchment in the lumped model. For example, a deviation is calculated for the average annual sum of model simulated $ETu/s$ and $Qu/s$ with the average annual sum of observed $ETu/s$ and $Qu/s$ for both SIMHYD and PIML models. In the Keesara catchment, all three cases of PIML model shows lesser deviation than the SIMHYD model (Figure 6(a)). Similar results are obtained in the Kantamal (Figure 6(b)) and Anandpur (Figure 6(c)) catchments. Also, we noted that the semi-distributed PIML model shows lesser deviation than lumped PIML model for the Kantamal (Figure 6(b)) and Anandpur (Figure 6(c)) catchments while Keesara catchment (Figure 6(a)) it shows comparable values which implies that semi-distributed structure can encapsulate spatial heterogeneity while performing better than lumped model structure. We did a similar analysis for Brady (Figure 6(d)) and Canyon (Figure 6(e))

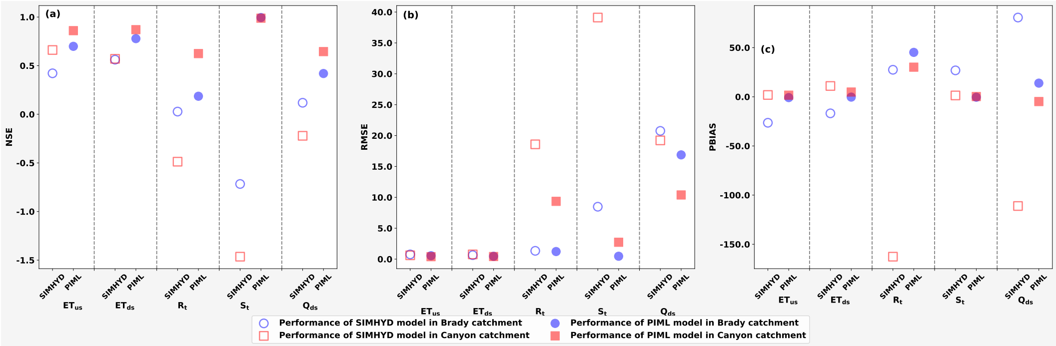


**Figure 5.** Performance assessment of semi-distributed with reservoir SIMHYD, and PIML model in testing period. (a) The NSE is plotted for the prediction of ET, reservoir storage (S), reservoir release (R) in the upstream and ET, Q for downstream parts of catchment. Hollow, and filled shows the performance of the SIMHYD, and PIML models, respectively. (b) Similarly, the RMSE is plotted. (c) The PBIAS is plotted for the ET, S, and R in the upstream part and ET, Q in the downstream part of the catchments.

catchments while accounting for reservoir storage and release. It considers two cases, which includes an upstream part in the semi-distributed model, the total catchment in the semi-distributed model. In Brady and Canyon catchments, the PIML models show lesser deviation in the both cases than its respective values for SIMHYD model. Overall, the PIML model shows consistent performance irrespective of the scale (lumped or semi-distributed) and catchment type (managed or unmanaged).

We noted that the *ET* dataset used in this study is the GLEAM model output. Thus we investigated deeper while calculating the average annual runoff coefficient and compared it with the observed. Similar to the previous deviation analysis, three cases, viz. upstream part in the semi-distributed model, total catchment in the semi-distributed model, and the total catchment in the lumped model, are considered for Keesara, Kantamal and Anandpur catchments. In the Keesara (Figure 6(f)) and Kantamal (Figure 6(g)) catchments, for all three cases, the PIML model shows a runoff coefficient close to the observed value than the SIMHYD model cases. However, the lumped PIML and semi-distributed PIML models show comparable performance in the Keesara catchment. While in the Kantamal catchment, semi-distributed PIML shows better agreement with observed than lumped PIML model. In the Anandpur (Figure 6(h)) catchment, for the upstream part SIMHYD model shows a runoff coefficient close to observed as compared to the PIML model. However, both lumped and semi-distributed PIML performs better in terms of runoff coefficient than respective SIMHYD model cases for the total catchment. Similar analysis is carried out for with reservoir case. For Brady and Canyon catchments we consider upstream part in the semi-distributed model, and total catchment in the semi-distributed model for runoff coefficient analysis. In the Brady (Figure 6(i)) and Canyon (Figure 6(j)) catchments, the PIML model shows a runoff coefficient closer to observed than the SIMHYD model for both upstream part and total catchment. This runoff coefficient analysis highlights that runoff is also modeled well in the PIML model compared to the SIMHYD model. Thus it shows robustness of the PIML model in predicting physically consistent outputs.

## 6 Conclusion

The PIML approach facilitates the synergistic use of interpretability from conceptual models and predictability from data-driven models. In this study, we have devel-

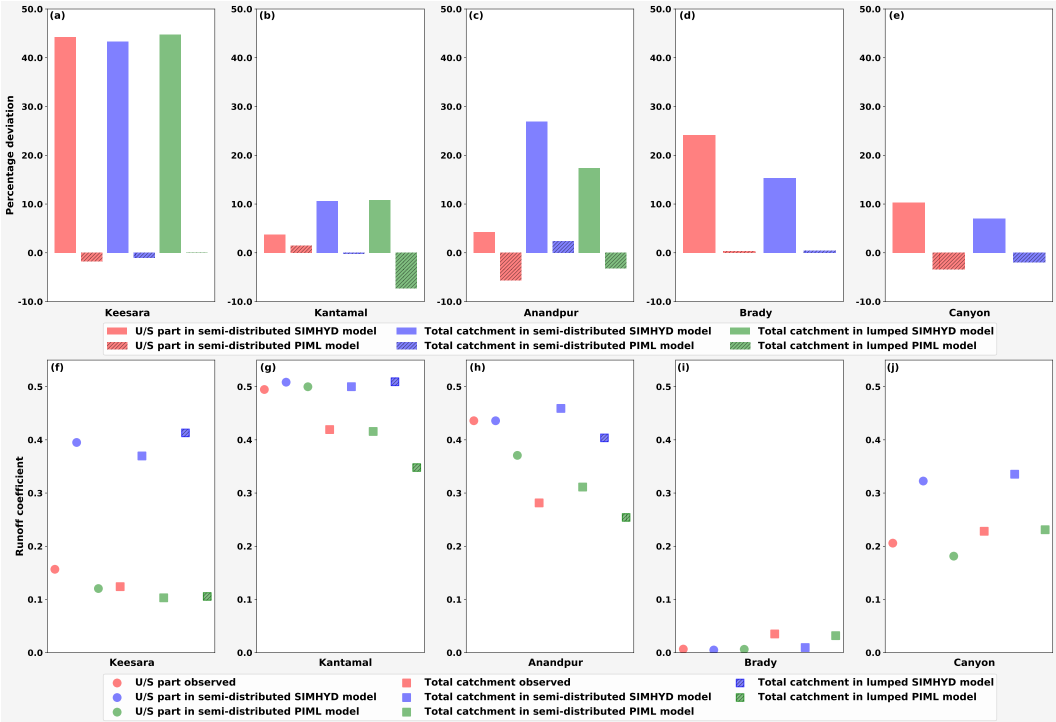


**Figure 6.** The first row shows comparison of deviation in the average annual sum of $ET$ and $Q$ with the average annual sum of observed data for respective variables in the testing period. It includes upstream part of catchment and total catchment in semi-distributed without reservoir model cases and total catchment in lumped model case for SIMHYD and PIML models. The catchments included in the analysis are: (a) Keesara, (b) Kantamal, (c) Anandpur, (d) Brady, and (e) Canyon; The second row compares average annual runoff coefficients of upstream part of the catchment and total catchment in semi-distributed without reservoir model cases as well as for total catchment in lumped model case for SIMHYD and PIML models with observed average annual runoff coefficient. The catchment used for this analysis are listed as: (f) Keesara, (g) Kantamal, (h) Anandpur, (i) Brady, and (j) Canyon.

oped the PIML model, which accounts for memory in the hydrological processes and provides interpretability through an intermediate variable. The predictors in the PIML model are selected based on the functional relationship shown by the conceptual (SIMHYD) model governing equations. Also, this study attempts to take advantage of long-term information learning capability in the LSTM model, which encapsulates the catchment response with temporal lag. We demonstrated three model cases considering different scales and mode of operation in the catchment. These three cases includes lumped model structure, semi-distributed model structures with and without reservoir. Our results shows that the PIML outperforms the conceptual as well as simple data-driven model. Also, water balance and runoff coefficient analysis shows that the PIML model predicts physically consistent outputs. The PIML is now materialized for hydrological processes as we demonstrated its application at both temporal (daily, monthly (Bhasme et al., 2022)) and spatial scales (lumped (Bhasme et al., 2022), semi-distributed) and also with managed and unmanaged catchments. We argue that our PIML modeling approach can make conceptual models more modular as it can be applied irrespective of the region for which it is developed. The application of PIML in different climatic as well as geographical regions shows its generalizability. The PIML approach has already shown flexibility in incorporating different ML methods in the conceptual model premise (Bhasme et al., 2022). Also, it shows the opportunity to build a flexible modeling framework similar to SUPERFLEX (Fenicia et al., 2011), where the modeler has choices for both modeling components, which accounts for physical processes and ML models to learn the complex interaction between the different components.

The current conceptual modeling approach is based on the mass balance where evapotranspiration (ET) is mainly dependent on precipitation and soil moisture. Researchers have shown the significance of soil moisture in flood modeling and forecasting (Wasko et al., 2020; Nanditha et al., 2022). However, the ET estimation largely rules the accuracy in the soil moisture estimation before the flood events, while the empirical relationship of actual ET with PET considers water balance and ignores other factors, including meteorological conditions (Fang et al., 2017). The inclusion of energy balance in modeling will serve the aforementioned purposes as hydrological processes are governed by both water balance as well as energy balance. ET prediction is a non-linear process, which can be handled better with the ML model (Walls et al., 2020; Cui et al., 2021). We can apply a similar approach as PIML while exploiting ML predictive ability in identifying complex non-linear relationships between ET and its governing factors in a separate ET modeling component. Further merging it in the overall model structure ensures both energy balance and mass balance.

## Acknowledgments

Funding for the project is provided by Scheme for Transformational and Advanced Research in Sciences of Ministry of Education implemented by Indian Institute of Science, Bangalore (Research Project ID: 367 titled 'Physics Guided Data Science Approach for Predictive Understanding of Hydrological Processes'). The authors thank IIT Gandhinagar colleagues Shekhar Goyal and Sarth Dubey for comments on the manuscript. The required precipitation dataset for Indian catchments is obtained from India Meteorological Department (IMD) (`https://www.imdpune.gov.in/`) while for US catchments it is sourced from Daymet (Daily Surface Weather Data on a 1-km Grid for North America, Version 4 R1). Actual and potential evapotranspiration datasets are obtained from the latest version of (v3.6a) of Global Land Evaporation Amsterdam Model (GLEAM) (`https://www.gleam.eu/`) datasets. The streamflow datasets for hydrological observation stations in India and US are obtained from India Water Resources Information System (India-WRIS; `https://indiawris.gov.in/wris/`) portal and United States Geological Survey (USGS) (`https://waterdata.usgs.gov/nwis`), respectively. The reservoir storage data is obtained from Texas Water Development Board (`https://www.waterdatafortexas.org/reservoirs/statewide`) and reservoir release is obtained from USGS for sites USGS 08145000 and USGS 08167800 for Brady Creek reservoir and Canyon lake, respectively and consideration of these stations for release data is consistent with ResOpsUS (Steyaert et al., 2022), a recently developed inventory of observed reservoir operations for conterminous United States (CONUS). All ML and PIML models code are available at GitHub (`https://github.com/pravin2408/PIML_daily_predictions_WRR`

## References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Retrieved from `https://www.tensorflow.org/` (Software available from tensorflow.org)

Ajami, N. K., Gupta, H., Wagener, T., & Sorooshian, S. (2004). Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of hydrology*, *298*(1-4), 112–135.

Aronica, G., & Cannarozzo, M. (2000). Studying the hydrological response of urban catchments using a semi-distributed linear non-linear model. *Journal of*

*Hydrology*, *238*(1-2), 35–43.
Bhasme, P., Vagadiya, J., & Bhatia, U. (2022). Enhancing predictive skills in physically-consistent way: Physics informed machine learning for hydrological processes. *Journal of Hydrology*, *615*, 128618.
Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., ... others (2019). Twenty-three unsolved problems in hydrology (uph)–a community perspective. *Hydrological sciences journal*, *64*(10), 1141–1158.
Chiew, F., Kirono, D., Kent, D., Frost, A., Charles, S., Timbal, B., ... Fu, G. (2010). Comparison of runoff modelled using rainfall from different downscaling methods for historical and future climates. *Journal of Hydrology*, *387*(1-2), 10–23.
Chiew, F., Peel, M., Western, A., et al. (2002). Application and testing of the simple rainfall-runoff model simhyd. *Mathematical models of small watershed hydrology and applications*, 335–367.
Cui, Y., Song, L., & Fan, W. (2021). Generation of spatio-temporally continuous evapotranspiration and its components by coupling a two-source energy balance model and a deep neural network over the heihe river basin. *Journal of Hydrology*, *597*, 126176.
Das, T., Bárdossy, A., Zehe, E., & He, Y. (2008). Comparison of conceptual model performance using different representations of spatial variability. *Journal of Hydrology*, *356*(1-2), 106–118.
David, C. H., Maidment, D. R., Niu, G.-Y., Yang, Z.-L., Habets, F., & Eijkhout, V. (2011). River network routing on the nhdplus dataset. *Journal of Hydrometeorology*, *12*(5), 913–934.
Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, *6*(2), 182–197.
Devia, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A review on hydrological models. *Aquatic Procedia*, *4*, 1001–1007.
Döll, P., Kaspar, F., & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: model tuning and validation. *Journal of Hydrology*, *270*(1-2), 105–134.
Ekka, A., Keshav, S., Pande, S., van der Zaag, P., & Jiang, Y. (2022). Dam-induced hydrological alterations in the upper cauvery river basin, india. *Journal of Hydrology: Regional Studies*, *44*, 101231.
Fang, Y.-H., Zhang, X., Corbari, C., Mancini, M., Niu, G.-Y., & Zeng, W. (2017). Improving the xin'anjiang hydrological model based on mass–energy balance. *Hydrology and Earth System Sciences*, *21*(7), 3359–3375.
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.
Fenicia, F., Kavetski, D., & Savenije, H. H. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. motivation and theoretical development. *Water Resources Research*, *47*(11).
Fenicia, F., Meißner, D., & McDonnell, J. J. (2022). Modeling streamflow variability at the regional scale:(2) development of a bespoke distributed conceptual model. *Journal of Hydrology*, *605*, 127286.
Fowler, K. J., Peel, M. C., Western, A. W., Zhang, L., & Peterson, T. J. (2016). Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models. *Water Resources Research*, *52*(3), 1820–1846.
Frame, J. M., Kratzert, F., Klotz, D., Gauch, M., Shelev, G., Gilon, O., ... Nearing, G. S. (2022). Deep learning rainfall–runoff predictions of extreme events. *Hydrology and Earth System Sciences*, *26*(13), 3377–3392.
Frame, J. M., Kratzert, F., Raney, A., Rahman, M., Salas, F. R., & Nearing, G. S.

(2021). Post-processing the national water model with long short-term memory networks for streamflow predictions and model diagnostics. *JAWRA Journal of the American Water Resources Association*, *57*(6), 885–905.

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., & Hochreiter, S. (2021). Rainfall–runoff prediction at multiple timescales with a single long short-term memory network. *Hydrology and Earth System Sciences*, *25*(4), 2045–2062.

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., ... others (2019). Mapping the world's free-flowing rivers. *Nature*, *569*(7755), 215–221.

Gutenson, J. L., Tavakoly, A. A., Wahl, M. D., & Follum, M. L. (2020). Comparison of generalized non-data-driven lake and reservoir routing models for global-scale hydrologic forecasting of reservoir outflow at diurnal time steps. *Hydrology and Earth System Sciences*, *24*(5), 2711–2729.

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

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

Jia, X., Zwart, J., Sadler, J., Appling, A., Oliver, S., Markstrom, S., ... others (2021). Physics-guided recurrent graph model for predicting flow and temperature in river networks. In *Proceedings of the 2021 siam international conference on data mining (sdm)* (pp. 612–620).

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., ... Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on knowledge and data engineering*, *29*(10), 2318–2331.

Khandelwal, A., Xu, S., Li, X., Jia, X., Stienbach, M., Duffy, C., ... Kumar, V. (2020). Physics guided machine learning methods for hydrology. *arXiv preprint arXiv:2012.02854*.

Khosravi, K., Golkarian, A., & Tiefenbacher, J. P. (2022). Using optimized deep learning to predict daily streamflow: A comparison to common machine learning algorithms. *Water Resources Management*, *36*(2), 699–716.

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.

Kratzert, F., Klotz, D., Hochreiter, S., & Nearing, G. S. (2021). A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall–runoff modeling. *Hydrology and Earth System Sciences*, *25*(5), 2685–2703.

Li, F., Zhang, Y., Xu, Z., Liu, C., Zhou, Y., & Liu, W. (2014). Runoff predictions in ungauged catchments in southeast tibetan plateau. *Journal of Hydrology*, *511*, 28–38.

Li, F., Zhang, Y., Xu, Z., Teng, J., Liu, C., Liu, W., & Mpelasoka, F. (2013). The impact of climate change on runoff in the southeastern tibetan plateau. *Journal of Hydrology*, *505*, 188–201.

Li, H., & Zhang, Y. (2017). Regionalising rainfall-runoff modelling for predicting daily runoff: Comparing gridded spatial proximity and gridded integrated similarity approaches against their lumped counterparts. *Journal of Hydrology*, *550*, 279–293.

Li, K., Huang, G., Wang, S., & Razavi, S. (2022). Development of a physics-informed data-driven model for gaining insights into hydrological processes in irrigated watersheds. *Journal of Hydrology*, *613*, 128323.

Liu, B., Tang, Q., Zhao, G., Gao, L., Shen, C., & Pan, B. (2022). Physics-guided long short-term memory network for streamflow and flood simulations in the lancang–mekong river basin. *Water*, *14*(9), 1429.

Liu, Z., Zhou, P., Chen, X., & Guan, Y. (2015). A multivariate conditional model for streamflow prediction and spatial precipitation refinement. *Journal of Geophysical Research: Atmospheres*, *120*(19), 10–116.

Lu, D., Konapala, G., Painter, S. L., Kao, S.-C., & Gangrade, S. (2021). Streamflow simulation in data-scarce basins using bayesian and physics-informed machine learning models. *Journal of Hydrometeorology*.

Martens, B., Miralles, D. G., Lievens, H., Schalie, R. v. d., De Jeu, R. A., Fernández-Prieto, D., ... Verhoest, N. E. (2017). Gleam v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925.

Miralles, D. G., Holmes, T., De Jeu, R., Gash, J., Meesters, A., & Dolman, A. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, *15*(2), 453–469.

Mostafaie, A., Forootan, E., Safari, A., & Schumacher, M. (2018). Comparing multi-objective optimization techniques to calibrate a conceptual hydrological model using in situ runoff and daily grace data. *Computational Geosciences*, *22*(3), 789–814.

Mpelasoka, F. S., & Chiew, F. H. (2009). Influence of rainfall scenario construction methods on runoff projections. *Journal of Hydrometeorology*, *10*(5), 1168–1183.

Nanditha, J., Rajagopalan, B., & Mishra, V. (2022). Combined signatures of atmospheric drivers, soil moisture, and moisture source on floods in narmada river basin, india. *Climate Dynamics*, 1–21.

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.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., ... Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research*, *57*(3), e2020WR028091.

Nevo, S., Morin, E., Gerzi Rosenthal, A., Metzger, A., Barshai, C., Weitzner, D., ... others (2022). Flood forecasting with machine learning models in an operational framework. *Hydrology and Earth System Sciences*, *26*(15), 4013–4032.

O'Sullivan, J., Ahilan, S., & Bruen, M. (2012). A modified muskingum routing approach for floodplain flows: theory and practice. *Journal of Hydrology*, *470*, 239–254.

Parisouj, P., Mohebzadeh, H., & Lee, T. (2020). Employing machine learning algorithms for streamflow prediction: a case study of four river basins with different climatic zones in the united states. *Water Resources Management*, *34*(13), 4113–4131.

Parisouj, P., Mokari, E., Mohebzadeh, H., Goharnejad, H., Jun, C., Oh, J., & Bateni, S. M. (2022). Physics-informed data-driven model for predicting streamflow: A case study of the voshmgir basin, iran. *Applied Sciences*, *12*(15). Retrieved from `https://www.mdpi.com/2076-3417/12/15/7464` doi: 10.3390/app12157464

Paul, P. K., Gaur, S., Kumari, B., Panigrahy, N., Mishra, A., & Singh, R. (2019). Diagnosing credibility of a large-scale conceptual hydrological model in simulating streamflow. *Journal of Hydrologic Engineering*, *24*(4), 04019004.

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of hydrology*, *279*(1-4), 275–289.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, *566*(7743), 195–204.

Ren-Jun, Z. (1992). The xinanjiang model applied in china. *Journal of hydrology*, *135*(1-4), 371–381.

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, *54*(11), 8558–8593.

Shin, M.-J., Guillaume, J. H., Croke, B. F., & Jakeman, A. J. (2015). A review of foundational methods for checking the structural identifiability of models: Results for rainfall-runoff. *Journal of Hydrology*, *520*, 1–16.

Siriwardena, L., Finlayson, B., & McMahon, T. (2006). The impact of land use change on catchment hydrology in large catchments: The comet river, central queensland, australia. *Journal of Hydrology*, *326*(1-4), 199–214.

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.

Swain, J. B., & Patra, K. C. (2017). Streamflow estimation in ungauged catchments using regionalization techniques. *Journal of Hydrology*, *554*, 420–433.

Thapa, S., Zhao, Z., Li, B., Lu, L., Fu, D., Shi, X., ... Qi, H. (2020). Snowmelt-driven streamflow prediction using machine learning techniques (lstm, narx, gpr, and svr). *Water*, *12*(6), 1734.

Thornton, M., Shrestha, R., Wei, Y., Thornton, P., Kao, S.-C., & Wilson, B. (2022). *Daymet: Daily surface weather data on a 1-km grid for north america, version 4.* ORNL Distributed Active Archive Center. Retrieved from https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=2129 doi: 10.3334/ORNLDAAC/2129

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.

Vaze, J., Post, D., Chiew, F., Perraud, J.-M., Viney, N., & Teng, J. (2010). Climate non-stationarity–validity of calibrated rainfall–runoff models for use in climate change studies. *Journal of Hydrology*, *394*(3-4), 447–457.

Wagena, M. B., Goering, D., Collick, A. S., Bock, E., Fuka, D. R., Buda, A., & Easton, Z. M. (2020). Comparison of short-term streamflow forecasting using stochastic time series, neural networks, process-based, and bayesian models. *Environmental Modelling & Software*, *126*, 104669.

Walls, S., Binns, A. D., Levison, J., & MacRitchie, S. (2020). Prediction of actual evapotranspiration by artificial neural network models using data from a bowen ratio energy balance station. *Neural Computing and Applications*, *32*(17), 14001–14018.

Wasko, C., Nathan, R., & Peel, M. C. (2020). Changes in antecedent soil moisture modulate flood seasonality in a changing climate. *Water Resources Research*, *56*(3), e2019WR026300.

Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, *55*(4), 1–37.

Wu, Y., Chen, Y., & Tian, Y. (2022). Incorporating empirical orthogonal function analysis into machine learning models for streamflow prediction. *Sustainability*, *14*(11), 6612.

Yaseen, Z. M., Kisi, O., & Demir, V. (2016). Enhancing long-term streamflow forecasting and predicting using periodicity data component: application of artificial intelligence. *Water resources management*, *30*(12), 4125–4151.

Zhou, Y., Cui, Z., Lin, K., Sheng, S., Chen, H., Guo, S., & Xu, C.-Y. (2022). Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques. *Journal of Hydrology*, *604*, 127255.