MID-AMERICA TRANSPORTATION CENTER

Report # MATC-UI: 004-53

Final Report

WBS: 25-1121-0005-004-53

A Machine Learning-Based System for Predicting Peak Flowrates of Nebraska Streams

Tirthankar Roy, PhD
Assistant Professor
Department of Civil and Environmental Engineering
University of Nebraska-Lincoln

Sudan Pokharel, MS
Graduate Research Assistant

David Admiraal, PhD
Associate Professor

2023

A Cooperative Research Project sponsored by
U.S. Department of Transportation- Office of the Assistant Secretary for Research and Technology

A Machine Learning-Based System for Predicting Peak Flowrates of Nebraska Streams

Tirthankar Roy, Ph.D.
Assistant Professor
Civil and Environmental Engineering
University of Nebraska-Lincoln

David Admiraal, Ph.D.
Associate Professor
Civil and Environmental Engineering
University of Nebraska-Lincoln

Sudan Pokharel, M.S.
Graduate Research Assistant
Civil and Environmental Engineering
University of Nebraska-Lincoln

A Report on Research Sponsored by

Mid-America Transportation Center
University of Nebraska–Lincoln

October 2023

Technical Report Documentation Page

Report No. 25-1121-0005-004-53	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A Machine Learning-Based System for Predicting Peak Flowrates of Nebraska Streams		5. Report Date October 2023
		6. Performing Organization Code
7. Author(s) Tirthankar Roy, PhD ORCID: 0000-0002-6279-8447 Sudan Pokharel, M.S., ORCID: 0009-0009-7010-698X David Admiraal, PhD ORCID: 0000-0002-6654-3740		8. Performing Organization Report No. 25-1121-0005-004-53
9. Performing Organization Name and Address Department of Civil and Environmental Engineering University of Nebraska-Lincoln 900 N 16th St Lincoln, NE 68588-0531		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 69A3551747107
12. Sponsoring Agency Name and Address Mid-America Transportation Center Prem S. Paul Research Center at Whittier School 2200 Vine St. Lincoln, NE 68583-0851		13. Type of Report and Period Covered Final Report April 2021 – June 2023
		14. Sponsoring Agency Code MATC TRB RiP No. 91994-93
15. Supplementary Notes		

16. Abstract

Accurate, early flood warnings are invaluable for ensuring transportation safety so that flood-prone roads can be closed well before they become hazardous. A flood forecasting system typically uses both hydrologic and hydraulic models. The latter requires peak flow information to calculate flood attributes, such as depth, velocity, and inundated area. For ungauged basins, which lack streamflow observations, accurate calculation of flood attributes is challenging. Regional regression equations are used for this purpose. Such equations provide peak flow estimates based on flow records and basin characteristics of nearby gauged basins. Regression equations used for Nebraska, however, are decades old. The three available sets of equations often produce results that vary by order of magnitude. Therefore, there is a serious need to improve the accuracy of peak flow prediction using recent datasets and advanced methods. In this study, we modeled daily streamflow and peak flow in Nebraska streams using new high-resolution datasets and two machine learning algorithms, Long Short-term Memory Network (LSTM) for daily streamflow and Random Forest (RF) for peak flow. A wide range of predictors were used in the study. Physically based constraints were imposed on the LSTM model for daily streamflow simulations, and the benefits were assessed. We showed that there is a value in adding physics-based constraints even though the constraints do not improve results in all cases. Therefore, constraints can be applied as needed. Additionally, adding newer datasets and advanced machine learning algorithms leads to improved estimates of peak flow.

17. Key Words Flooding, Transportation Safety, Peak Flow, Machine Learning, LSTM, physics-based machine learning, streamflow prediction, Peak flow Estimation, Random Forest, Regional Regression, Flood, Resilient Infrastructure.	18. Distribution Statement		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 58	22. Price

Table of Contents

List of Figures

List of Tables

Acknowledgements

We acknowledge MATC for providing funding for this research.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

This report is based on two of our manuscripts from this project, as cited below. One of these manuscripts has been published already, and the other one is currently under review. The report uses texts and content from the manuscripts.

Published paper:

Pokharel, S., T. Roy, and D. Admiraal (2023), Effects of mass balance, energy balance, and storage-discharge constraints on LSTM for streamflow prediction, *Environmental Modeling & Software*, doi:10.1016/j.envsoft.2023.105730.

Manuscript under review:

Pokharel, S., T. Roy, and D. Admiraal, Machine Learning-based Peak flow Estimation for Nebraska Streams, under review in *Journal of Hydrology*.

Abstract

Accurate, early flood warnings are invaluable for ensuring transportation safety so that flood-prone roads can be closed well before they become hazardous. A flood forecasting system typically uses both hydrologic and hydraulic models. The latter requires peak flow information to calculate flood attributes, such as depth, velocity, and inundated area. For ungauged basins, which lack streamflow observations, accurate calculation of flood attributes is challenging. Regional regression equations are used for this purpose. Such equations provide peak flow estimates based on flow records and basin characteristics of nearby gauged basins. Regression equations used for Nebraska, however, are decades old. The three available sets of equations often produce results that vary by order of magnitude. Therefore, there is a serious need to improve the accuracy of peak flow prediction using recent datasets and advanced methods. In this study, we modeled daily streamflow and peak flow in Nebraska streams using new high-resolution datasets and two machine learning algorithms, Long Short-Term Memory Network (LSTM) for daily streamflow and Random Forest (RF) for peak flow. A wide range of predictors were used in the study. Physically based constraints were imposed on the LSTM model for daily streamflow simulations, and the benefits were assessed. We showed that there is a value in adding physics-based constraints even though the constraints do not improve results in all cases. Therefore, constraints can be applied as needed. Additionally, adding newer datasets and advanced machine learning algorithms leads to improved estimates of peak flow.

Chapter 1 General Introduction

Nebraska is prone to hydrologic extremes such as floods and droughts, as intermittent extreme weather events are formed because of short-duration convergent airflow of contrasting characteristics. It is not surprising that the place is called the 'Heartland of Extremes' (Dewey and Mogil, 2017). The most recent severe flooding occurred in the Spring of 2019, where a rain-on-snow (ROS) event caused the rapid melting of large snowpacks (Flanagan et al. 2019). Rapid snowmelt and ROS events are known contributors to flooding in the Midwest (Villarini et al. 2011). The 2019 event caused an estimated loss exceeding $3 billion (as of August 2019), and lives were also lost (Flanagan et al. 2019). Transportation systems, including roads, bridges, and train lines, experienced significant damage.

Accurate and early flood predictions are crucial for ensuring transportation safety. Based on these predictions, authorities can close transportation systems likely to be impacted by flooding, which can help mitigate detrimental impacts. A flood forecasting system typically uses a combination of hydrologic and hydraulic models. The hydrologic model quantifies watershed outflow contributions, whereas the hydraulic model calculates the propagation of floodwater. In ungauged basins, due to the lack of streamflow observations, the hydraulic model is more complex, requiring the use of regional regression equations. These equations, which estimate peak flow rates for a wide range of return periods, are developed based on the information from gauged basins in the vicinity, assumed to be representative of the ungauged basins under consideration. There are three sets of equations available for Nebraska; however, they are decades old. Furthermore, the estimates from these three sets do not agree, often varying by order of magnitude. Therefore, it is crucial that these equations be replaced with superior techniques so that flood forecasting can be carried out with higher accuracy. In this study, we developed a system to model peak flow rates for different return periods in Nebraska streams

using state-of-the-art datasets and a machine learning algorithm. We also carried out daily streamflow simulations using another machine learning algorithm with physics-based constraints.

This report combines the content from two of our manuscripts that originated from this project. These are included as separate Chapters (2 and 3).

Chapter 2 Physics-based Machine Learning for Streamflow Estimation

The information relayed in this chapter is from a published paper: "Pokharel, S., T. Roy, and D. Admiraal (2023), Effects of mass balance, energy balance, and storage-discharge constraints on LSTM for streamflow prediction, *Environmental Modeling & Software*, doi:10.1016/j.envsoft.2023.105730."

We investigated the effects of physics-based constraints, added to the loss function of a Long Short-Term Memory (LSTM) network, on its performance in daily streamflow prediction. Three types of constraints (mass balance, energy balance, and storage-discharge relationship), along with their combinations, are tested across 34 river basins in Nebraska. We found that the addition of constraints improves the model performance in several basins, but there are also cases where the performance drops or does not differ significantly. Mass and energy balance constraints improve the performance in 38% and 32% of catchments, respectively, while storage-discharge constraints improve the performance in 12% of catchments. The combination of mass and energy balance constraints has a positive effect on 41% of catchments, while the combination of mass balance and storage-discharge constraints improves the performance in 26% of catchments. We recommend the use of constraints in cases where they boost the LSTM performance.

2.1 Introduction

Accurate streamflow prediction is of critical importance for various sectors, such as water resources planning and management, agricultural water management, disaster mitigation efforts, etc. (Dong et al., 2020; Gauch, Mai, et al., 2021; Ghimire et al., 2021; Hunt et al., 2022). As the demand for water resources continues to increase, the focus on modeling streamflow has intensified. With the proliferation of high-resolution mapping and other remote sensing datasets together with the availability of datasets at high spatial and temporal resolutions, we have seen a

rise in novel methods for modeling streamflow across different scales. Machine learning (ML) models, in particular, have shown significant promise in this area, with applications including flood forecasting (Castangia et al., 2023; Krajewski et al., 2017; Lin et al., 2013; Nevo et al., 2021; Yuan et al., 2018), flood early warning systems (Moon et al., 2019; Pyayt et al., 2014; Sanders et al., 2022), flood inundation (Chang et al., 2018; Kim & Han, 2020; Tanoue et al., 2016), flood prediction (Tongal & Booij, 2018), runoff prediction (Kratzert et al., 2018), and streamflow prediction (Parisouj et al., 2020; Tongal & Booij, 2018; Xiang & Demir, 2020; Yaseen et al., 2016). Despite the advancement, due to their restricted interpretability (Anderson & Radić, 2022; Bhasme et al., 2021; Gilpin et al., 2019; Reichstein et al., 2019), physical inconsistencies (Karpatne et al., 2017; Willard et al., 2020), and enduring problems with equifinality (Troch et al., 2015), ML algorithms are often criticized by domain experts.

Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997) have been widely implemented in the field of hydrology of late. These networks can store and regulate information over time (Gauch, Kratzert, et al., 2021; Hochreiter & Schmidhuber, 1997; Kratzert et al., 2019; Yang et al., 2023), which makes them particularly suitable for simulating the memory effects of different types of hydrological variables. While traditional recurrent neural networks (RNNs) are limited in their ability to remember long-term sequences in modeling time-dependent and sequential data (Bengio et al., 1994), LSTMs can leverage information from both short-term and long-term dependencies, which is advantageous for hydrologic applications like streamflow prediction (Khandelwal et al., 2020; Kratzert et al., 2018; Majeske et al., 2022; Song et al., 2020).

Currently, there have been efforts to combine physical understanding with ML models to improve their performance and interpretability (Daw et al., 2017, 2019; Hoedt et al., 2021;

Karpatne et al., 2017; Young et al., 2017). For example, to estimate lake temperature, Daw et al., (2017) presented a physics-guided neural network (PGNN) model, which fuses a physics-based model with a neural network. They established a framework for developing neural networks that utilize physical equations and principles, in which a crucial link between the temperature, density, and depth of water is incorporated into the physics-based loss function of the LSTM model. Jia et al., (2019) extended this work by including a physics-based penalty for energy conservation in the loss function to ensure consistency between the lake thermal energy gain over time and the net thermodynamic fluxes into and out of the lake, thereby capturing more complex and general physical relationships. It is our understanding that there has been limited research on adding physics-based constraints into the loss function of the ML model and what type of constraints to include for streamflow prediction. Incorporating physics-based constraints, such as water balance, energy balance, and storage discharge relationship, into the loss function of a ML model for streamflow prediction is important because it allows the model to consider real-world physical processes that affect the amount of water available for streamflow. These constraints play a vital role in accurately predicting streamflow by considering the interactions between various components of the hydrological system, such as precipitation, evaporation, and storage. Understanding these elements is essential for effective streamflow prediction since they all influence the amount of water that is available for streamflow. However, there is a lack of research on the specific types of physics-based constraints that should be included in the loss function of a machine learning model for streamflow prediction.

In this study, we present a framework for modeling streamflow in 34 basins in Nebraska using high-resolution datasets that incorporates three types of physics-based constraints (water balance, energy balance, and storage-discharge relationship) into the loss function of an LSTM

model. We evaluate the effectiveness of these constraints and examine their individual and combined effects on streamflow prediction by independently training the LSTM model for each catchment. We aim to provide a new approach for combining machine learning and physics-based constraints in streamflow prediction that can contribute to the field of hydrology.

2.2 Study area and data

2.2.1 Study area

The study area for this research was the Hydrologic Unit Code (HUC) 8 basins of Nebraska. Out of the 72 HUC-8 basins in the state, 34 were selected for this analysis based on the availability of stream gauge data for at least 20 years. The locations of the selected HUC-8 basins within Nebraska are shown in Figure 2.1. The selected basins represent a range of geographies and hydrological conditions within the state. Stream gauge data for these basins were used as the primary source of observations for the study. Basins that did not have a gauge station at the outlet or had limited years of data were excluded from the analysis. This study aims to provide a framework for improving streamflow prediction in these basins by incorporating physics-based constraints into the loss function of LSTM.

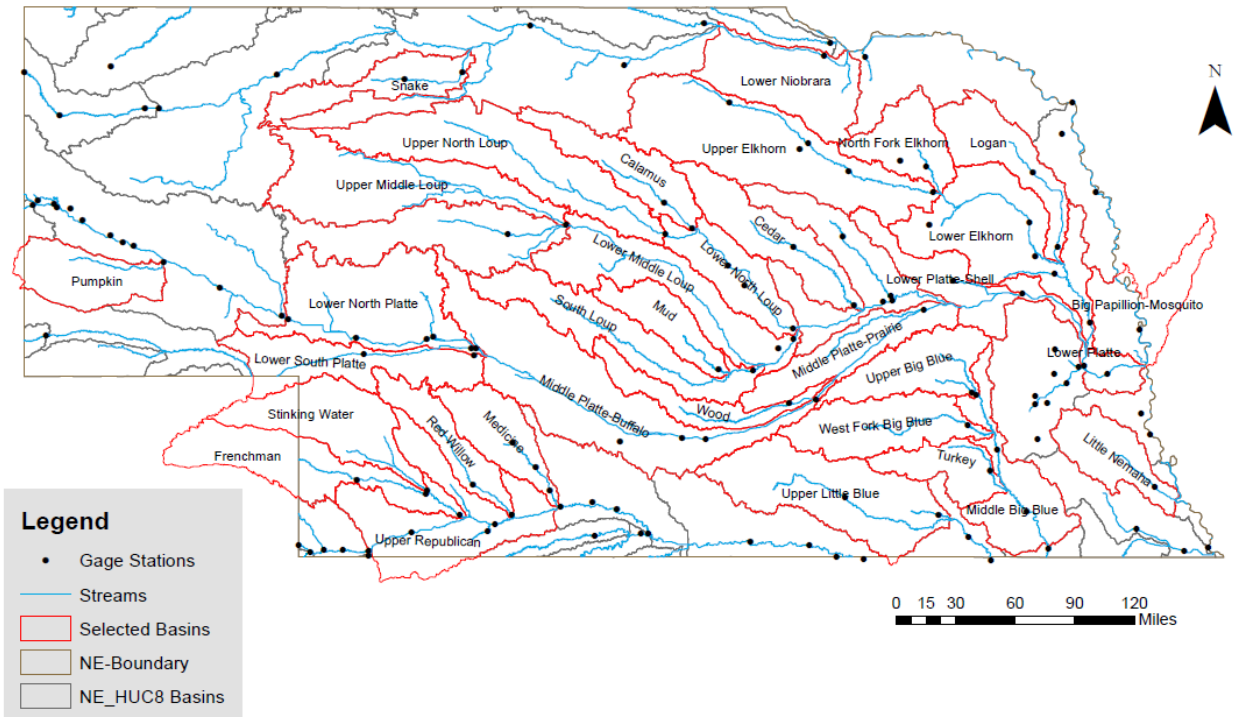

Figure 2.1 Nebraska HUC-8 Basins, red color basins are the basins used for analysis.

2.2.2 Data

The atmospheric and land data for this study was obtained from the ERA5 dataset (Hersbach et al., 2020) which is the fifth-generation atmospheric reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 provides global coverage on a 30 km grid with 137 levels of atmospheric resolution from the ground up to 80 km in altitude and includes hourly data for a wide range of atmospheric, land, and oceanic climate variables. In addition to ERA5, streamflow data were collected from the United States Geological Survey (USGS) and the Nebraska Department of Natural Resources (NeDNR). The variables considered in this study include wind speed components, dew point temperature, temperature, skin temperature, boundary layer height, convective available potential energy, actual evaporation, leaf area index, snow albedo, snow depth, soil temperature, streamflow, surface solar and

thermal radiation, surface pressure, surface sensible heat flux, total column water, total precipitation, and volumetric soil water. The gridded ERA5 variables were lumped over the basin extent to form a time series. The variables listed in the study are relevant in predicting streamflow because they all play a role in the hydrological cycle and understanding these variables and their interactions is crucial in predicting streamflow. For example, precipitation, evaporation, and snow elements in the hydrological cycle affect the amount of water available for streamflow. Similarly, temperature and soil-related variables affect the evaporation and infiltration rate which affects the streamflow. The wind components affect the evaporation rate and the transport of moisture, which are important factors in the formation of precipitation. The selection of the datasets for this study was motivated by reviewing previous studies in streamflow prediction (Dalkiliç & Hashimi, 2020; Damavandi et al., 2019; Ha et al., 2021; Hadi & Tombul, 2018; Ni et al., 2020; Pellicciotti et al., 2010; Peng et al., 2017; Samit Thapa, Zebin Zhao, Bo Li, 2020; Shortridge et al., 2016; Tongal & Booij, 2018).

2.3 Methodology

2.3.1 Long Short-term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a type of Recurrent Neural Network (RNN) that are capable of learning both short-term and long-term dependencies, as originally introduced by Hochreiter and Schmidhuber (1997). LSTM networks were developed to address the problem of vanishing and exploding gradients in traditional RNNs, allowing them to effectively model sequences with long time lags. The architecture of an LSTM network is shown in Figure 2.2 and consists of three main gate mechanisms: the forget gate, the input gate, and the output gate. The forget gate determines which pieces of the long-term memory should be discarded based on the previous hidden state and the current input data. The input gate

determines what new information should be added to the network's long-term memory given the previous hidden state and current input data. The output gate determines the next hidden state and is used to make final predictions. For a more detailed explanation of LSTM networks, see Hrnjica & Mehr, (2020).

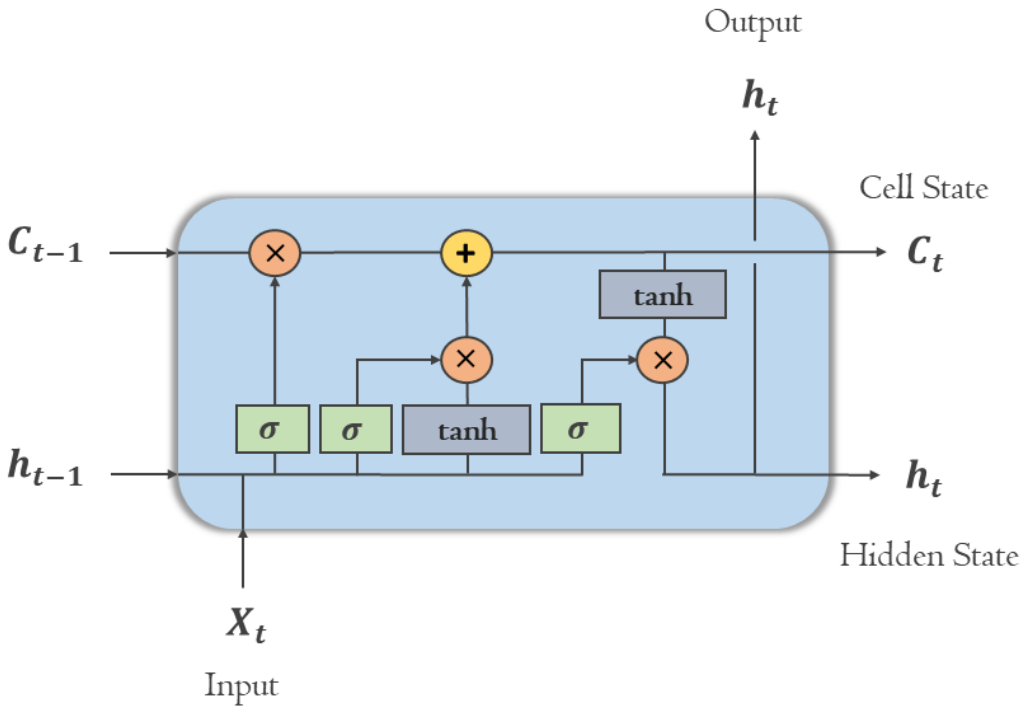

Figure 2.2 An LSTM architecture

The ability of a deep learning model, such as an LSTM network, to effectively learn and make predictions is highly dependent on the appropriate selection of hyperparameters. These hyperparameters, including the number of nodes or neurons in the network, the dropout rate, the activation function, the optimization algorithm, the learning rate, and the loss function, can significantly impact the model's performance. In this study, we implemented a network architecture comprising three stacked LSTM layers, including an input layer and two hidden layers. The input layer contained 80 neurons, while the hidden layers contained 50 and 10

neurons, respectively. The output layer was composed of a single neuron. We used '*relu*' activation function for all the stacked layers of LSTM and '*linear*' for the output layer. To prevent overfitting, a dropout rate of 30% (0.3) was applied before and after the final hidden layer. Overfitting occurs when a model fits too closely to the training data and is unable to generalize to new datasets. The use of dropout layers, which randomly ignore a subset of neurons during training, can help mitigate this issue. We used a lookback window of seven days for all independently trained LSTM models.

2.3.1.1 Training of LSTM model

In this study, the models were trained using a mini-batch training approach with a batch size of 50 and an epoch size of 200. The Adam optimization algorithm was used for training, which combines the momentum and Root Mean Square Propagation (RMSP) gradient descent techniques and is known to be effective for a wide range of deep learning models, including LSTMs. The mean-square error (MSE) was used as the loss function for the standalone LSTM model. For models that included physics-based constraints, the MSE loss function was modified, as described in later sections. To ensure the input features, including meteorological variables and discharge, were properly scaled, the LeCun et al. (2012) and Minns & Hall (2009) normalization method was applied, which involves subtracting the mean and dividing by the standard deviation. The mean and standard deviation were calculated using only the training period data. The normalized output of the network was then transformed using the normalization parameters from the training period to produce the final discharge predictions.

2.3.2 Physics-based loss function for prediction of streamflow

The loss function is a key component of any neural network, as it measures the error between the network's predictions and the true values. Mean Squared Error (MSE) is a widely

used loss function for regression problems in machine learning, and it is calculated as the mean of the squared differences between the actual values (y_true) and the predicted values (y_pred) over a dataset of size n.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_{true} - y_{pred})^2$$

where,
MSE = Mean square error
n = number of data points
y_{true} = actual(true) values
y_{pred} = predicted values

In this study, MSE was employed as the loss function for the standalone LSTM model. However, to incorporate physics-based constraints into the model, the loss function was modified as follows:

$$\text{Loss} = Loss_{TRN}(y_{true} - y_{pred}) + \lambda LossPHY(Ypred)$$

Where the training loss $Loss_{TRN}$ measures a supervised error such as MSE between true (y_{true}) and predicted labels (y_{pred}). The hyperparameter λ is a penalty factor that controls the relative importance of the physical constraint term L_PHY in the overall loss function. By adjusting λ appropriately, it is possible to enforce consistency with known physical laws in the model's predictions.

2.3.2.1 Conservation of Mass (Water Balance-WB)

Conservation of mass, also known as the law of mass balance, dictates that the inflow to a system must be equal to the outflow plus any changes in storage within the system. In the context

of a watershed, water is received through precipitation and snowmelt, and is lost through processes such as evaporation, streamflow, and groundwater recharge. The difference between the amount of water received and lost determines the amount of water stored within the basin.

$$\Delta S = P-E-Q-GW$$

Where ΔS is the change in storage, P is precipitation, Q is runoff, E is evaporation, and GW is net groundwater flows out of the catchment.

To apply the conservation of mass principle to a watershed, it is necessary to consider the physical variables that govern the hydrological cycle, including precipitation, surface runoff, groundwater infiltration, evaporation, and transpiration. By assuming that the watershed is a closed system, it is possible to use the conservation of mass principle to track the amount of water stored within the system.

In this study, we developed two approaches based on the conservation of mass principle, which are described in further detail below.

Approach 1:

In our first approach, we incorporated a physical constraint into the loss function of the LSTM model to ensure consistency with the conservation of mass principle. This principle dictates that the inflow to a system must be equal to the outflow plus any changes in storage within the system over a given time period. In a watershed, long-term storage refers to the total amount of water that is stored over an extended period of time, typically several months to several years. To enforce this constraint, we defined a condition such that the difference between the long-term simulated and observed storage changes must be less than or equal to 10% of the observed storage change. Simulated storage is calculated by subtracting the sum (over the entire

data length) of evaporation and simulated discharge in the training dataset from the total precipitation, whereas observed storage is determined by subtracting the sum of evaporation and actual observed discharge in the training dataset from the total precipitation. Due to the complexity and high variability of hydrological systems, a 10% difference between simulated and observed storage changes is considered acceptable in many applications (Safeeq et al., 2021). In addition, measurement errors and uncertainties in input data can also result in differences between simulated and observed results. The 10% difference implies that the model is able to reasonably approximate the observed storage changes. If this condition is met, the regular MSE loss function is used. However, if the condition is not met, the product of the absolute difference between the simulated and observed storage changes and a weight factor is added to the loss function to penalize the model.

Loss = Weight*(absolute difference between the simulated and observed change in storage) + MSE

This approach allows us to ensure that the LSTM model adheres to the principle of mass conservation in its predictions of streamflow.

Approach 2:

In the second approach, we introduced a physical constraint that limits the change in the simulated long-term storage to be less than 1% of total precipitation to ensure a water balance closure. The simulated storage is calculated by subtracting the sum of evaporation and simulated discharge in the training dataset from the total precipitation. A change in simulated storage of less than 1% of total precipitation is considered an acceptable level of precision for many hydrological applications. This is because precipitation is a major input in the water balance

equation, and small changes in precipitation can result in large changes in water storage. For example, if the total precipitation is 1000 mm, a change in simulated storage of less than 1% would be less than 10 mm. This small change in storage is a relatively small margin of error and suggests that the model is able to reasonably ensure a water balance closure.

If the condition is satisfied, the standard MSE loss function is applied. On the other hand, if the condition is not met, the loss function is modified to include a penalty term, which is calculated by multiplying the absolute change in simulated storage by a weight factor and adding it to the MSE loss. The resulting loss function can be expressed as:

$$\text{Loss} = \text{Weight}*(\text{absolute change in simulated storage}) + \text{MSE}$$

This approach utilizes the principle that the net change in storage for long time periods in a watershed tends to be close to zero and aims to enforce this constraint through the use of the penalty term in the loss function.

2.3.2.2 Energy balance (EB)

The principle of energy conservation dictates that energy cannot be created nor destroyed. In the context of a watershed, this implies that net radiation (R_n) (balance among incoming shortwave, outgoing shortwave, incoming longwave, and outgoing longwave radiations) either increases the temperature (sensible heat, H), or causes phase change (latent heat, LE), or is lost from the system. This can be represented as:

$$R_n = H + LE + Loss$$

where an example of $Loss$ would be the energy lost due to ground heat flux. The ground heat flux term tends to be insignificant over a 24-hour period, so it can be ignored in the loss function formulation. Thus, the energy balance equation becomes:

$$R_n - H = LE$$

In the water balance equation, the incoming component is precipitation, and the outgoing component is evaporation and streamflow. Evaporation serves as the link between the water and energy balance. To ensure consistency with the principle of energy conservation, we formulated a loss function equating the difference between net radiation and sensible heat to latent heat (total evaporation) with a tolerance of 1% of the net radiation sum. We assumed a tolerance of 1% of the sum of net radiation because energy balance is complex and is dependent on many factors such as temperature, pressure, radiation, and small changes in these factors can result in large changes in the energy balance. Therefore, a small change in the energy balance suggests that the model is able to accurately capture the energy balance of the system. Additionally, a 1% tolerance allows some of the model uncertainties and is neither overly nor under conservative.

If this condition is met, the normal MSE loss function is used. However, if the condition is not met, the model is penalized by taking the absolute difference in incoming and outgoing energy, multiplying it by a weight, and adding it to the MSE. The resulting loss function is as follows:

Loss = Weight*(absolute difference in energy) + MSE

2.3.2.3 Storage Discharge (SD)

This constraint is based on the storage-discharge relationship, which is given as:

$$S = KQ$$

where S is the storage, Q is the discharge, and K is the storage-time constant. The calculation of K for each time step involves determining the ratio of the simulated and observed storage values to the corresponding simulated and observed discharge values. The final K value for the entire watershed is obtained by averaging the K values across all time steps. The difference between the average simulated K (K_{sim}) and the average observed K (K_{obs}) is calculated and is then divided by the K_{obs}, to capture the relative change in the ratio. To incorporate this physical constraint into the loss function of our model, we added a condition such that the ratio should be less than or equal to 1%. A small difference in the storage time constant suggests that the model can accurately capture the dynamics of the system and simulate the observed storage changes.

If this condition is met, the normal MSE loss function is used. Otherwise, we penalized the model by taking the absolute change in this ratio, multiplying it by a weight, and adding the resulting value to the MSE loss. The modified loss function can be expressed as:

$$\text{Loss} = \text{Weight}*(K_{sim} - K_{obs}) / K_{obs} + \text{MSE}$$

2.3.3 Evaluation Metrics

The Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) is a measure of the accuracy and efficiency of a model in predicting variables that differ from the mean. It represents the proportion of the original variance accounted for by the model, and is calculated as follows:

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_m^t - Q_o^t)}{\sum_{t=1}^{T}(Q_o^t - \overline{Q_o})}$$

Where Q_o^t and Q_m^t are observed and modeled discharges at time t. $\overline{Q_o}$ is the mean of observed discharges. The NSE can range from -∞ to 1, with a value of 1 indicating perfect correspondence between simulation and observation, a value of 0 indicating that the model simulation has the same explanatory power as the mean of the observations, and a value of less than 0 indicating that the model is a worse predictor than the mean of the observations.

2.4 Results and Discussion

In this study, the predictability performance of standalone LSTM models and LSTM models with the addition of physics-based constraints in the loss function was evaluated for one-day-ahead streamflow prediction at 34 basins. The model predictions were used to evaluate the model's ability to simulate streamflow data during the testing phase.

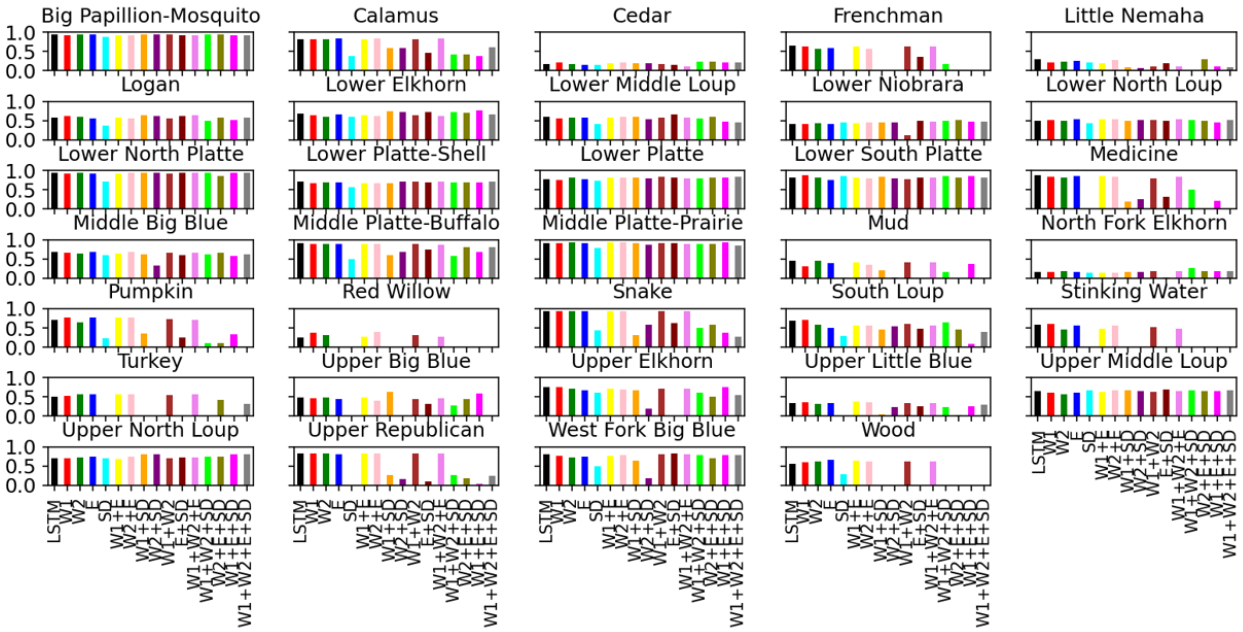

Figure 2.3 Comparison of standalone LSTM with LSTM including physics-based constraints and their combination. Y-axis shows each basin NSE value. X-axis shows models. A negative NSE value is not included in the plot.

Figure 2.3 shows the prediction performance of standalone LSTM models and LSTM models with physics-based constraints in the loss function. For simplicity, LSTM is represented as LSTM, Water balance Approach 1 is represented as W1, Water balance Approach 2 is represented as W2, Energy balance is represented as E, and Storage Discharge is represented as SD. The prediction efficiency of standalone LSTM models at all basins was found to be between $0.16 \leq \text{NSE} \geq 0.95$. Out of the 34 basins, nine basins had low prediction efficiency for streamflow, with an NSE value of less than 0.5.

To examine the effect of physics-based constraints on prediction accuracy, we developed 15 LSTM models, each incorporating a single physical constraint or a combination of physics-based constraints. The impact of adding physics-based constraints is seen to be minimal where there is already good performance from the LSTM model. For example, in Big Papillion-

Mosquito, the NSE is 0.95 from the LSTM model, indicating less room for improvement in prediction.

The results of our analysis showed that the impact of physics-based constraints on prediction accuracy varied among the different basins studied (Figure 2.3). In some cases, the incorporation of physics-based constraints resulted in an improvement in prediction accuracy, while in other cases it had no effect or even decreased prediction accuracy. For example, the incorporation of the W1 constraint improved prediction accuracy for the Red Willow basin by 43% but had no effect on the Calamus basin. In contrast, the imposition of the W2 constraint improved prediction accuracy for the Red Willow basin by 18% but decreased prediction accuracy by 15% for the Little Nemaha basin and had no effect on the Calamus basin. The incorporation of the E constraint decreased prediction accuracy for the Red Willow and Little Nemaha basins by 162% and 14%, respectively. However, the incorporation of the E constraint improved prediction accuracy for the West Fork Big Blue basin by 20%. Additionally, the imposition of the SD constraint improved prediction accuracy in some basins, such as Lower Niobrara, but had no effect on the majority of the basins studied.

In our analysis of 34 basins, the incorporation of the W1 constraint improved prediction efficiency for 38% of the basins, with an average increase in efficiency of 9%. Additionally, the combination of W1 with E and W2 improved prediction efficiency for 41% and 38% of the basins, respectively, with average increases in efficiency of 7% and 6%. The incorporation of the W2 constraint improved prediction efficiency for 32% of the basins, with an average increase in efficiency of 7%. The combination of W2 with E and W1 increased the percentage of basins with improved prediction efficiency from 32% to 41% for both combinations, with average increases in efficiency of 9% and 6%, respectively.

It was observed that for a subset of catchments, the imposition of constraints on the LSTM model resulted in minimal or no effect on model efficiency. Specifically, in 15% of catchments, the imposition of W2, in combination with W1 and E constraints, did not yield a statistically significant change in efficiency as measured by the NSE metric. Furthermore, in 9% of catchments, the imposition of W1, in combination with W2 and E constraints, W1 and SD constraints, W1 and W2 constraints, and W2, E, and SD constraints, resulted in a similar lack of change in NSE. Notably, the imposition of SD constraints, as well as combinations of SD constraints with W2, resulted in zero percent catchment with no effect.

The incorporation of the SD constraint improved prediction efficiency for 12% of the basins, with an average increase in efficiency of 5%. The combination of SD with W1, W2, and E increased the number of basins with improved prediction efficiency, with the percentage increasing from 12% to 26% for combinations with W1 and W2, and 29% for combinations with E.

Overall, these results suggest that multiple physics-based constraints can be used separately or in combinations depending on the characteristics and behaviors of the particular catchment, in order to improve prediction efficiency (Figure 2.4).

Our results showed that the incorporation of the SD constraint decreased prediction efficiency for 88% of the basins, with an average decrease in efficiency of 60%. Additionally, the combination of SD with other physics-based constraints was found to decrease the efficiency of those constraints in improving prediction accuracy. For example, in the Cedar basin, the incorporation of the W1 constraint alone improved prediction efficiency by 22%. However, when combined with the SD constraint, the improvement in efficiency was reduced to only 8%, indicating a 14% decrease.

These findings suggest that caution should be exercised when implementing the SD constraint in streamflow prediction models.

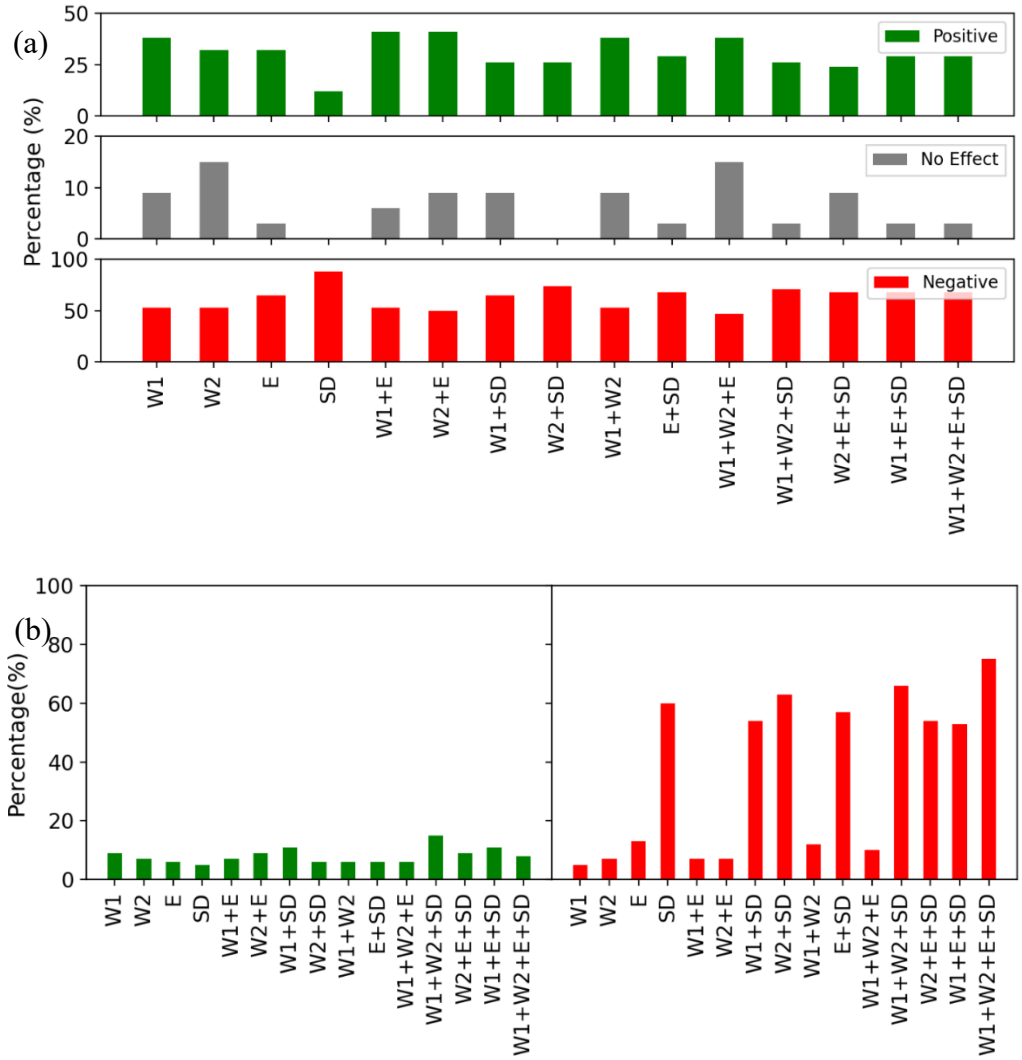

Figure 2.4 Percentage of catchments where the performance improved/fall after imposing different constraints and their combinations. Green, red, and gray bars show the percentage of catchment where the performance improved, fall, and has no effect after imposing different constraints and their combinations. (b) Green and red bars show the average percentage improvement and fall after imposing different constraints and their combinations for catchments.

2.5 Conclusions

In this study, we investigated the effects of incorporating physics-based constraints, including water balance, energy balance, and storage-discharge relationship, into LSTM for predicting daily streamflow in 34 HUC-8 basins in Nebraska. The performance of the model was thoroughly assessed for cases with and without the inclusion of these constraints and their combinations.

Our results showed that the inclusion of certain physics-based constraints can lead to improved performance of the LSTM model in streamflow prediction. In particular, the use of water balance constraints was found to have the most significant positive impact on model performance with an increase in efficiency for 38% of catchments. Additionally, the combination of water balance (W1) and energy balance constraints resulted in the greatest improvement among all constraint combinations examined with an increase in efficiency from 38% to 41% of catchments. On the other hand, the storage discharge constraint and its combinations displayed less consistent results and may be more relevant in catchments where flow is dominated by catchment characteristics such as topography, soil types, land use, and hydrography. These characteristics can affect the hydrological response of a catchment area, such as the amount and timing of streamflow, and are important to consider when modeling and predicting streamflow in a catchment. It was also noted that the specific method of imposing constraints can affect model performance. For example, W1 improves efficiency for 38% of the catchment, while W2 improves efficiency for 32%. These findings highlight the potential value of using physics-based constraints in machine learning models for streamflow prediction, particularly in data-scarce regions.

These findings suggest that the application of constraints on LSTM models may not always result in a significant improvement across all catchments and that further research is needed to understand the conditions under which constraints are most effective. Additionally, the results indicate that the specific constraints combination and their effect may vary depending on the catchment. While it may not always result in a significant improvement in the overall performance, the incorporation of physics-based constraints will enhance the model's ability to capture underlying physical processes that govern streamflow, which can potentially lead to the improved generalization capability of the model. Thus, physics-based constraints can increase the robustness of the model to unseen data and make it more representative of real-world scenarios. This is an important step towards physically relevant streamflow prediction by infusing physical laws into the LSTM model via a simple but powerful alteration to the loss function. This approach provides a good balance between capturing the watershed-specific uniqueness through the training of LSTM models for each catchment and incorporating generalizable knowledge across all catchments through the use of constraints (Beven, 2014; Troch et al., 2015).

In practice, it is important to first test the efficacy of constraints for the problem at hand. Individual constraints and their combinations need to be thoroughly assessed before being used. It is important to approach this p rocess with a mindset of continuous evaluation and improvement because the effectiveness of constraining can change depending on the variables under consideration, datasets used, unique and generalizable characteristics of the catchments, among others.

Overall, this study provides valuable insights into the use of LSTM models for streamflow prediction and the potential benefits and limitations of imposing constraints on their

loss function. Additionally, this study highlights the importance of considering both the unique and generalizable characteristics of a catchment when imposing constraints on the LSTM model. While this study has focused on the use of these specific physics-based constraints, it is important to note that other forms of physical relationships can also be explored in LSTM and other machine-learning models for streamflow prediction. Finally, it would be beneficial to dig deeper into the factors influencing the success or failure of different constraints and combinations of constraints to predict flow in different catchments. This type of research can help identify common characteristics that affect the effectiveness of these constraints and their combinations in streamflow prediction. This can open new possibilities for improving the performance and robustness of these models and make them more useful for predictions and decision-making.

Chapter 3 Peak Flow Estimation Using Machine Learning

The information relayed in this chapter is from a published paper: Pokharel, S., T. Roy, and D. Admiraal, Machine Learning-Based Peak Flow Estimation for Nebraska Streams, under review in *Journal of Hydrology*.

Accurate estimation of peak flow is vital for the design of infrastructure and for ensuring public safety. Regional regression equations are widely used to estimate peak flow in Nebraska streams; however, equations from multiple studies often yield results that vary by orders of magnitude, casting doubt on their reliability. This highlights the need for a more advanced and unified approach to improving the overall accuracy of peak flow estimation. In this study, we modeled peak flow using a suite of Random Forest models for different return periods in 40 Nebraska streams using high-resolution datasets covering atmospheric, hydrologic, soil, and basin morphometric characteristics. The newly developed system outperformed the existing regression equations which were previously fitted to an older dataset for all return periods considered. This was evident when the average Kling-Gupta Efficiency (KGE) medians from both systems were compared for 100 randomly selected training-testing basin sets. The newly developed system exhibits an average median KGE value of 0.15 across all return periods considered while in contrast, the regression equations yielded an average median KGE value of -0.30. The new system can aid in better decision-making related to the mitigation of flood disasters and the design of resilient infrastructure.

3.1 Introduction

Floods are the most frequent, widespread hazards that significantly impact infrastructure, human safety, and livelihoods (Desai and Ouarda, 2021; Doocy et al., 2013; Hirabayashi et al., 2013; Li et al., 2022; Merz et al., 2021; Rasheed et al., 2022; Tanoue et al., 2016). According to the National Centers for Environmental Information (NCEI) (2023), between 2021 and 2022, the

cost of flood-related damages rose from $155.3 billion to $165 billion. These rising costs parallel an increase in the occurrence and severity of extreme flooding events observed in many global regions over recent decades (Berghuijs et al., 2017; Blöschl et al., 2019; Marijnissen et al., 2019; Prieto et al., 2020). The state of Nebraska has experienced multiple catastrophic flooding events in the recent past, notably the great floods of 1993 and 2019 (NWS, 2022). In 2019, rapid melting of record snowfall accumulation between January and March caused much of the catastrophic flooding in Nebraska and Iowa, with an estimated $3 billion in damages and three lives lost (Schwartz, 2019). The rain-on-snow (ROS) event in that spring, combined with the rapid melting of snowpack, was what caused this catastrophic flooding. Rapid snowmelt and ROS events are frequently to blame for large floods in the Midwest (Rasiya et al., 2023; Velásquez et al., 2023; Villarini et al., 2011). According to Flanagan et al. (2020), the flood damaged roads, bridges, and rail systems, amongst other modes of transportation.

Peak flow and flooding are closely related because peak flow describes the magnitude and extent of the flooding. Information about peak flow is required to accurately estimate flood damage, which is a crucial consideration in flood response and water management decisions (Beard, 1997; Kienzler et al., 2015). Additionally, peak flow rates are used in designing infrastructure such as bridges, culverts, dams, and levees on or near streams (Ding and Haberlandt, 2017). In Nebraska, USGS and others have developed various methods for computing the peak flow rates for different return periods. Furness (1955) developed an equation to calculate peak flow for basins with a minimum drainage area of 100 mi^2 (259 km^2) over a return period of up to 50 years. The state of Nebraska was divided into two regions based on soil properties. The equation used to calculate the peak flow was primarily dependent on drainage area, while all other factors influencing the peak flow were lumped into a coefficient C.

$$Q_{2.33} = CA^{0.7}$$

where, $Q_{2.33}$ is the mean annual peak flow (cfs) to the 2.33 year recurrence interval, C is the peak flow coefficient, and A is the drainage area (mi^2). Hutchison (1962) developed methods for calculating peak flow for uncontrolled and unregulated streams with drainage areas of no more than 300 mi^2 (777 km^2). The process involves three steps: determining the total and contributing drainage areas and the flood region, computing the average annual flood for the location, and using a ratio to relate the mean annual peak flow to the desired recurrence interval. The accuracy of flood magnitudes for selected recurrence intervals was found to be dependent on the number of stations used and the length of each record. Beyond 25 years, the regional frequency curves cannot be extrapolated with confidence. Furthermore, the mean annual flood curves for the drainage area cannot be extended beyond the limit, as per this method.

Patterson (1966) developed frequency curves to estimate the size and frequency of floods with recurrence intervals between 1.1 and 50 years for most streams in the Missouri River basin above Sioux City, Iowa. The relationship between the mean annual flood and floods with different recurrence intervals was represented by the frequency curves, which predicted peak discharge using a dimensionless frequency curve and a basin characteristic relation. However, it was not possible to use the composite frequency curves to extrapolate beyond a recurrence interval of 50 years. Like Patterson (1966), Matthai (1968) also created frequency curves to calculate peak flow magnitudes between 1.1 and 50 years. Beckman (1976) developed a method to estimate floods with recurrence intervals of up to 100 years. Nebraska was divided into five hydrological regions using regression techniques based on soil types for regions 1 and 2 and basin divides for regions 3 and 5. Flood magnitude and frequency were determined for major

controlled streams like the North Platte, South Platte, Platte, and Republican Rivers, using peak flow data from 303 gauge stations with 13 or more years of data. Log-Pearson Type III (LP3) was implemented for station data analysis. The regression equation for the five regions included factors such as drainage area, mean annual precipitation, main stream length, main stream slope, maximum 24-hour 50-year rainfall, and temperature. However, the equation does not apply to regulated or urbanized streams and cannot be used for streams in adjacent states. The standard error of estimate ranged from 22% in the Big Blue region to 60 to 102% in the western part, but the equation was later revised, improving the peak flow estimation. Soenksen et al. (1999) presented the updated peak flow frequency analysis for selected streamflow gaging stations in Nebraska. The scope of this study (Soenksen et al. 1999) was to update the old regression equations developed by Beckman (1976), create a new set of equations, and evaluate those equations. Geographic Information System (GIS) and digital spatial data were used to quantify the drainage basin characteristics. Peak flow frequency analyses were done for unregulated streamflow-gaging stations with at least ten years of annual peak-flow record through 1993 and located in or within about 50 miles of Nebraska using the LP3 frequency distribution and the guidelines in Bulletin 17B of the Interagency Advisory Committee on Water Data (ISCWD, 1982). For seven regions, eight regression equations were developed. The standard error ranged from 12 to 64%.

Strahm and Admiraal (2005) used GIS and data from 273 gauge stations to develop regional regression equations for estimating peak flow magnitude for the seven hydrologic regions of Nebraska identified by Soenksen et al. (1999). The weighted least squares method was used for the regression analysis and sets of equations were developed to estimate peak flow rates with a recurrence interval of 2 to 500 years. The regional regression helped in estimating peak

flow in ungauged basins, which is a big challenge in the field of hydrology (Sivapalan et al., 2003). The equations were created using nearby watershed basins with comparable hydrologic properties. Even though these equations incorporated more detailed information about the hydroclimatology and basin characteristics compared to their older counterparts, they are two decades old and do not account for changes that have occurred in recent decades.

Regionalization plays a crucial role in addressing data scarcity issues in locations without adequate hydrological measurements (Buytaert and Beven, 2009; Song et al., 2022). To date, many studies have explored various approaches to regionalization, aiming to provide accurate estimation in ungauged regions. Studies by Hrachowitz et al. (2013) and Razavi and Coulibaly (2012) have significantly contributed to the discourse on regionalization in the context of flood frequency estimation. These studies have highlighted the challenges of extrapolating hydrological relationships across different climatic and geographic settings and pointed out the need for flexible methodologies that account for the complexities of diverse regions. Early studies Ouarda et al. (2006) and Shu and Ouarda (2008) highlighted the importance of considering regional context when estimating flood frequencies. Recent work by Brunner et al. (2018) and Han et al. (2022) have emphasized the dynamic nature of hydrological processes and the necessity of adaptive techniques to capture these dynamics.

Our study takes a distinctive regionalization approach by considering a machine learning model to capture peak flow for different return periods from a wide range of hydroclimatic and basin morphometric factors. We employed random forest, which has been widely used in water resources applications (Tyralis et al., 2019). In contrast to the traditional regionalization methods that partition a large region based on predefined characteristics and employ different sets of

regression equations in each part, we built a single model that used data from across different locations within the state.

The remainder of this paper is structured as follows: Section 3.2 describes the study area and datasets utilized for the current analysis. Section 3.3 details the methodological procedures carried out, including the random forest modeling approach. Section 2.4 presents the results obtained from the random forest model and regression model, as well as a discussion of key findings. Section 2.5 summarizes the conclusions and implications of this study.

3.2 Study area and data

3.2.1 Study area

The study area encompasses 40 Hydrologic Unit Code 8 (HUC-8) basins located within the state of Nebraska, characterized by diverse physical and geographical features. The region includes a range of landscapes, such as the Great Plains, the Sand Hills, and the Ogallala Aquifer. The selection of these basins was based on the availability of gauge stations and the length of discharge data. We only included the basins that have at least 20 years of discharge data. Basin areas ranged from 1,366 to 13,352 km^2, with each exhibiting unique hydrological characteristics, such as different soil types, land-use patterns, and water availability. The locations of the selected HUC-8 basins within Nebraska are shown in Figure 3.1.

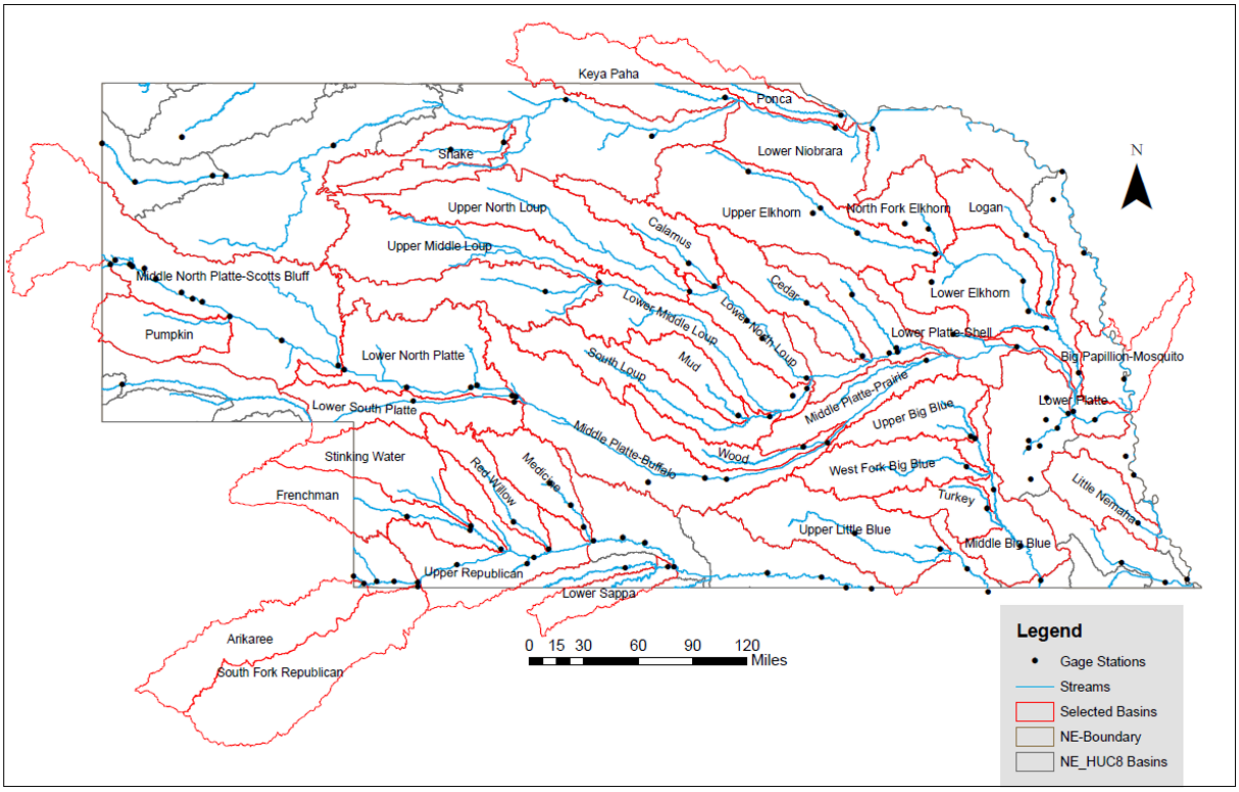

Figure 3.1 HUC-8 basins in Nebraska. Basins demarcated with red are the ones used in this study.

3.2.2 Data

The datasets (Table 3.1) used in this study were collected from three different sources. The atmospheric and land data was obtained from the ERA5-Land (9 km grid) and ERA5 (30 km grid) (Hersbach et al., 2020; Muñoz-Sabater et al., 2021), which is the fifth-generation atmospheric reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF). Streamflow data were collected from the United States Geological Survey (USGS) and the Nebraska Department of Natural Resources (NeDNR). The mean, mode, and quantiles from 10^{th} to 90^{th} for each of the dynamic variables (Table 3.1) were calculated and used in the analysis. Descriptive statistics and quantiles of input variables are used to predict streamflow quantiles because they provide useful information about the distributional properties of the

variables. We also included basin morphometric and soil characteristics data such as area, basin slope, main channel slope, relative relief, shape factor, compactness ratio, elongation ratio, rotundity of the basin, average minimum permeability, average permeability, and average maximum soil slope. All of these datasets were generated using GIS techniques.

Table 3.1 Description of the datasets.

Variable Name	**Units**	**Source**	**Classification**
10m u Wind Component	$m\ s^{-1}$	ERA5-Land	Dynamic
10m v Wind Component	$m\ s^{-1}$	ERA5-Land	Dynamic
2m Dew point Temperature	K	ERA5-Land	Dynamic
2m Temperature	K	ERA5-Land	Dynamic
Convective available potential energy	$J\ kg^{-1}$	ERA5	Dynamic
Total Evaporation	m	ERA5-Land	Dynamic
Leaf Area Index, High Vegetation	$m^2\ m^{-2}$	ERA5-Land	Dynamic
Leaf Area Index, Low Vegetation	$m^2\ m^{-2}$	ERA5-Land	Dynamic
Skin Temperature	K	ERA5-Land	Dynamic
Snow Albedo	(0-1)	ERA5-Land	Dynamic
Snow depth	m	ERA5-Land	Dynamic
Soil Temperature layer 1	K	ERA5-Land	Dynamic
Surface Net Solar Radiation	$J\ m^{-2}$	ERA5-Land	Dynamic
Surface Net Thermal Radiation	$J\ m^{-2}$	ERA5-Land	Dynamic
Surface Pressure	Pa	ERA5-Land	Dynamic
Surface Sensible Heat Flux	$J\ m^{-2}$	ERA5-Land	Dynamic
Total Column Water	$kg\ m^{-2}$	ERA5	Dynamic
Total Precipitation	m	ERA5-Land	Dynamic
Volumetric Soil Water Layer 1	$m^3\ m^{-3}$	ERA5-Land	Dynamic
Streamflow	Cfs	USGS/NeDNR	Dynamic
Basin Area	mi^2	GIS	Static
Basin Slope	$ft\ mi^{-1}$	GIS	Static
Main Channel Slope	$ft\ mi^{-1}$	GIS	Static
Relative Relief	$ft\ mi^{-1}$	GIS	Static
Shape Factor	Dimensionless	GIS	Static
Compactness Ratio	Dimensionless	GIS	Static
Elongation Ratio	Dimensionless	GIS	Static
Rotundity	Dimensionless	GIS	Static
Average Permeability	$in\ hr^{-1}$	GIS	Static
Average Maximum Soil Slope	%	GIS	Static

3.3 Methodology

3.3.1 Data Preprocessing

The calculation of the return periods was carried out according to the guidelines outlined in the US Interagency Committee on Water Data Bulletin 17B (ISCWD, 1982). The annual maximum discharge values for each of the selected catchments were fitted to an LP3 distribution based on the method of moments (Singh, 1998). These distributions were then used to determine peak flows for the return periods of 2, 5, 10, 25, 50, 100, 200, and 500 years.

As can be seen from Table 3.1, our predictors included both static and dynamic variables. Examples of static variables would be basin slope, basin area, etc., while precipitation, evapotranspiration, etc., will fall under the category of dynamic variables. For each of the dynamic variables, we included their mean, mode, and nine quantiles (10^{th} to 90^{th} with an interval of 10) derived from the time series. The predictors considered in this analysis can either increase or decrease streamflow. For example, as evaporation increases, streamflow will decrease. On the other hand, streamflow will increase with an increase in precipitation. Therefore, in some cases, the lower values of a predictor could be more closely related to the higher values of streamflow. To incorporate this countering effect, we first assessed the direction of the change in streamflow consequent to a change in the predictor value and accordingly calculated the quantiles either of their annual minimum or maximum values.

3.3.2 Random Forest

Random Forest (RF) (Breiman, 2001a) is a powerful and versatile machine learning (ML) algorithm widely applied for supervised learning tasks (regression and classification). It is an ensemble method that utilizes multiple decision trees to generate accurate and robust predictions. The primary goal of RF is to decrease the prediction variance by averaging numerous decision

trees that are trained on diverse subsets of the same training data (Ghimire et al., 2023; Hastie et al., 2009). The algorithm is particularly suitable for modeling non-linear and complex variable interactions and is less susceptible to multicollinearity in the data, especially for prediction problems (Tyralis et al., 2019). It should be noted that multicollinearity can still influence the interpretation of variable importance. Figure 3.2 shows the proposed methodology for this study.

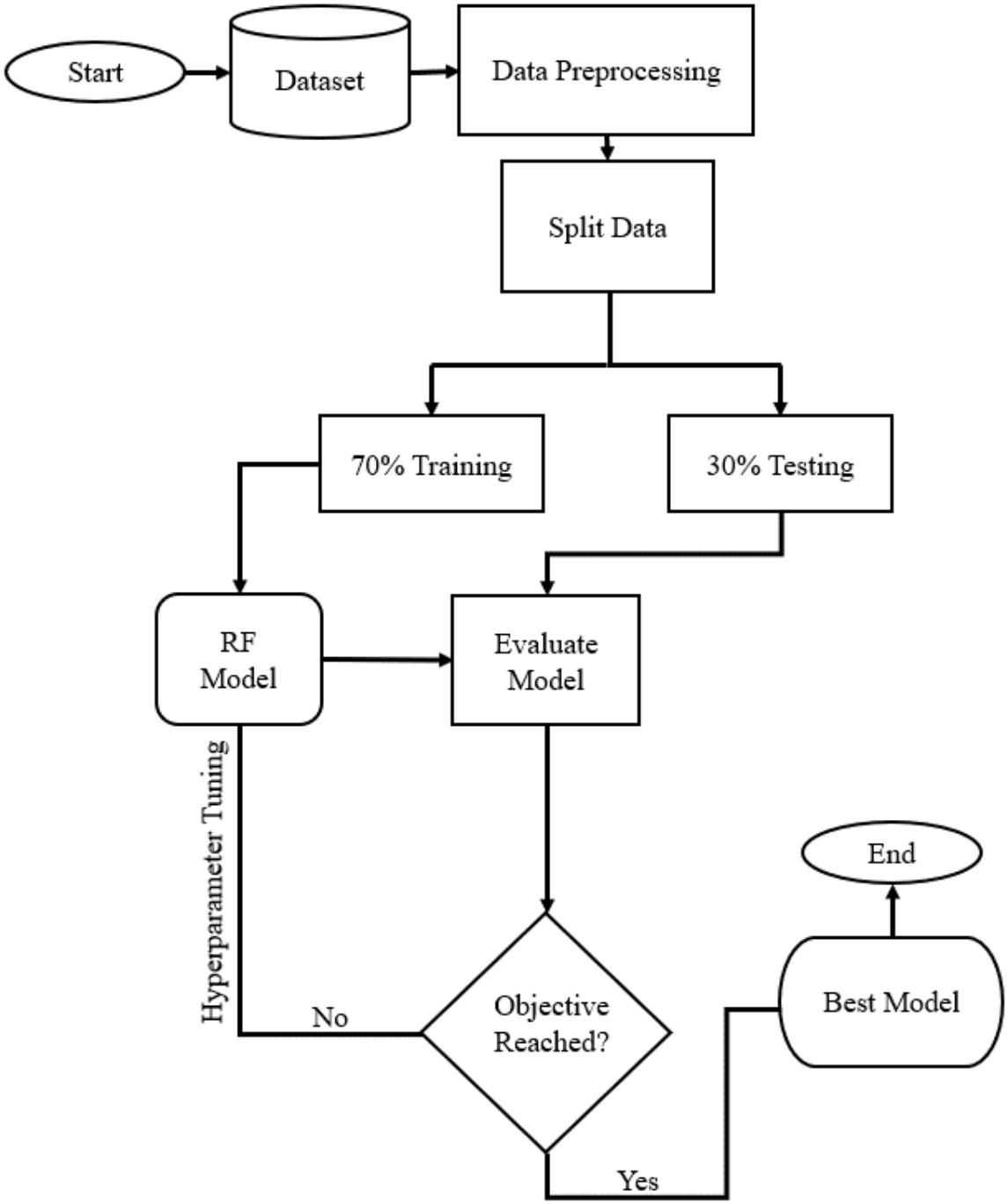

Figure 3.2 Workflow schematic of the proposed RF model.

In RF, a collection of multiple decision trees forms a random forest, with each of these trees learning from a subset of the original features. The key concept behind this approach is that

each individual tree is a weak learner (owing to exposure to partial data), while the entire forest can provide precise outcomes while addressing high variance (Figure 3.3), making it less susceptible to overfitting compared to other machine learning algorithms. For a more comprehensive mathematical formulation and information on RF, refer to Breiman, (2001a, 2001b) and Liaw & Wiener (2002)

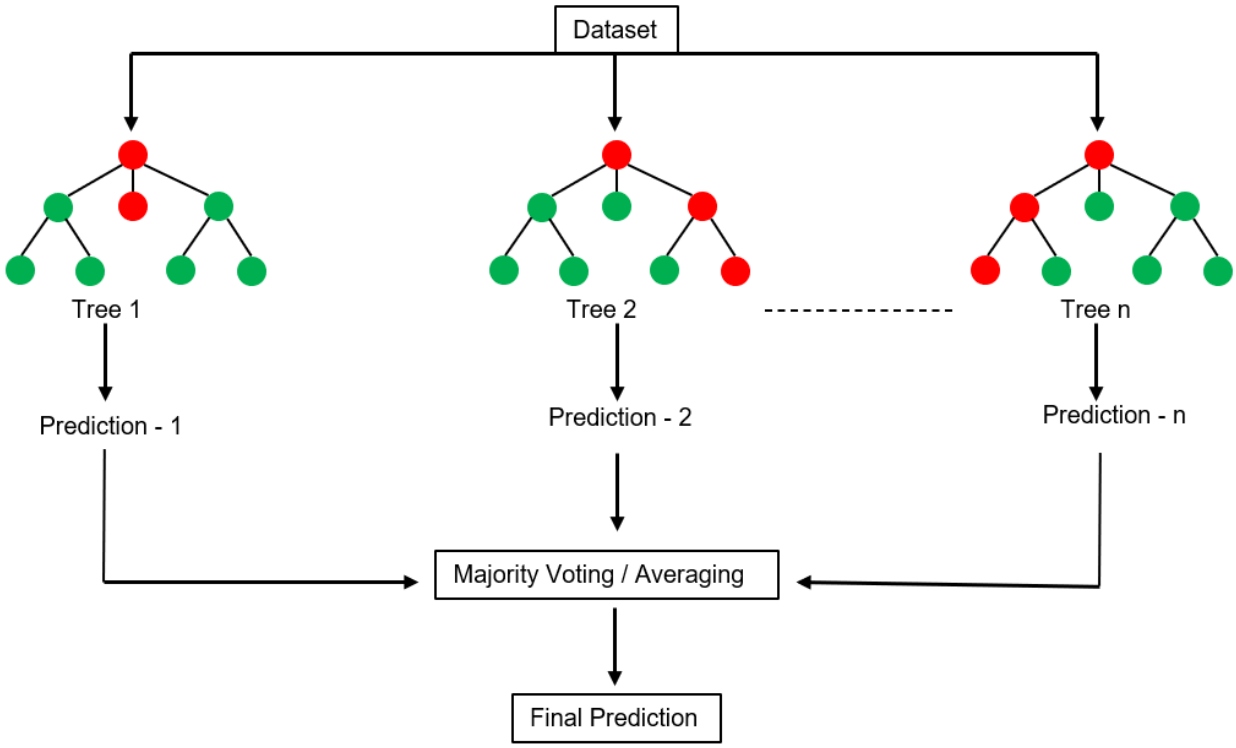

Figure 3.3 Graphical illustration of an RF architecture.

The efficiency of RF in learning from the data and making predictions largely depends on the appropriate selection of hyperparameters, which control the training process in the ML models. Hyperparameter tuning can be accomplished either manually or through automated methods such as grid search or random search, as explored in prior research (Belete & Huchaiah, 2021; Wu et al., 2019). Regardless of the chosen method, a thorough consideration of the impact

of each hyperparameter on the model's accuracy is important, and a range of hyperparameters should be tested to derive the optimal set. The number of trees in the forest is one of the crucial hyperparameters in RF, which can enhance the model's accuracy but can also lead to increased computational complexity and training time. The maximum depth of the trees, which regulates the model's complexity, is another significant hyperparameter. A shallower tree may cause underfitting, while a deeper tree, on the other hand, can lead to overfitting. Overfitting happens when an ML model performs very well on the data it was trained on but fails to generalize to new unseen data (Pokharel et al., 2023). Another important hyperparameter in RF is the minimum sample leaf, which represents the minimum number of data points allowed in a leaf node.

In this study, RF models were constructed for each return period's peak discharge, and the hyperparameters were manually tuned, with the performance assessed using the Kling-Gupta Efficiency (KGE) Gupta et al., 2009). KGE integrates three common types of errors in a model, i.e., bias, variability, and timing (correlation), and is written as:

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}$$

where r is the linear correlation coefficient between the observations and simulations. γ is the ratio of the coefficients of variation, and β is the ratio of the means. KGE can range from -∞ to 1, with a value of 1 indicating perfect correspondence between simulations and observations. The final used hyperparameters to estimate the peak flow across different return periods are shown in Table 3.2.

Table 3.2 Hyperparameters used in RF models across different return periods.

Return Period	Number of trees	Maximum depth	Minimum sample leaf
T2	10	4	Default
T5	1000	4	5
T10	1000	4	5
T25	1000	4	5
T50	1000	4	5
T100	1000	4	2
T200	1000	4	2
T500	1000	4	2

We implemented separate RF models for all return periods considered, and the data in each case was split into training (70%) and testing (30%) sets. For each return period, the number of total data points was 40, which is equal to the total number of catchments. Since the RF model is a tree-based algorithm, there was no need for feature scaling. We also randomly sampled the training and testing sets a hundred times to identify the best-performing model with the highest KGE. Moreover, during each iteration (total 100), we compared the predictions from the recent regression equations (developed by Strahm and Admiraal (2005) using older datasets) with the observed peak streamflow and computed the corresponding KGEs. These KGE values were subsequently compared against the KGE values obtained from the RF model in each case.

3.4 Results and Discussion

In this section, we first present the best-performing RF model for estimating peak flow for each return period. We then compare the performance of our RF-based system with the existing regression equation-based system (Strahm and Admiraal, 2005) for seven hydrological regions to estimate peak flow in Nebraska. Note that our goal was not to compare RF with regression equations since it is well established that RF would demonstrate superior performance in this case (e.g., Breiman, 2001; Prieto et al., 2020). This is why we did not recreate the

regression equations with the same data used in the RF models. Our comparison instead is between two different systems, the existing one that is widely used in practice and a new one that has been developed in this study using high resolution datasets and RF. The new system allows us to produce more accurate estimates of peak flow, and therefore, has the potential to better inform practice.

3.4.1 Performance of RF model

In this study, the RF models were developed for peak discharge corresponding to each return period, with the optimal model obtained through 100 iterations with randomly selected basins for the train-test splits. The performance of the RF model for predicting peak streamflow was evaluated using the KGE metric for different return periods ranging from 2 to 500 years (Figure 3.4 and Figure 3.5). The KGE values for the RF model ranged from 0.71 to 0.81, with a median KGE of 0.76 across different return periods (Figure 3.4). Figure 3.5 displays the relationship between target and predicted values for eight return periods ranging from 2 to 500 years. Each subplot contains a 1:1 line representing perfect model agreement. The close alignment of the data points along the 1:1 line illustrates the strong correlation between targets and predictions across all return intervals. The highest KGE value was obtained for the 2-year return period peak discharge, while the lowest was obtained for the 5-year return period peak discharge. For return periods ranging from 10 to 50 years, KGE values ranged from 0.75 to 0.79. Longer return periods of 100 to 500 years yielded KGE values ranging from 0.72 to 0.74. These findings indicate that the RF models have consistently good predictive performance across all return periods.

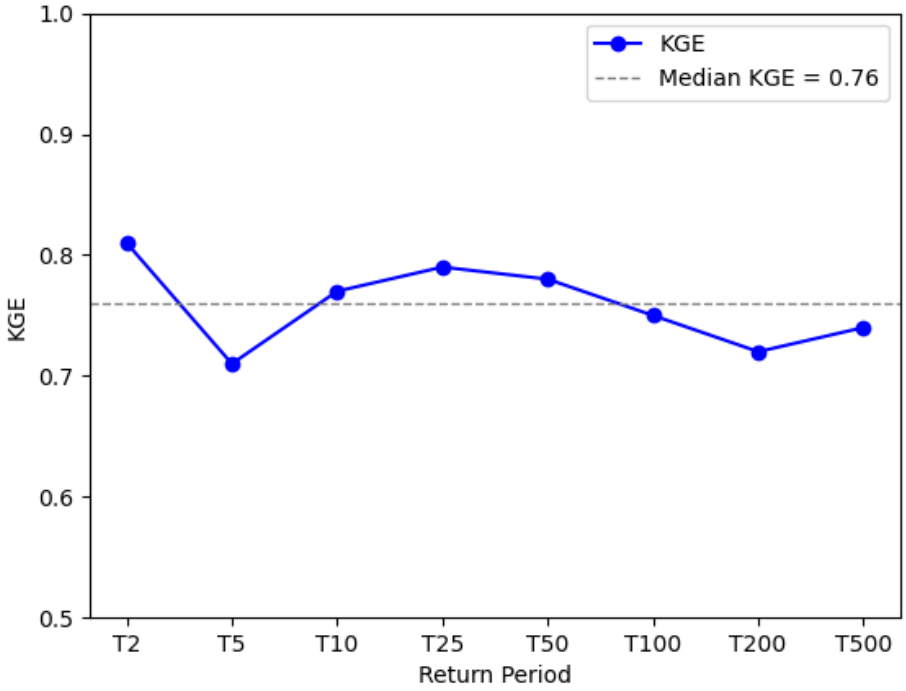

Figure 3.4 KGE values for different return periods obtained from the RF model.

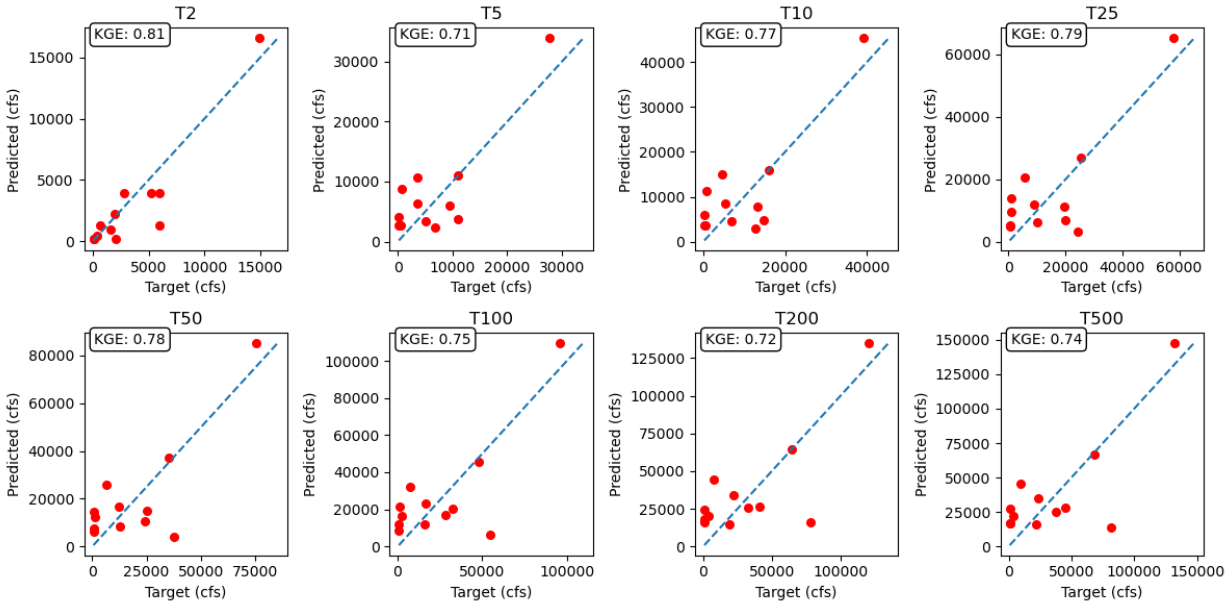

Figure 3.5 Scatter plot of best performing RF model predictions across different return periods. X-axis represents the target peak discharge and Y-axis represents the predicted peak discharge for different return periods, as indicated in the subplots.

3.4.2 Comparison of RF performance with regression equations

We compared the performance of RF models to an existing system of regional regression equations developed by Strahm and Admiraal (2005) for estimating peak flow for different return periods in Nebraska. To accomplish this, 100 iterations of the RF model were performed for each return period, and the KGE scores for all testing basins were recorded. Furthermore, for those 100 runs, the existing regional regression equations were also used to compute the peak flow discharge for different return periods in the testing basins. The observed and predicted peak flow discharge data were analyzed, and the KGE scores for all the testing basins obtained from the 100 iterations were computed. It is important to note that the comparison involved evaluating the outcomes of two distinct systems: A regression model fitted using older datasets, and the RF model trained with recent datasets. The focus of the comparison was to evaluate the efficacy of the newly developed RF model with existing regression equations. Figure 3.6 shows the boxplot distribution of KGE values obtained for all test runs.

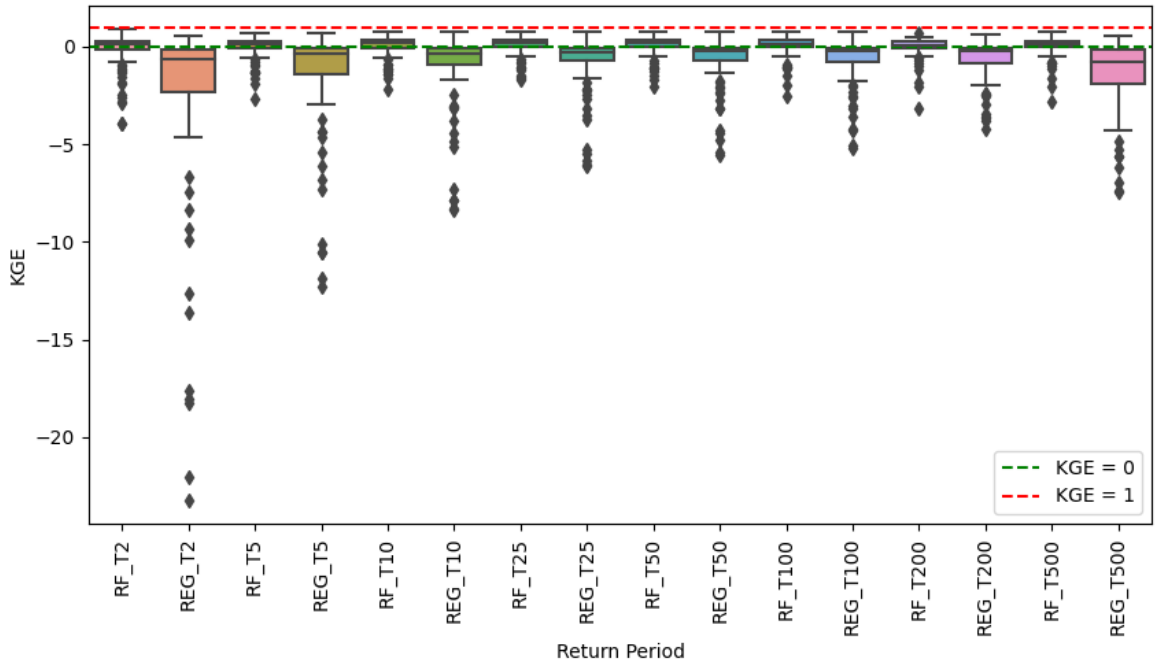

Figure 3.6 Boxplot showing the KGE values obtained from the 100 runs of the RF model and regression equations across different return periods. The red dotted line represents the upper limit of KGE (i.e., 1), whereas the green dotted line represents the zero value. Zero value is highlighted just to compare the median KGE values. RF_Tx represents the Random Forest model for x-year return period. REG_Tx represents the existing regression equation developed by Strahm and Admiraal, (2005) for x-year return period.

In Figure 3.7, the median KGE values are plotted, which were derived from 100 test runs for both RF models and regression equations, corresponding to each return period. It is important to note that Figure 3.4 shows the KGE value of the RF model that exhibited the highest performance. Conversely, Figure 3.6 illustrates the median KGE values obtained from 100 test runs for each return period. As can be seen from Figure 3.6 and Figure 3.7 for all return periods, the RF model outperformed the regression equations, with the median KGE value for all return period peak flow above zero, while the regression equations median KGE value was below zero. The average median KGE value across all return periods considered for the RF model was 0.15 while in contrast, the regression equations yielded an average median KGE value of -0.30. Also,

it can be seen that for the 500-year return period, the median KGE drops sharply, indicating low performance of the regression equations to capture the peak flow for high return periods. The RF models, on the other hand, performed consistently across all return periods.

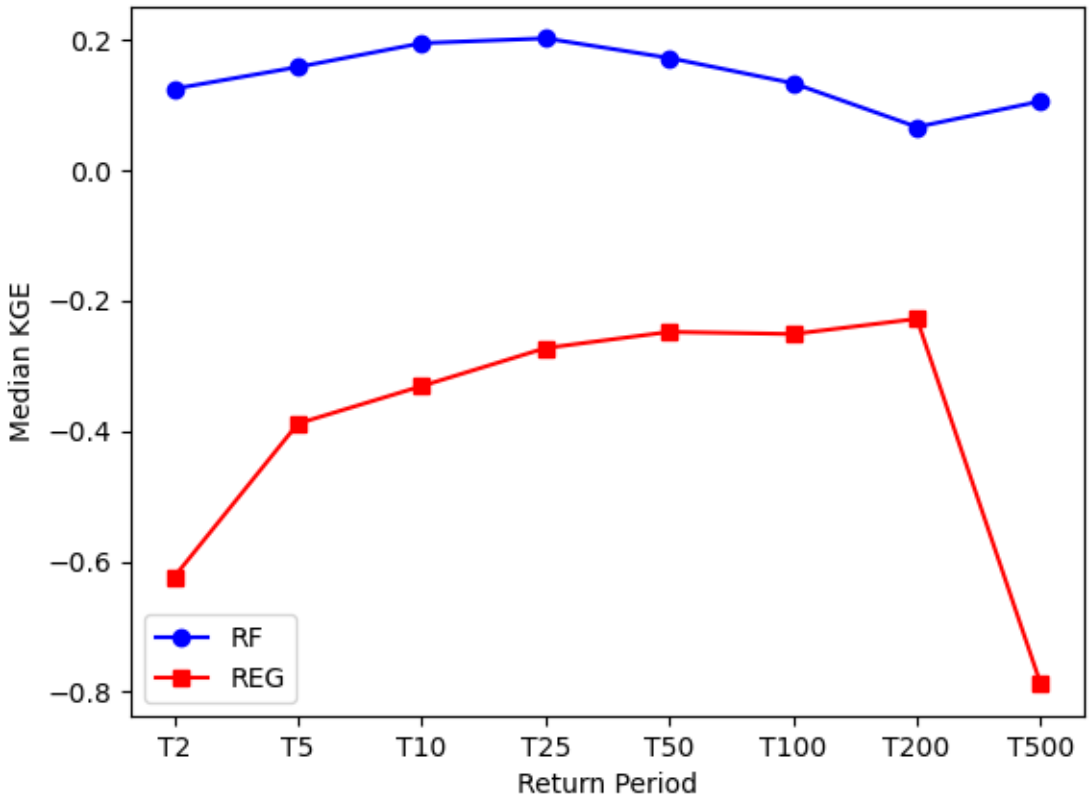

Figure 3.7 Median KGE for RF model (blue) and regression equation (red) across different return periods.

The interquartile ranges (IQRs) in the case of RF models (Figure 3.6) were shorter than the regression equations, indicating lower variability in the KGE values in the former. Moreover, in the case of regression equations, the lower outliers in the KGE values stretched much further compared to their counterparts in the RF models. These results also highlight the favorable performance of the RF models in comparison to the regression equations and have significant implication for engineering practices.

It is evident from the results that using data from across different locations within the state, even if those places are not alike, leads to a more accurate estimation of peak flows compared to the traditional regional regression equations built using datasets from similar regions. This is particularly significant due to the changing climate causing unpredictable weather patterns in different regions (Lee, 2023). As this trend continues, the response to extreme events of a basin can be used as an experience to provide insights into other basins. This notion matches with the perspectives of researchers such as Sivapalan (2006) and Wagener et al. (2020) who suggest shifting the focus from local distinction to a broader perspective in hydrological studies. Our study highlights the intrinsic worth of data synergy (Fang et al., 2022) in precisely estimating peak flow across diverse return periods. Data synergy involves the advantageous outcome of pooling and combining diverse datasets from different regions. By incorporating diverse datasets from various regions within the state, our model demonstrates remarkable performance.

3.5 Conclusions

In this study, we employed a suite of RF models to accurately estimate peak flow for different return periods in 40 Nebraska streams. We also compared the performance of the RF models with that of the existing regional regression equations developed by Strahm and Admiraal (2005) used for peak flow estimation. Our results demonstrated that the RF models outperformed the regression equations (fitted to older datasets) for peak flow estimations across all return periods, with median KGE scores above zero, while cores for the regression equations were below zero. The RF models showed KGE values greater than 0.75 for 2, 10, 25, 50, and 100-year return periods, while for 5, 200, and 500-year return periods, KGE values were between 0.5 and 0.75. Additionally, the variability in the KGE values was comparatively lower in RF

models when both systems were tested for 100 random training-testing basin sets. These findings suggest that the RF models outperformed the existing regional regression equations fitted using older datasets.

The outcomes of this study have important implications for stakeholders and decision-makers in infrastructure design and disaster management. Accurate prediction of peak flow can help guide the design and construction of crucial infrastructure like bridges, dams, and levees, ensuring its resilience and safety during flood events. Additionally, the findings of the study can benefit the development of emergency response plans and warning systems, which are critical in lowering the danger of loss of life and property damage during floods. Furthermore, the results can aid transportation authorities in taking preventative measures such as road closures, emergency evacuations, and the development of alternative routes that can minimize the disruption to daily life and business activities, thereby improving the socio-economic resilience of the affected communities.

Overall, the use of high-resolution datasets and the deployment of ML methods in this study have helped us achieve better performance in peak flow estimation compared to the existing regression equations. Our findings highlight the potential of ML techniques in improving the accuracy of hydrological estimations and advancing water management practices.

Chapter 4 Dissemination

Peer-reviewed Journals

Pokharel, S., Roy, T., Admiraal, D., 2023. Effects of mass balance, energy balance, and storage-discharge constraints on LSTM for streamflow prediction. Environ. Model. Softw. 166, 105730. https://doi.org/10.1016/j.envsoft.2023.105730 (**Impact Factor: 4.9**)

Pokharel, S., Roy, T., Admiraal, D., 2023. Machine Learning-based Peak flow Estimation for Nebraska Streams, under review in Journal of Hydrology (**Impact Factor: 6.4**)

Conference Presentations

Pokharel, S., T. Roy, and D. Admiraal (2023), Enhancing Peakflow Estimation in Nebraska with Machine Learning, *Nebraska Water Conference*, Oct 3-4, Omaha.

Pokharel, S., T. Roy, and D. Admiraal (2023), Enhancing Transportation Safety through Improved Peak Streamflow Prediction using Machine Learning Techniques, *EGU General Assembly*, Apr 23-28, Vienna.

Pokharel, S., T. Roy, and D. Admiraal (2022), A Physics-guided Machine Learning Scheme for Predicting Peak Flow in Streams, *AGU Fall Meeting*, Dec 12-16, Chicago.

Pokharel, S., D. Admiraal, and T. Roy (2022), A Physics-Based Machine Learning Scheme for Predicting Peak Flows in Nebraska Streams, *Student Research Days*, UNL, Lincoln.

Seminar Presentations

Pokharel, S., Admiraal, D., Roy, T. (2022, March 4). A Physics-Based Machine Learning Scheme for Predicting Peak Flows in Nebraska Streams [PowerPoint slides]. *Environmental and Water Resources Engineering Seminar* Series, University of Nebraska-Lincoln, Lincoln.

Codes/Software/Data

Data: The datasets are collected from multiple sources. Please refer to "Getting the Prediction" word document for the name and source of the datasets used. This file is available on the following OneDrive link. The hydrometeorological datasets are obtained from ERA5. The folder does not include the ERA5 data since that can be easily downloaded from its original source. Datasets are stored in a physical hard drive as well.

Codes: Relevant codes developed in this project are uploaded in the OneDrive folder under the subfolder named "Code". To run the codes, please follow the instruction provided in the word document "Getting the Prediction". All the codes are written in Python using Spyder IDE from Anaconda. This software system (and others) can be used to reproduce the results.

OneDrive Folder Link:

https://uofnelincoln-my.sharepoint.com/:f:/g/personal/spokharel2_unl_edu/EgOXBxiSnXhArmDTJMNMZsUBIPZD89IHV57IhrM-IxYwmw?e=vEofoA

References

Anderson, S., Radi´ c, V., 2022. Interpreting deep machine learning for streamflow modeling across glacial, nival, and pluvial regimes in southwestern Canada. Frontiers in Water 4, 92. https://doi.org/10.3389/FRWA.2022.934709/BIBTEX.

Bao, Z., Zhang, J., Liu, J., Fu, G., Wang, G., He, R., Yan, X., Jin, J., Liu, H., 2012. Comparison of regionalization approaches based on regression and similarity for predictions in ungauged catchments under multiple hydro-climatic conditions. J. Hydrol. 466–467, 37–46. https://doi.org/10.1016/J.JHYDROL.2012.07.048

Beard, L.R., 1997. Estimating Flood Frequency and Average Annual Damage. J. Water Resour. Plan. Manag. 123, 84–88. https://doi.org/10.1061/(asce)0733-9496(1997)123:2(84)

Beckman, E.., 1976. Magnitude and frequency of floods in Nebraska,Water-Resources Investigations Report 76-109,128p.

Belete, D.M., Huchaiah, M.D., 2021. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. https://doi.org/10.1080/1206212X.2021.1974663 44, 875–886. https://doi.org/10.1080/1206212X.2021.1974663

Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Network. 5 (2), 157–166. https:// doi.org/10.1109/72.279181.

Berghuijs, W.R., Aalbers, E.E., Larsen, J.R., Trancoso, R., Woods, R.A., 2017. Recent changes in extreme floods across multiple continents. Environ. Res. Lett. 12, 114035. https://doi.org/10.1088/1748-9326/AA8847

Beven, K., 2014. What we see now: event-persistence and the predictability of hydro-eco-geomorphological systems. Ecol. Model. 298, 4–15. https://doi.org/10.1016/J.ECOLMODEL.2014.07.019.

Bhasme, P., Vagadiya, J., Bhatia, U., 2021. Enhancing predictive skills in physically- consistent way: physics informed machine learning for hydrological processes, 1–20. http://arxiv.org/abs/2104.11009.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R.A.P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G.T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G.B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T.R., Kohnová, S., Koskela, J.J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J.L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K., Živković, N., 2019. Changing climate both increases and decreases European river floods. Nat. 2019 5737772 573, 108–111. https://doi.org/10.1038/s41586-019-1495-6

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Brunner, M.I., Furrer, R., Sikorska, A.E., Viviroli, D., Seibert, J., Favre, A.C., 2018. Synthetic design hydrographs for ungauged catchments: a comparison of regionalization methods. Stoch. Environ. Res. Risk Assess. 32, 1993–2023. https://doi.org/10.1007/s00477-018-1523-3

Buytaert, W., Beven, K., 2009. Regionalization as a learning process. Water Resour. Res. 45, 11419. https://doi.org/10.1029/2008WR007359

Castangia, M., Grajales, L.M.M., Aliberti, A., Rossi, C., Macii, A., Macii, E., Patti, E., 2023. Transformer neural networks for interpretable flood forecasting. Environ. Model. Software 160, 105581. https://doi.org/10.1016/J.ENVSOFT.2022.105581.

Chang, M.J., Chang, H.K., Chen, Y.C., Lin, G.F., Chen, P.A., Lai, J.S., Tan, Y.C., 2018. A support vector machine forecasting model for typhoon flood inundation mapping and early flood warning systems. Water 2018 10 (12), 1734. https://doi.org/ 10.3390/W10121734. Page 1734, 10.

Dalkiliç, H.Y., Hashimi, S.A., 2020. Prediction of daily streamflow using artificial neural networks (ANNs), wavelet neural networks (WNNs), and adaptive neuro-fuzzy inference system (ANFIS) models. Water Supply 20 (4), 1396–1408. https://doi.org/ 10.2166/WS.2020.062.

Damavandi, H.G., Shah, R., Stampoulis, D., Wei, Y., Boscovic, D., Sabo, J., 2019. Accurate prediction of streamflow using long short-term memory network: a case study in the Brazos river basin in Texas. Int. J. Environ. Sustain Dev. 10 (10), 294–300. https://doi.org/10.18178/IJESD.2019.10.10.1190.

Daw, A., Karpatne, A., Watkins, W., Read, J., Kumar, V., 2017. Physics-guided Neural Networks (PGNN): an Application in Lake Temperature Modeling. http://arxiv. org/abs/1710.11431.

Daw, A., Thomas, R.Q., Carey, C.C., Read, J.S., Appling, A.P., Karpatne, A., 2019. Physics-guided architecture (PGA) of neural networks for quantifying uncertainty in lake temperature modeling. In: Proceedings of the 2020 SIAM International Conference on Data Mining. SDM, pp. 532–540. https://doi.org/10.1137/ 1.9781611976236.60, 2020.

Desai, S., Ouarda, T.B.M.J., 2021. Regional hydrological frequency analysis at ungauged sites with random forest regression. J. Hydrol. 594, 125861. https://doi.org/10.1016/J.JHYDROL.2020.125861

Ding, J., Haberlandt, U., 2017. Estimation of instantaneous peak flow from maximum mean daily flow by regionalization of catchment model parameters. Hydrol. Process. 31, 612–626. https://doi.org/10.1002/HYP.11053

Dong, L., Fang, D., Wang, X., Wei, W., Damaˇ seviˇ cius, R., Scherer, R., Woʹ zniak, M., 2020. Prediction of streamflow based on dynamic sliding window LSTM. Water 2020 12 (11), 3032. https://doi.org/10.3390/W12113032. Page 3032, 12.

Doocy, S., Daniels, A., Murray, S., Kirsch, T.D., 2013. The human impact of floods: a historical review of events 1980-2009 and systematic literature review. PLoS Curr. 5. https://doi.org/10.1371/CURRENTS.DIS.F4DEB457904936B07C09DAA98EE8171A

Fang, K., Kifer, D., Lawson, K., Feng, D., Shen, C., 2022. The Data Synergy Effects of Time-Series Deep Learning Models in Hydrology. Water Resour. Res. 58. https://doi.org/10.1029/2021WR029583

Flanagan, P.X., Mahmood, R., Umphlett, N.A., Haacker, E., Ray, C., Sorensen, W., Shulski, M., Stiles, C.J., Pearson, D., Fajman, P., 2020. A hydrometeorological assessment of the historic 2019 flood of Nebraska, Iowa, and South Dakota. Bull. Am. Meteorol. Soc. 101, E817–E829. https://doi.org/10.1175/BAMS-D-19-0101.1

Furness, L.., 1955. Floods in Nebraska,magnitude and frequency,Lincoln,Nebraska,Nebraska Department of Roads and Irrigation,103p.

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., Hochreiter, S., 2021a. Rainfall- runoff prediction at multiple timescales with a single Long Short-Term Memory network. Hydrol. Earth Syst. Sci. 25 (4), 2045–2062. https://doi.org/10.5194/hess- 25-2045-2021.

Gauch, M., Mai, J., Lin, J., 2021b. The proper care and feeding of CAMELS: how limited training data affects streamflow prediction. Environ. Model. Software 135, 104926. https://doi.org/10.1016/J.ENVSOFT.2020.104926.

Ghimire, P., Pokharel, S., Kim, K., Barutha, P., 2023. MACHINE LEARNING-BASED PREDICTION MODELS FOR BUDGET FORECAST CEES 2023 | 2 nd International Conference on MACHINE LEARNING-BASED PREDICTION MODELS FOR BUDGET FORECAST IN 0–6.

Ghimire, S., Yaseen, Z.M., Farooque, A.A., Deo, R.C., Zhang, J., Tao, X., 2021. Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. Sci. Rep. 11 (1), 1–26. https://doi.org/10.1038/s41598-021-96751-4, 2021 11:1.

Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., Kagal, L., 2019. Explaining explanations: an overview of interpretability of machine learning. In: Proceedings - 2018 IEEE 5th International Conference on Data Science and Advanced Analytics. DSAA, pp. 80–89. https://doi.org/10.1109/DSAA.2018.00018, 2018.

Golian, S., Murphy, C., Meresa, H., 2021. Regionalization of hydrological models for flow estimation in ungauged catchments in Ireland. J. Hydrol. Reg. Stud. 36, 100859. https://doi.org/10.1016/J.EJRH.2021.100859

Goswami, M., O'Connor, K.M., Bhattarai, K.P., 2007. Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment. J. Hydrol. 333, 517–531. https://doi.org/10.1016/J.JHYDROL.2006.09.018

Guo, Y., Zhang, Y., Zhang, L., Wang, Z., 2021. Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review. Wiley Interdiscip. Rev. Water. https://doi.org/10.1002/wat2.1487

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, 80–91. https://doi.org/10.1016/J.JHYDROL.2009.08.003

Ha, S., Liu, D., Mu, L., 2021. Prediction of yangtze river streamflow based on deep learning neural network with el ni˜ no–southern oscillation. Sci. Rep. 11 (1), 1–23. https://doi.org/10.1038/s41598-021-90964-3.

Hadi, S.J., Tombul, M., 2018. Forecasting daily streamflow for basins with different physical characteristics through data-driven methods. Water Resour. Manag. 32 (10), 3405–3422. https://doi.org/10.1007/S11269-018-1998-1/FIGURES/7.

Han, X., Mehrotra, R., Sharma, A., Rahman, A., 2022. Incorporating nonstationarity in regional flood frequency analysis procedures to account for climate change impact. J. Hydrol. 612, 128235. https://doi.org/10.1016/J.JHYDROL.2022.128235

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Springer Series in Statistics. https://doi.org/10.1007/978-0-387-84858-7

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Hor´anyi, A., Mu˜ noz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., et al., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146 (730), 1999–2049. https://doi.org/10.1002/QJ.3803.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S., 2013. Global flood risk under climate change. Nat. Clim. Chang. 3, 816–821. https://doi.org/10.1038/nclimate1911

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/NECO.1997.9.8.1735.

Hoedt, P.-J., Kratzert, F., Klotz, D., Halmich, C., Holzleitner, M., Nearing, G., Hochreiter, S., Klambauer, G., 2021. MC-LSTM: mass-conserving LSTM. https ://arxiv.org/abs/2101.05186v3. 7S.

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H. V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB)-a review. Hydrol. Sci. J. 58, 1198–1255. https://doi.org/10.1080/02626667.2013.803183

Hrnjica, B., Mehr, A.D., 2020. Energy Demand Forecasting Using Deep Learning. EAI/ Springer Innovations in Communication and Computing, pp. 71–104. https://doi. org/10.1007/978-3-030-14718-1_4.

Hundecha, Y., Zehe, E., Bárdossy, A., 2002. Regional parameter estimation from catchment properties prediction in ungauged basins 309, 20–22.

Hunt, K.M.R., Matthews, G.R., Pappenberger, F., Prudhomme, C., 2022. Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States. Hydrol. Earth Syst. Sci. Discuss. https://doi.org/10.5194/ hess-2022-53. February, 1–30.

Hutchison, N.E., 1962. Floods in Nebraska on Small Drainage Areas Magnitude and Frequency.

ISCWD, 1982. Guidelines for determining flood flow frequency (revised). [WWW Document].

Jarboe, J.E., Haan, C.T., 1974. Calibrating a water yield model for small ungaged watersheds. Water Resour. Res. 10, 256–262. https://doi.org/10.1029/WR010I002P00256

Jia, X., Willard, J., Karpatne, A., Read, J., Zwart, J., Steinbach, M., Kumar, V., 2019. Physics guided RNNs for modeling dynamical systems: a case study in simulating lake temperature profiles. Proceedings 558–566. https://doi.org/10.1137/ 1.9781611975673.63.

Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2017. Theory-guided data science: a new paradigm for scientific discovery from data. IEEE Trans. Knowl. Data Eng. 29 (10), 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168.

Khandelwal, A., Xu, S., Li, X., Jia, X., Stienbach, M., Duffy, C., Nieber, J., Kumar, V., 2020. Physics Guided Machine Learning Methods for Hydrology. http://arxiv. org/abs/2012.02854.

Kienzler, S., Pech, I., Kreibich, H., Müller, M., Thieken, A.H., 2015. After the extreme flood in 2002: Changes in preparedness, response and recovery of flood-affected residents in Germany between 2005 and 2011. Nat. Hazards Earth Syst. Sci. 15, 505–526. https://doi.org/10.5194/nhess-15-505-2015

Kim, H. Il, Han, K.Y., 2020. Inundation map prediction with rainfall return period and machine learning. Water 2020 12 (6), 1552. https://doi.org/10.3390/W12061552. Page 1552, 12.

Krajewski, W.F., Ceynar, D., Demir, I., Goska, R., Kruger, A., Langel, C., Mantilllla, R., Niemeier, J., Quintero, F., Seo, B.C., Smallll, S.J., Weber, L.J., Young, N.C., 2017. Real-time flood

forecasting and information system for the state of Iowa. Bull. Am. Meteorol. Soc. 98 (3), 539–554. https://doi.org/10.1175/BAMS-D-15-00243.1.

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. https://doi.org/10.5194/HESS-22-6005-2018.

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 Resour. Res. 55 (12), 11344–11354. https://doi.org/ 10.1029/2019WR026065.

LeCun, Y.A., Bottou, L., Orr, G.B., Müller, K.R., 2012. Efficient Backprop. Lecture Notes In Computer Science (Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics), 7700 LECTURE NO, 9–48. https://doi.org/10.1007/978-3- 642-35289-8_3/COVER.

Lee, 2023. Future Global Climate: Scenario-based Projections and Near-term Information, in: Climate Change 2021 – The Physical Science Basis. Cambridge University Press, pp. 553–672. https://doi.org/10.1017/9781009157896.006

Li, Z., Gao, S., Chen, M., Gourley, J.J., Liu, C., Prein, A.F., Hong, Y., 2022. The conterminous United States are projected to become more prone to flash floods in a high-end emissions scenario. Commun. Earth Environ. 3, 1–9. https://doi.org/10.1038/s43247-022-00409-6

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest 2.

Lin, G.F., Chou, Y.C., Wu, M.C., 2013. Typhoon flood forecasting using integrated two- stage Support Vector Machine approach. J. Hydrol. 486, 334–342. https://doi.org/ 10.1016/J.JHYDROL.2013.02.012.

Majeske, N., Zhang, X., Sabaj, M., Gong, L., Zhu, C., Azad, A., 2022. Inductive predictions of hydrologic events using a long short-term memory network and the soil and water assessment tool. Environ. Model. Software 152, 105400. https://doi.org/10.1016/J. ENVSOFT.2022.105400.

Marijnissen, R., Kok, M., Kroeze, C., Van Loon-Steensma, J., 2019. Re-evaluating safety risks of multifunctional dikes with a probabilistic risk framework. Nat. Hazards Earth Syst. Sci. 19, 737–756. https://doi.org/10.5194/NHESS-19-737-2019

Matthai, H.., 1968. Magnitude and Frequency of Urban Floods in the United States. 534.

Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J.C.J.H., Bates, P., Bertola, M., Kemter, M., Kreibich, H., Lall, U., Macdonald, E., 2021. Causes, impacts and patterns of disastrous river floods. Nat. Rev. Earth Environ. 2021 29 2, 592–609. https://doi.org/10.1038/s43017-021-00195-3

Minns, A.W., Hall, M.J., 2009. Artificial Neural Networks as Rainfall-Runoff Models. https://doi.org/10.1080/02626669609491511. Https://Doi.Org/10.1080/ 02626669609491511, 41(3), 399–417.

Moon, S.H., Kim, Y.H., Lee, Y.H., Moon, B.R., 2019. Application of machine learning to an early warning system for very short-term heavy rainfall. J. Hydrol. 568, 1042–1054. https://doi.org/10.1016/J.JHYDROL.2018.11.060.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo, C., Thépaut, J.N., 2021. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth Syst. Sci. Data 13, 4349–4383. https://doi.org/10.5194/ESSD-13-4349-2021

NCEI, 2023. U.S. Billion-dollar Weather and Climate Disasters, 1980 - present [WWW Document]. Natl. Centers Environ. Inf. https://doi.org/10.25921/STKW-7W73

Nevo, S., Morin, E., Rosenthal, A.G., Metzger, A., Barshai, C., Weitzner, D., Kratzert, F., Elidan, G., Dror, G., Begelman, G., Nearing, G., Noga, H., Shavitt, I., Yuklea, L., Royz, M., Giladi, N., Levi, N.P., Gilon, O., Maor, R., et al., 2021. Flood forecasting with machine learning models in an operational framework. November 1–31.

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.

NWS, 2022. Flooding in Nebraska.

Ouarda, T.B.M.J., Cunderlik, J.M., St-Hilaire, A., Barbet, M., Bruneau, P., Bobée, B., 2006. Data-based comparison of seasonality-based regional flood frequency methods. J. Hydrol. 330, 329–339. https://doi.org/10.1016/J.JHYDROL.2006.03.023

Pagliero, L., Bouraoui, F., Diels, J., Willems, P., McIntyre, N., 2019. Investigating regionalization techniques for large-scale hydrological modelling. J. Hydrol. 570, 220–235. https://doi.org/10.1016/J.JHYDROL.2018.12.071

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 Resour. Manag. 34 (13), 4113–4131. https://doi.org/ 10.1007/S11269-020-02659-5/FIGURES/10.

Patterson, J., 1966. Magnitude and Frequency of Floods in Minnesota.

Pellicciotti, F., Bauder, A., Parola, M., 2010. Effect of glaciers on streamflow trends in the Swiss Alps. Water Resour. Res. 46 (10) https://doi.org/10.1029/2009WR009039.

Peng, T., Zhou, J., Zhang, C., Fu, W., 2017. Streamflow forecasting using empirical wavelet transform and artificial neural networks. Water 2017 9 (6), 406. https://doi.org/10.3390/W9060406. Page 406, 9.

Pokharel, S., Roy, T., Admiraal, D., 2023. Effects of mass balance, energy balance, and storage-discharge constraints on LSTM for streamflow prediction. Environ. Model. Softw. 166, 105730. https://doi.org/10.1016/j.envsoft.2023.105730

Prieto, C., Le Vine, N., Kavetski, D., García, E., Medina, R., 2019. Flow Prediction in Ungauged Catchments Using Probabilistic Random Forests Regionalization and New Statistical Adequacy Tests. Water Resour. Res. 55, 4364–4392. https://doi.org/10.1029/2018WR023254

Prieto, C., Patel, D., Han, D., 2020. Preface: Advances in flood risk assessment and management. Nat. Hazards Earth Syst. Sci. 20, 1045–1048. https://doi.org/10.5194/NHESS-20-1045-2020

Pyayt, A.L., Kozionov, A.P., Kusherbaeva, V.T., Mokhov, I.I., Krzhizhanovskaya, V.V., Broekhuijsen, B.J., Meijer, R.J., Sloot, P.M.A., 2014. Signal analysis and anomaly detection for flood early warning systems. J. Hydroinf. 16 (5), 1025–1043. https:// doi.org/10.2166/HYDRO.2014.067.

Rasheed, Z., Aravamudan, A., Gorji Sefidmazgi, A., Anagnostopoulos, G.C., Nikolopoulos, E.I., 2022. Advancing flood warning procedures in ungauged basins with machine learning. J. Hydrol. 609, 127736. https://doi.org/10.1016/J.JHYDROL.2022.127736

Rasiya, S., Velasquez, N., Rojas, M., Mantilla, R., Harvey, K., Ceynar, D., Quintero, F., Krajewski, W.F., Roy, T., 2023. Applicability of a flood forecasting system for Nebraska watersheds. Environ. Model. Softw. 164, 105693. https://doi.org/10.1016/j.envsoft.2023.105693

Razavi, T., Coulibaly, P., 2012. Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods. J. Hydrol. Eng. 18, 958–975. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000690

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. https://doi.org/10.1038/s41586-019- 0912-1, 2019 566:7743.

Safeeq, M., Bart, R.R., Pelak, N.F., Singh, C.K., Dralle, D.N., Hartsough, P., Wagenbrenner, J.W., 2021. How realistic are water-balance closure assumptions? A demonstration from the southern sierra critical zone observatory and kings river experimental watersheds. Hydrol. Process. 35 (5), e14199 https://doi.org/10.1002/ HYP.14199.

Sanders, W., Li, D., Li, W., Fang, Z.N., 2022. Data-Driven flood alert system (FAS) using extreme gradient boosting (XGBoost) to forecast flood stages. Water 2022 14 (5), 747. https://doi.org/10.3390/W14050747. Page 747, 14.

Schwartz, M., 2019. Nebraska Faces Over $1.3 Billion In Flood Losses, NPR News on March 21, 2019. [WWW Document]. URL https://www.npr.org/2019/03/21/705408364/nebraska-faces-over-1-3-billion-in-flood-losses

Shortridge, J.E., Guikema, S.D., Zaitchik, B.F., 2016. Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in

seasonal watersheds. Hydrol. Earth Syst. Sci. 20 (7), 2611–2628. https://doi.org/10.5194/HESS-20-2611-2016.

Shu, C., Ouarda, T.B.M.J., 2008. Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. J. Hydrol. 349, 31–43. https://doi.org/10.1016/J.JHYDROL.2007.10.050

Singh, V.P., 1998. Log-Pearson Type III Distribution 252–274. https://doi.org/10.1007/978-94-017-1431-0_15

Sivapalan, M., 2006. Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, in: Encyclopedia of Hydrological Sciences. Wiley. https://doi.org/10.1002/0470848944.hsa012

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Conell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., Zehe, E., 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. Hydrol. Sci. J. 48, 857–880. https://doi.org/10.1623/hysj.48.6.857.51421

Soenksen, P.J., Miller, L.D., Sharpe, J.B., Watton, J.R., 1999. Peak-Flow Frequency Relations and Evaluation of the Peak-Flow Gaging Network in Nebraska Peak-Flow Frequency Relations and Evaluation of the Peak-Flow Gaging Network in Nebraska.

Song, T., Ding, W., Wu, J., Liu, H., Zhou, H., Chu, J., 2020. Flash flood forecasting based on long short-term memory networks. Water (Switzerland) 12 (1). https://doi.org/ 10.3390/w12010109.

Song, Z., Xia, J., Wang, G., She, D., Hu, C., Hong, S., 2022. Regionalization of hydrological model parameters using gradient boosting machine. Hydrol. Earth Syst. Sci. 26, 505–524. https://doi.org/10.5194/HESS-26-505-2022

Strahm, B.J., Admiraal, D.M., 2005. W355 Nebraska Hall Sponsored by The Nebraska Department of Roads 1500 Nebraska Highway 2 Department of Civil Engineering University of Nebraska – Lincoln W348 Nebraska Hall U . S . Department of Transportation Research and Special Programs Administration 1.

Sun, A.Y., Wang, D., Xu, X., 2014. Monthly streamflow forecasting using Gaussian Process Regression. J. Hydrol. 511, 72–81. https://doi.org/10.1016/J.JHYDROL.2014.01.023

Tanoue, M., Hirabayashi, Y., Ikeuchi, H., 2016. Global-scale river flood vulnerability in the last 50 years. Sci. Rep. 6 https://doi.org/10.1038/SREP36021.

Thapa, Samit, Zhao, Zebin, Bo Li, L., 2020. Snowmelt-Driven Streamflow Prediction Using. Water. Tongal, H., Booij, M.J., 2018. Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. J. Hydrol. 564 (July), 266–282. https://doi.org/10.1016/j.jhydrol.2018.07.004.

Troch, P.A., Lahmers, T., Meira, A., Mukherjee, R., Pedersen, J.W., Roy, T., Vald´ es- Pineda, R., 2015. Catchment coevolution: a useful framework for improving predictions of hydrological change? Water Resour. Res. 51 (7), 4903–4922. https:// doi.org/10.1002/2015WR017032.

Tyralis, H., Papacharalampous, G., Langousis, A., 2019. A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. Water 2019, Vol. 11, Page 910 11, 910. https://doi.org/10.3390/W11050910

Velásquez, N., Quintero, F., Koya, S.R., Roy, T., Mantilla, R., 2023. Snow-detonated floods : Assessment of the U.S. midwest march 2019 event. J. Hydrol. Reg. Stud. Reg. Stud. 47. https://doi.org/10.1016/j.ejrh.2023.101387

Villarini, G., Smith, J.A., Baeck, M.L., Krajewski, W.F., 2011. Examining Flood Frequency Distributions in the Midwest U.S.1. JAWRA J. Am. Water Resour. Assoc. 47, 447–463. https://doi.org/10.1111/J.1752-1688.2011.00540.X

Wagener, T., Gleeson, T., Coxon, G., Hartmann, A., Howden, N., Pianosi, F., Rahman, M., Rosolem, R., Stein, L., Woods, R., 2020. On doing large-scale hydrology with Lions: Realising the value of perceptual models and knowledge accumulation. https://doi.org/10.31223/OSF.IO/ZDY5N

Willard, J., Jia, X., Xu, S., Steinbach, M., Kumar, V., 2020. Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems 1 (1), 1–34. http://arxiv.org/abs/2003.04919.

William Farmer, B.H., Kiang, J.E., Feaster, T.D., Eng, K., n.d. Techniques and Methods 4-A12 Regionalization of Surface-Water Statistics Using Multiple Linear Regression Book 4, Hydrologic Analysis and Interpretation.

Wu, J., Chen, X.Y., Zhang, H., Xiong, L.D., Lei, H., Deng, S.H., 2019. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. J. Electron. Sci. Technol. 17, 26–40. https://doi.org/10.11989/JEST.1674-862X.80904120

Xiang, Z., Demir, I., 2020. Distributed long-term hourly streamflow predictions using deep learning – a case study for State of Iowa. Environ. Model. Software 131, 104761. https://doi.org/10.1016/J.ENVSOFT.2020.104761.

Yang, M., Yang, Q., Shao, J., Wang, G., Zhang, W., 2023. A New Few-Shot Learning Model for Runoff Prediction: Demonstration in Two Data Scarce Regions. Environmental Modelling & Software, 105659. https://doi.org/10.1016/J. ENVSOFT.2023.105659.

Yaseen, Z.M., Jaafar, O., Deo, R.C., Kisi, O., Adamowski, J., Quilty, J., El-Shafie, A., 2016. Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. J. Hydrol. 542, 603–614. https://doi.org/10.1016/J. JHYDROL.2016.09.035.

Young, A.R., 2006. Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model. J. Hydrol. 320, 155–172. https://doi.org/10.1016/J.JHYDROL.2005.07.017

Young, C.C., Liu, W.C., Wu, M.C., 2017. A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. Appl. Soft Comput. 53, 205–216. https://doi.org/10.1016/J.ASOC.2016.12.052.

Yuan, X., Chen, C., Lei, X., Yuan, Y., Muhammad Adnan, R., 2018. Monthly runoff forecasting based on LSTM–ALO model. Stoch. Environ. Res. Risk Assess. 32 (8), 2199–2212. https://doi.org/10.1007/s00477-018-1560-y.