Neural Computing and Applications (2022) 34:18917–18931
https://doi.org/10.1007/s00521-022-07500-1

ORIGINAL ARTICLE

On the use of machine learning to account for reservoir management rules and predict streamflow

Achraf Tounsi[1] · Marouane Temimi[1] · Jonathan J. Gourley[2]

Received: 5 February 2022 / Accepted: 1 June 2022 / Published online: 25 June 2022

Abstract
This study aims to develop a Machine Learning (ML)-based technique to infer reservoir management rules and predict downstream discharge values. The case study is the Hackensack River Watershed in New Jersey, USA. A Long Short-Term Memory (LSTM) model was used to predict streamflow values at the USGS station at New Milford, right downstream of Oradell reservoir. A good agreement between observed and simulated streamflow values was obtained during the 2020–2021 testing period. An NSE value of 0.93 was determined with the 48-h precipitation lead time, suggesting that the 48-h precipitation forecast mostly drives releases Oradell reservoir. The developed model was tested during Hurricane Ida. The analysis revealed that a similar NSE of 0.95 was obtained with a 48-h precipitation lead time followed by the 12-h lead time model, which was based on the watershed response time. In addition, the conducted feature analysis revealed that only four out of the seven upstream USGS stations in the watershed have a significant impact on the model's performance. This work implies that ML can capture reservoir management rules and predict reservoir releases using precipitation and upstream flow data as input variables. This study lays the groundwork for a generalization of the method over the CONUS to infer reservoirs' operation rules for streamflow simulation.

Keywords Machine learning · LSTM · Streamflow · MRMS · Precipitation · Water reservoir

1 Introduction

Flooding is a natural disaster with consequences that are particularly severe in urban areas where significant damages to infrastructure are often reported [1]. Predicting hydrological processes in urban areas is challenging, especially in the presence of reservoirs that are commonly used for flood control and impact mitigation. The presence of water reservoirs along streams, which is common in or around urban areas, and their releases depending on specific management rules dictated by their managers is an example that illustrates the complexity of modeling hydrological processes in urban areas. To accurately predict urban flooding and its risk, it is vital to account for anthropogenic factors, especially those introduced by streamflow regulation that alters natural streamflow regimes through reservoir management [2, 3].

Several hydrological models have been developed to study watershed hydrology with applications in rural and urban areas [4, 5]. Models like SWMM, HEC-RAS [6], TUFLOW [7], CSIRO TVD, and AutoRAPID [8] are often used to simulate urban hydrological processes while accounting for the combination of riverine and urban flooding. The models are usually developed in 1D, 2D, or a combination of 1D-2D configurations. The 1D modeling component usually covers the streams, the subsurface drainage network, and major drainage center lines in a watershed. The 2D modeling component simulates the lateral expansion of flood inundation. The 1D and 2D components have usually coupled both ways. Introducing

✉ Achraf Tounsi
atounsi@stevens.edu

Marouane Temimi
mtemimi@stevens.edu

Jonathan J. Gourley
jj.gourley@noaa.gov

[1] Department of Civil, Environmental, and Ocean Engineering, Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ 07030, USA

[2] National Oceanic and Atmospheric Administration (NOAA), National Severe Storms Laboratory, Norman, OK, USA

additional components like water reservoirs in such modeling configurations increases the level of complexity of the models. In addition, the management rules of reservoirs are usually not publicly available. Water reservoirs are managed and operated by different public and private entities. Depending on their size and capacity, they may have different primary functions like flood protection and/or water supply. Reservoir operators often have to optimize the operation of the reservoir to maximize the benefits in terms of water supply, flood protection, and hydropower production, which leads to complex management protocols that are hard to infer and incorporate into hydrological and hydraulic models. In the presence of water reservoirs in a watershed, the performance of the models mentioned above tends to be limited in simulating the spatial and temporal dynamics of urban hydrological processes [9].

The optimum operation of water reservoirs is the main challenge facing water resources managers and decision-makers. Optimizing operational procedures is a requirement for the effective management and planning of complex water resources systems [10, 11]. Reservoirs' management strategies are usually defined depending on the primary function of the reservoir as well as the prevailing and future hydrological and meteorological conditions. When reservoir management rules and protocols are not available, the analysis of water releases (output variable) and information on hydrological and meteorological variables can determine the missing rules assuming consistency in the management strategy. The unknown reservoirs' management rules add to the uncertainty in the model's parameters and precipitation forecast, which dictates releases in anticipation of extreme events [12, 13].

More recently, data-driven techniques of ML methods, such as neural networks (NNs) [14], support vector machines (SVMs) [15–18], and wavelet transform (WT) [19, 20], have been used in streamflow forecasting studies and acquired wide attention. Several studies demonstrated that ML methods could capture nonlinear behavior despite the complexity of the involved processes [20–23]. Several researchers examined different ML techniques with a focus on reservoir controls and their usefulness in predicting streamflow. Allawi et al. [24] investigated the usage of the Shark Machine Learning Algorithm (SMLA) to determine the optimal reservoir release policy and investigated the model's effectiveness in terms of dam and reservoir physical characteristics. Raman and Chandramouli [25] employed a feedforward neural network (FFN) to formulate an operating rule for a single reservoir for irrigation. Deka and Chandramouli et al. [26] trained an FFN to estimate optimum releases from a system of three reservoirs. Chang et al. [27] used an adaptive network-based fuzzy inference system to estimate the optimal water release from a single reservoir in Taiwan. Moreover, AI techniques like fuzzy logic, neural networks, genetic algorithms, support vector machines, and genetic programming have been widely used in several hydrologic applications [24, 28]. For instance, Shiri et al. [29] developed an extreme learning machine, genetic programming, and artificial neural network (ANN) models for predicting water levels in Urmia Lake. Emamgholi Zadeh et al. [30] used the ANN model for forecasting groundwater levels in the Bastam Plain. All these studies demonstrated the potential of using ML algorithms for hydrologic applications.

Recently, hydrologists have started to use Long Short-Term Memory (LSTM) in different applications. Introduced by Hochreiter and Schmidhuber [31], LSTM models are a powerful tool for solving time-series prediction problems [32]. LSTMs are proven to capture various time series data trends and behaviors and learn their long-range dependencies with high performance [33]. Zhang et al. [34] used three ML models, which are (1) the ANN, (2) support vector regression, and (3) LSTM in simulations of reservoir operations at the Gezhouba Dam. They reported that the LSTM had the best overall performance. Also, Kratzert et al. [35] built an LSTM model that describes the rainfall-runoff behavior of a large number of complex catchments on a daily scale. The study by Ni et al. [36] focused on developing two hybrid models based on the traditional LSTM model for monthly streamflow and rainfall forecasting. Hu et al. [37] showed that the LSTM model outperforms the ANN model for flood forecasting up to 6 h ahead. Chen et al. [1] proposed an LSTM neural network model for flood forecasting where the daily discharge and rainfall for 1-, 2-, and 3-day flowrate forecasting combinations were used as input data. The same study reported 0.99, 0.95, and 0.87 Nash–Sutcliffe efficiency values (NSE) corresponding to the three forecasting cases. An LSTM was found to be a viable option for flood forecasting on the Da River in Vietnam [38]. Yuan et al. [39] proposed an LSTM neural network and an Ant Lion Optimizer model (LSTM–ALO), which provides an effective method for monthly runoff forecasting in the Astor River Basin for the period between 1974 and 2009.

Overall, the previous studies demonstrated the potential of ML, particularly the LSTM method, in hydrological applications, especially in deducing reservoir management rules and using them in the prediction of downstream river discharge. The previous investigations showed a sensitivity of ML techniques to the quality of input variables, mainly the precipitation forecast. Releases from water reservoirs are driven by precipitation forecasts as well as the current conditions in the reservoir like water level and available storage capacity and the watershed like antecedent moisture conditions. In addition, the operation of reservoirs depends on the main function of the dam, like water

supply, hydropower production, or flood control and mitigation. In urban or rural areas, the reservoir's location can dictate different management rules as the impact of major releases on downstream locations is not the same. The risk is usually higher in urban areas, especially when parts of the floodplains are developed and densely populated.

This work aims to evaluate the capability of the LSTM model to infer reservoir operating rules and forecast river discharge at downstream stations. The case study is the Hackensack River watershed in northern New Jersey which comprises the Oradell reservoir that is used for water supply and flood protection in northern New Jersey. This study proposes using a novel application of the LSTM method to infer reservoir management rules. Then, downstream discharge values are predicted using a trained version of the model. In addition to developing a data-driven method to predict streamflow values, it is expected that the feature analysis of the LSTM model reveals the input variables that control water releases from a reservoir which can be expanded to watersheds where reservoir management rules are not known.

2 Methodology

2.1 Study area

The study domain covers the upstream subwatershed within the Hackensack River Basin that comprises the Oradell Reservoir (Fig. 1). The subwatershed includes flood-prone areas with a dominant dense residential land use type. The total drainage area of the reservoir located at the Borough of Oradell, Bergen County, New Jersey, is 292. 7 square kilometers (km^2). The reservoir is formed by a hollow concrete dam completed in 1922. The overall capacity at the spillway level is 13.275 million cubic meters (m^3) with an elevation of 7.06 ms [40]. Seven sluice gates regulate the reservoir release.

The study period extends between November 2014 and September 2021 and includes Hurricane Ida, which struck the entire northeastern region and caused significant flooding with a water elevation at the Oradell dam of 3.36 m leading to an overflow and flooding [40]. The area can be considered an important testing ground for studying a complex estuarine system with one of the most densely developed floodplains on the eastern seaboard [13]. The downstream area is considered highly urbanized, with vital assets which were heavily damaged due to Hurricane Ida in 20 21 [13]. Critical facilities are located around the river's floodplain, such as the New Jersey Transit's main storage and maintenance facility, the Port Authority Trans-Hudson (PATH) stations, and Teterboro Airport [41]. The coastal location of the watershed and the tidal influence on its streamflow add to the complexity of the modeling and simulation of hydrological and hydraulic processes, which requires a coupling of land and ocean models to simulate downstream w ater levels accurately.

2.2 Datasets

One key input variable of the developed LSTM model is rainfall over the Hackensack Watershed obtained from the radar-based Multi Radar/Multi-Sensor (MRMS) precipitation dataset [42]. In this study, a 7-year mosaic of MRMS rainfall measurements from 2014 to 2021 was used. The product is provided in Cartesian coordinates available every 2 min at a 1-km spatial resolution [43]. MRMS was deployed operationally in 2014 at the National Center for Environmental Prediction (NCEP). The system was developed to produce severe weather, transportation, and precipitation products for improved decision-making capability to improve hazardous weather forecasts and warnings, hydrology, aviation, and numerical weather prediction.

The upstream USGS stations included in the model are the Oradell Reservoir at Oradell, New Jersey (Monitoring station number 01378480), Lake Tappan at Old Tappan, New Jersey (Monitoring station number 01376950), Hackensack River at West Nyack New York (Monitoring station number 01376800) Pascack Brook at Park Ridge New Jersey (Monitoring station number 01377370), Woodcliff Lake at Hillsdale New Jersey (Monitoring station number 01377450), Pascack Brook at Westwood New Jersey (Monitoring station number 01377500), and the Hackensack River at Rivervale New Jersey (Monitoring station number 01377000) (Fig. 1). Information on the USGS stations can be obtained from [40]. Having a total drainage area of 76.6 km^2, the Pascack Brook at Westwood, New Jersey, is in the Westwood Borough, Bergen County, New Jersey, on the right bank, 22.9 m upstream from the bridge on Harrington Avenue in Westwood, 152.4 m downstream from Musquapsink Brook, and 3.70 km upstream from the mouth. Its flow is regulated by Woodcliff Lake, 4.8 km above the station. As for the Hackensack River at Rivervale New Jersey sta tion, it is in the River Vale Township, Bergen County, New Jersey, on the right bank upstream of the bridge on Westwood Avenue in Rivervale, 2.4 km upstream from Pascack Brook, 6.59 km downstream of Lake Tappan, and 7.4 km upstream from Oradell Dam having a total drainage area of 150.2 km^2. The flow of the station is regulated by Deforest Lake and Lake Tappan. Lake Tappan is an earthen dam with the Old Tappan station for the measuring stage. The reservoir's total capacity is 79,500 m^3 with an elevation of 16.8 m. The Woodcliff Lake at Hillsdale reservoir is an earthen dam with a drainage area of 50.2 km^2, a total

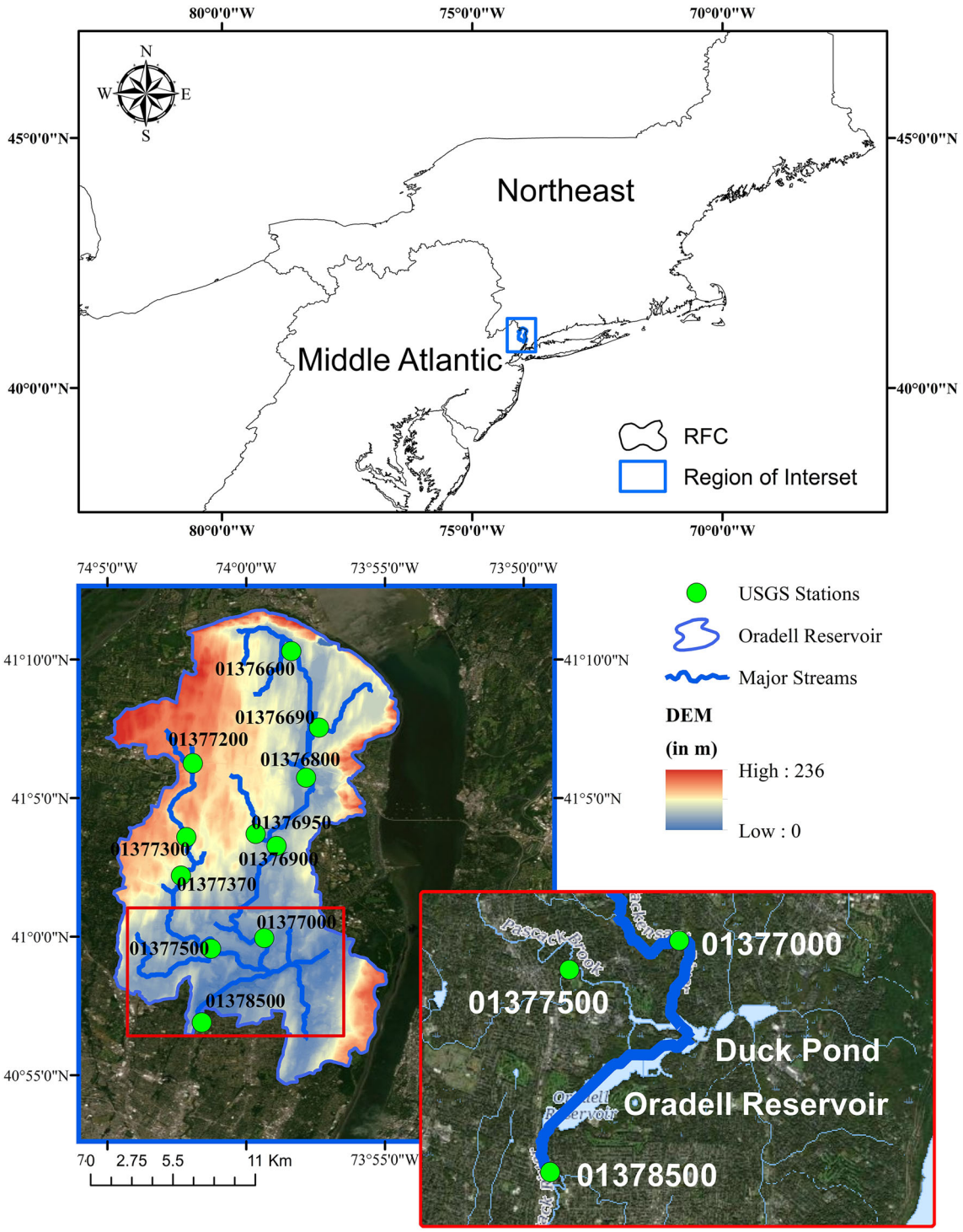

Fig. 1 Oradell reservoir and used USGS stations (with their labels)

capacity of 3.3 million m^3, and an elevation of 29 m. The Pascack Brook at Park Ridge is a continuous recording discharge station covering a drainage area of 34.7 km^2. Finally, the Hackensack River at West Nyack station covers a drainage area of 30.7 km^2. Its flow is regulated by Deforest Lake. When the discharge at the station is higher than 42.5 m^3/s, an overflow is reported due to an overtopping of the Klein Avenue Levee 480 m upstream of the gauge [40].

The downstream station that records releases from the Oradell Reservoir is the Hackensack River at New Milford (01 378 500). The station has a total drainage area of 292.7

km^2. The developed LSTM model will be verified downstream of the reservoir at the New Milford station site. It is worth noting that there is another USGS station downstream of New Milford on the Hackensack River in Hackensack. The flow at the Hackensack USGS station is tidally influenced, and streamflow is not reported at that station. The New Milford Station is the first downstream station on the Hackensack River where the discharge is reported. Immediately downstream of the New Milford station on the Hackensack River, one can find the Teterboro airport, an airport in the New York City area that is managed by the Port Authority of New York and New Jersey, which also reports precipitation continuously.

The dataset is divided into three parts: training set, validation set (holdout set), and testing set. These three sets are essential to avoid overfitting and model selection bias. Figure 2 illustrates the selection of data for the output variable. In the training set, which represents a sample of data used to fit the model, approximately 70% of the period from November 2014 to March 2019 was included. As for the validation set, the data between March 2019 and June 2020 is used. The goal of the validation set is to track progress through validation loss and accuracy. Finally, for the test data, which is the sample data used to provide an unbiased evaluation of a final model fit on the training dataset, data from June 2020 to the second week of September 2021 is selected.

2.3 ML configuration

The Recurrent Neural Network (RNN) is one class of ANNs that uses a recursive approach and can be used on sequential data. One key feature of this type of network is the network delay recursion which allows the description of a system's dynamic performance [44]. As shown in Fig. 3a, the network's output at time t is the result of the signal delay recursion, which associates both the input at time t and the recursive signals before the input at time t. Despite its ability to analyze short-term sequential data, the RNN has a drawback when learning long-range dependencies, which is critical for some time series forecasting problems, such as those considered in this study. To overcome this flaw, LSTM models were introduced as the long-term dependencies' saver in time series data thanks to their purpose-built memory cell [45]. In addition, LSTM models introduce a solution to the vanishing gradient problem related to long-term context memorization [31] as their architecture trims the gradients in the network using continuous error flows through constant error carousels within special multiplicative units [44]

The cells are composed of a sigmoid neural net layer and a multiplication operation, as depicted in Fig. 3b [44]. An input at time step t is (X_t), the hidden state from the previous time step that is introduced to the LSTM block (H_t), then the hidden state (S_t) is computed as follows:

1. The forget gate (f_t) will decide which information will be excluded from the cell state.

$$f_t = \sigma\left(X_t U^f + S_{t-1} W^f + b_f\right) \tag{1}$$

2. Decide which new information will be saved in the cell state. The input gate (i_t) first decides which values are to be updated. Then, a *tanh* layer creates a vector of new candidate values C_t.

$$i_t = \sigma\left(X_t U^i + S_{t-1} W^i + b_i\right) \tag{2}$$

$$\widetilde{C_t} = \tanh(X_t U^c + S_{t-1} W^c + b_c) \tag{3}$$

3. Update the old cell state (C_{t-1}) into the new state C_t:

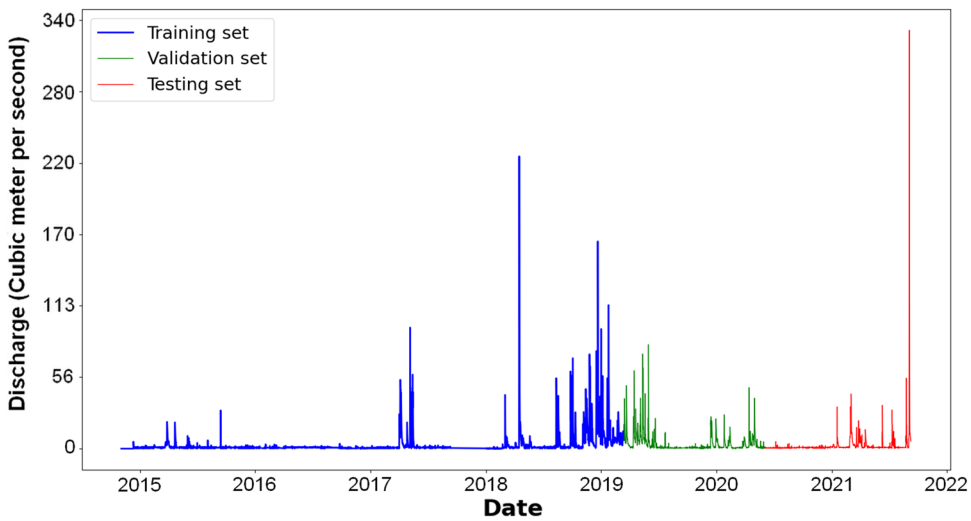

Fig. 2 Temporal distribution of Hackensack River discharge observation at New Milford used for training, validation, and test

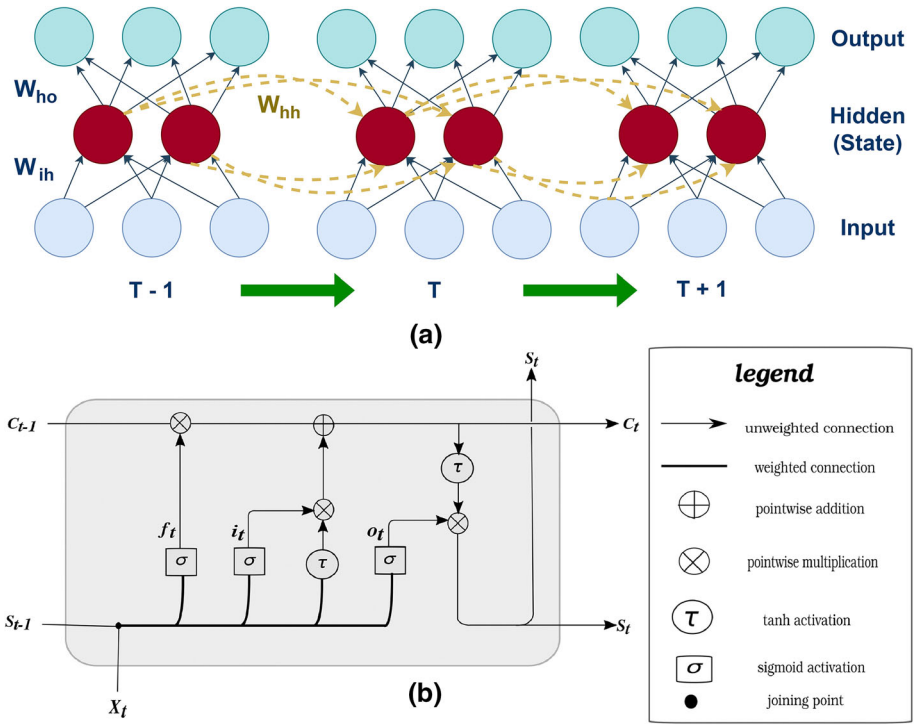

Fig. 3 **a** Processing of time sequence in RNN and **b** LSTM block, where f_t, i_t, o_t are forget, input, and output gates, respectively, adapted from [44]

$$C_t = C_{t-1} \otimes f_t + i_t \otimes \widetilde{C_t} \tag{4}$$

4. Decide what is going to be produced as output. The output gate (o_t) decides what parts of the cell state will be produced as output. Then, the cell state goes through the *tanh* layer (to push the values to be between -1 and 1) and multiplies by the output gate.

$$o_t = \sigma(X_t U^o + S_{t-1} W^o + b_o) \tag{5}$$
$$S_t = o_t \otimes \tanh(C_t)$$

where (U_f, U_i, U_o, U_c), (W_f, W_i, W_o, W_c), and (b_f, b_i, b_o, b_c) are input weights, recurrent weights, and biases, respectively. X_t, S_t and C_t are input, hidden, and cell state at time step t, respectively. S_{t-1} and C_{t-1} are the hidden and cell state at time step $t-1$, respectively. $\otimes$, $\oplus$ and σ are pointwise multiplication, pointwise addition, and sigmoid activation, respectively.

This study uses an LSTM model with a sequence input single-output (SISO) configuration. The sigmoid function calculates the output values of the hidden layer and the output layer, while the Adam algorithm is used for model training. Adam's optimization algorithm was used to correct model performance. The optimizer combines the best properties of the AdaGrad and RMSProp algorithms and can resolve sparse gradients on noisy problems such as ours. The NSE metric is taken as the objective function. The performance of the LSTM model depends on its configuration parameters. For instance, the window size is crucial for the model's best performance and counts as an additional hyperparameter that requires tuning. The risk of losing long-term precipitation information and the long-term impact cannot be captured inside the time window.

The input variables comprise MRMS rainfall values that correspond to varying lead times, ranging from none, 6, 12, 18, 24, 30, 36, 42, 48, 54, 60, 66, and 72 h. The MRMS rainfall data product is quality controlled, and biases are mitigated using rain gauge corrections. MRMS is used in this study with the introduced lag times to mimic the quantitative precipitation forecasts and infer reservoir management rules without the bias that uncertainty in rainfall prediction could introduce. The interest in this study is to determine the most relevant lead times of precipitation forecasts that are driving discharge releases and not necessarily quantifying the impact of uncertainty in the precipitation forecasts and their impact on decision-making. Using the bias-corrected, high-resolution MRMS product minimizes the possible contribution of uncertainty in quantitative precipitation forecasts. It is worth noting that the use of actual precipitation forecasts from coarser and mid-range products like the Global Forecast System (GFS) model, which the reservoir operators usually use, is not supposed to change the determined reservoir management rules dramatically. Their bias with respect to MRMS, if any, should not lead to a significant change in the simulated streamflow and the reservoir inflow, especially during extreme events, which are usually well predicted days in advance. In addition, the LSTM input variables

include the water level in the reservoir as a proxy for water storage held by the Oradell dam.

A feature analysis that includes the various lead times should reveal the specific lead time that is more influential in deciding the reservoir releases. This is conducted by analyzing features' importance and their weights in the LSTM model. It is expected that the relationship between the lead time of the predicted rainfall event and the decision to release water from the reservoir varies depending on the magnitude of the rainfall event. This analysis should reveal which precipitation forecast lead time drives the release of streamflow and, therefore, how early the reservoir manager decides to release water ahead of predicted events. Heavy rainfall events can appear in the forecast days in advance. However, the decision to release water from the reservoir to leave room for the resulting inflow should be taken by the operator prior to the occurrence of the event. Reservoir operators must manage competing requirements. On the one hand, they have to minimize the lead time to reduce the uncertainty of the precipitation forecast. On the other hand, they need to leave enough time to partially empty the reservoir, which requires time. This is especially true in the case of consecutive events with short intervals separating them and operational constraints that limit the evacuation of water volume. The operators of the reservoirs balance these two competing requirements and determine the optimum timing to perform a release.

In order to tune the LSTM models, four primary hyperparameters and model structure parameters were used: (1) The batch size, which refers to the number of training examples utilized in one iteration, (2) the epoch, which indicates the number of passes of the entire training dataset the ML algorithm has completed, (3) the number of units, which means the dimension of the inner cells in LSTM and (4) the dropout, which is a regularization technique for reducing overfitting in ANNs by preventing complex co-adaptations on training data. In order to determine the LSTM weights, the values are discovered via the ADAM (Adaptive Moment Estimation) optimization procedure. The algorithm can be used instead of the classical stochastic gradient descent procedure to update the weights iteratively based on training data by calculating an exponential moving average of the gradient and squared gradient. To control the amount of the updated weights, the learning rate is set to 0.001.

2.4 Model validation metrics

In this study, different statistical metrics were used to evaluate the performance of the proposed model. Table 1 presents the metrics employed in this study. Observed and simulated streamflow data are represented by O and S, respectively.

3 Results and discussion

The performance of the LSTM model with respect to different lead times is summarized in Table 2. The performance metrics of the test datasets reported in Table 2 correspond to the testing period from June 2020 through September 2021. The results reported in Table 2 show that the best performances of the model were obtained with the 12- and 48-h lead times, with slightly better performance in the case of the 48-h lead time. The other models that use different lead time values also showed an acceptable agreement between the simulated streamflow values and the observed ones, but they still underperform compared to the 48-h lead time case.

The higher performance with some specific lead times may be dictated by the magnitude of rainfall event in the forecast, which drives the decision to release water from the reservoir or not. Major and high magnitude rainfall events often have strong signals and can be forecast days in advance. The information on the magnitude of the event in the forecast is also compounded with the available storage in the reservoir. Releasing water from the reservoir to increase the active storage capacity may take hours to a few days. This is usually required when the available storage capacity is insufficient to store the predicted total runoff.

An examination of the feature impact graph provided in Fig. 4 suggests that the Hackensack River discharge value at Rivervale, the station right upstream of the reservoir, is among the most influential variables that affect the determination of New Milford station discharge using the LSTM, followed by the Hackensack River West Nyack, Pascack Brook Westwood, Oradell water level, and the 48-h lead time precipitation values. In order to gain a deeper understanding of how precipitation lead times affect output, the weight importance of all 13 lead time precipitation values was evaluated. Figure 4 summarizes the results of the feature analysis using the SHapley Additive exPlanations (SHAP), which can be defined as a game-theoretic approach that explains the output of any machine learning model by connecting optimal credit allocation with local explanations based on the traditional Shapley values from game theory and their related extensions [46]. The results show that the 48-h lead time rainfall accumulations have the highest weight importance, followed by the 12-h lead time.

The tuning of the LSTM model hyperparameters using Adam optimized to determine the following optimum values for batch, epoch, units, and dropout, which correspond to 4096, 300, 200, and 0.2, respectively. Figure 5 summarizes the models' average loss value and its envelope versus the number of epochs. All models converge to

Table 1 Model evaluation metrics

Metric	Acronym	Equation
Peak ratio	PR	$\frac{\mathrm{Max(O)}-\mathrm{Max(S)}}{\mathrm{Max(O)}}$
Mean absolute error	MAE	$\frac{\sum_i^n \lvert S_i - O_i \rvert}{n}$
Coefficient of determination	R^2	$1-\frac{\sum_i \left(O_i-\widehat{O}\right)^2}{\sum_i \left(O_i-\overline{O}\right)^2}$
Mean squared error squared	MSE^2	$\left(\frac{1}{n}\sum_i (O_i - S_i)^2\right)^2$
Mean squared error	MSE	$\frac{1}{n}\sum_i (O_i - S_i)^2$
Nash–Sutcliffe model efficiency coefficient	NSE	$1-\frac{\sum_i (O_i-S_i)^2}{\sum_i \left(O_i-\overline{O}\right)^2}$
Kling-Gupta efficiency	KGE	$1-\sqrt{(R-1)^2+\left(\frac{\overline{O}}{std(O)}\right)^2+\left(\frac{\overline{S}}{std(S)}\right)^2}$
Correlation coefficient	R	$\frac{\sum (S-\overline{S})(O-\overline{O})}{\sqrt{\sum \left(S-\overline{S}\right)^2}\sqrt{\sum \left(O-\overline{O}\right)^2}}$
Linear regression	Alpha, Beta	$O = \mathrm{Alpha} \times S + \mathrm{Beta}$

Table 2 Performance of the LSTM models

Time	Test data	PR	MAE	R^2	MSE^2	MSE	NSE	KGE	R	Alpha	Beta
0	Hurricane Ida	1.05	9.28	0.94	428.75	20.71	0.89	0.77	0.97	0.87	0.76
	Testing period	1.09	2.22	0.92	29.40	5.42	0.86	0.72	0.96	0.69	1.13
6	Hurricane Ida	0.92	11.38	0.93	490.49	22.15	0.84	0.70	0.96	0.84	0.68
	Testing period	0.89	2.76	0.92	50.29	7.09	0.76	0.55	0.96	0.56	1.33
12	Hurricane Ida	0.97	11.67	0.96	609.03	24.68	0.9	0.72	0.98	0.83	0.68
	Testing period	0.96	1.39	0.93	20.42	4.52	0.9	0.7	0.97	0.9	0.64
18	Hurricane Ida	0.77	9.69	0.94	411.76	20.29	0.88	0.84	0.97	0.89	0.84
	Testing period	0.8	3.29	0.92	31.61	5.62	0.84	0.76	0.96	0.8	− 0.14
24	Hurricane Ida	0.63	12.10	0.95	784.43	28.01	0.87	0.74	0.97	0.71	0.67
	Testing period	0.54	1.51	0.91	26.86	5.18	0.87	0.73	0.96	0.78	0.83
30	Hurricane Ida	0.85	11.88	0.92	524.57	22.90	0.90	0.77	0.96	0.94	0.73
	Testing period	0.88	1.86	0.91	21.50	4.64	0.89	0.57	0.95	0.82	0.4
36	Hurricane Ida	0.92	10.70	0.94	444.81	21.09	0.87	0.73	0.97	1.01	0.81
	Testing period	0.88	2.93	0.90	37.24	6.10	0.82	0.55	0.95	0.77	0.02
42	Hurricane Ida	0.79	13.75	0.92	641.67	25.33	0.88	0.75	0.96	0.92	0.68
	Testing period	0.65	2.03	0.91	26.38	5.14	0.87	0.63	0.96	0.87	0.36
48	Hurricane Ida	1.02	8.59	0.95	344.32	18.56	0.92	0.85	0.97	0.95	0.83
	Testing period	1.07	1.28	0.94	17.76	4.21	0.91	0.73	0.97	0.86	0.68
54	Hurricane Ida	1.09	11.28	0.94	687.71	26.22	0.91	0.68	0.97	0.71	0.72
	Testing period	1.14	2.08	0.92	24.03	4.90	0.88	0.50	0.96	0.86	0.36
60	Hurricane Ida	0.71	8.83	0.92	419.23	20.48	0.90	0.76	0.96	0.83	0.76
	Testing period	0.79	1.34	0.90	23.56	4.85	0.88	0.77	0.95	1.01	0.74
66	Hurricane Ida	1.15	19.14	0.92	1532.16	39.14	0.88	0.67	0.96	0.6	0.67
	Testing period	1.3	3.03	0.88	30.94	5.56	0.85	0.61	0.94	0.83	1.46
72	Hurricane Ida	0.92	10.06	0.90	456.87	21.37	0.89	0.67	0.95	0.87	0.83
	Testing period	0.89	2.76	0.87	31.63	5.62	0.85	0.62	0.93	0.91	0.09

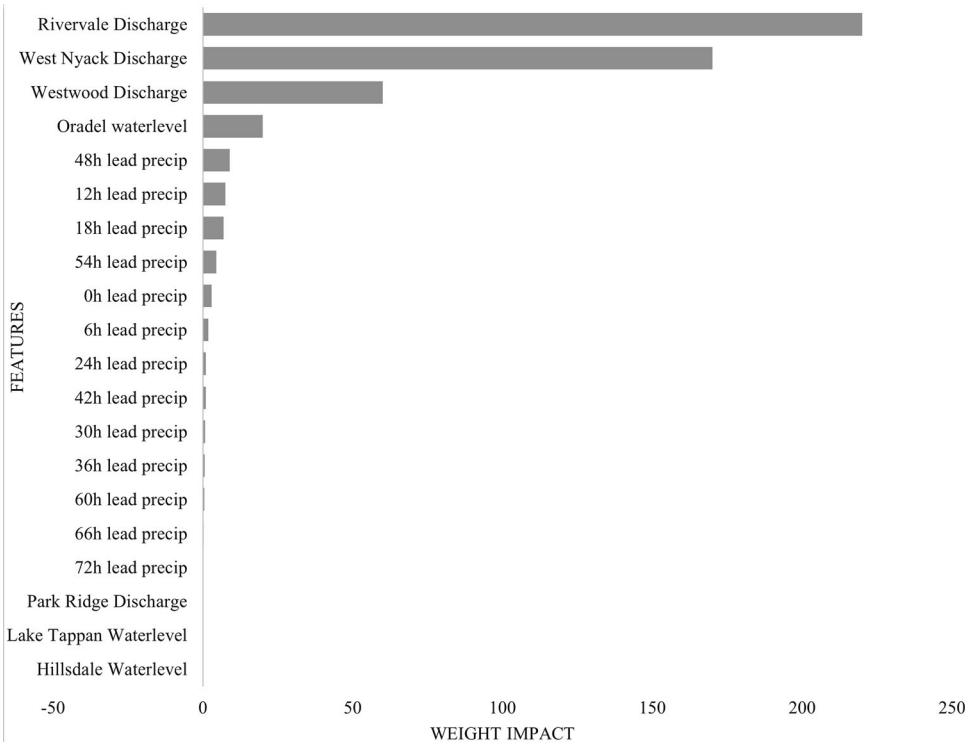

Fig. 4 Feature importance analysis

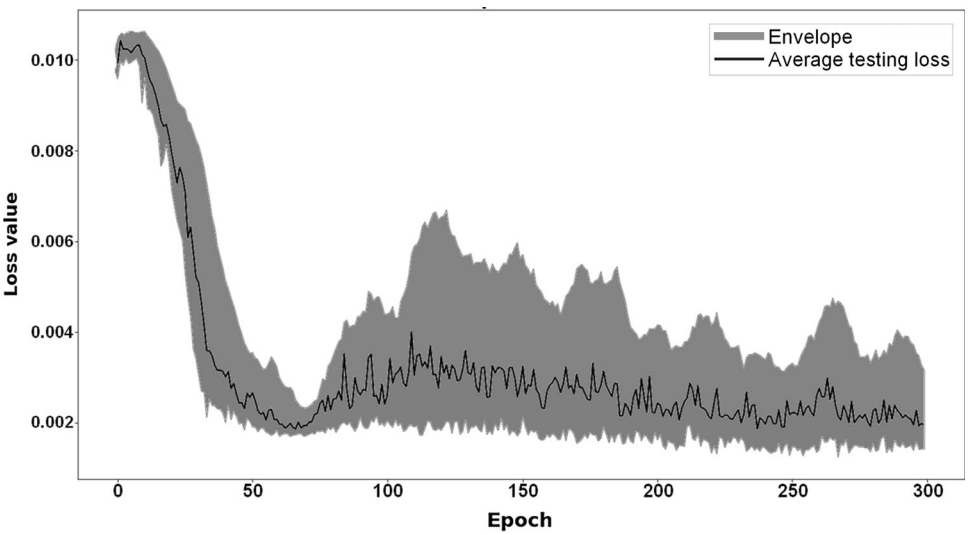

Fig. 5 Validation loss vs. epochs envelope for all 13 LSTM models

their minimums from epoch 120 to stabilize throughout epoch 300.

Figure 6 shows a comparison between the simulated streamflow at the New Milford station for each of the lead times and the observed streamflow for the 2020 and 2021 testing periods, excluding Hurricane Ida, which will be assessed separately in the subsequent section. The comparison corroborates the performance metrics reported in Table 2 and discussed above. Overall, the simulated streamflow and the observed one are in agreement. The analysis of the ten rainfall events in the figure showed that the 48-h lead time event performs better than the other models whenever there is a significant accumulation of rain (in three cases, the accumulation of rain exceeded 10 mm). During low streamflow periods, the LSTM model captured the small variability of observed flow, which should correspond to releases due to frequent rainfall events and operation rules to maintain targeted live storage in the reservoir. The LSTM model was able to simulate the peaks of streamflow for events (i.e., reservoir releases) of different magnitudes during the test period (peak ratios for models respectively 0 h, 6 h, 12 h, 18 h, 24 h, 30 h, 36 h,

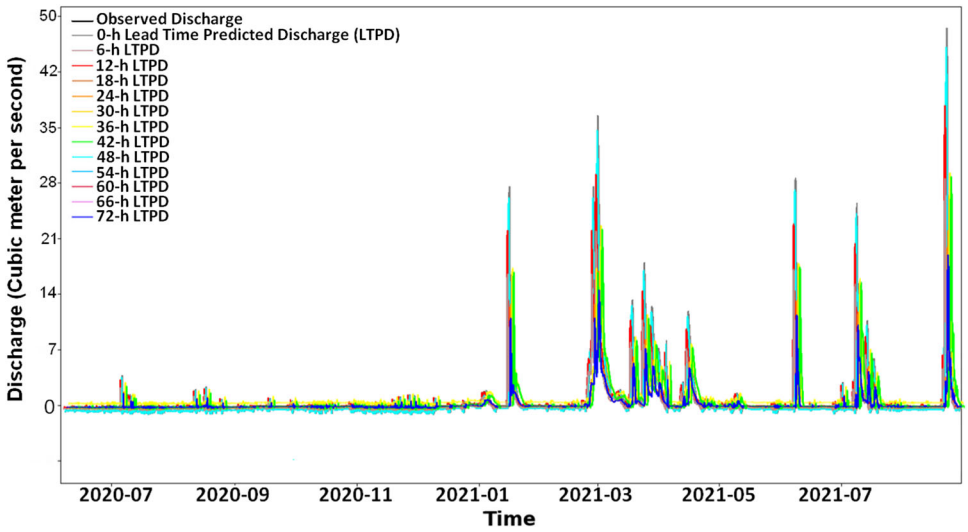

Fig. 6 Time series of simulated and observed discharge during the testing period excluding Hurricane Ida

42 h, 48 h, 54 h, 60 h, 66 h, 72 h lead times: 1.09, 0.89, 0.96, 0.8, 0.54, 0.88, 0.88, 0.65, 1.07, 1.14, 0.79, 1.3, 0.89). The model seems to perform better during more frequent events (lower magnitude) that correspond to streamflow peaks lower than ten m^3/s, possibly due to an abundance of similar events in the training sample, which can lead to better training of the LSTM and, therefore, better performance. At the same time, the performance of the LSTM model during major events is preeminent as it nearly predicted the exact value of the discharge value with an average peak ratio of 0.9.

All 13 models with a 6-h lead time difference captured the variability of the New Milford discharge value with some biases that can be attributed to the models' training and parameterization. For instance, all models exhibited a bias in the prediction of peak flow, as shown in Table 2. The 48-h scenario, considered the best-performing model, has a peak ratio of 1.07. The rest of the models slightly overestimated or underestimated the peak flow. The temporal evolution of the discharge time series is in agreement with all models' simulations. The 12- and 48-h lead time models had NSE scores of 0.91 and 0.92, MAE values of 1.39 and 1.27 m^3/s, and RMSE values of 4.52 and 4.21 m^3/s, respectively, over the testing data between 06/2020 and 09/2021.

The Hackensack watershed and the entire northeast region in the US were impacted significantly by Hurricane Ida on September 1st and 2nd, 2021. It is essential to investigate the performance of the developed LSTM model during this extreme event. Hurricane Ida caused significant flooding, which led to considerable human and economic losses. The total rainfall event over the watershed reached 177 mm in certain regions (Fig. 7). The spatial distribution of the total rainfall over the watershed from the MRMS records (Fig. 7) showed that most of the rainfall impacted the southeastern and downstream parts of the watershed. The concentration of heaviest rainfall records in MRMS around the watershed outlet fosters a rapid hydrologic response compared to other events with evenly distributed rainfall or with heavier rain over upstream locations of the watersheds.

On the other hand, the lowest precipitation occurred in the northwestern part of the watershed, with values varying between 94 and 115 mm. A rain gauge observation at Teterboro airport (Fig. 7) recorded a total rainfall of 172 mm over the same record period between 3 pm on September 1st, 2021, and 3 am on September 2nd, 2021, retrieved from MRMS observation. The results showed a good agreement between remote sensing-based measurements and in situ observations. According to NOAA Atlas 14 Point Precipitation Frequency Estimates (https://hdsc.nws.noaa.gov/hdsc/pfds/pfds_map_cont.html), the total rainfall of 12-h duration (matching with Ida event) in the southwestern part of the watershed that experienced a total rainfall of 177 mm, has a 100-year return period. In contrast, the northwestern part of the watershed that recorded a total rainfall of 94 mm during 12-h has a 5-year return period.

Figure 8 displays the observed streamflow values upstream and downstream of the reservoir at New Milford station and the reported rainfall in the MRMS product (with zero lag) during the hurricane. The New Milford station discharge increased 36 h prior to Ida to levels ranging between 1.05 and 3.2 m^3/s. The same goes for the other stations, especially the Pascack and Rivervale stations which also showed a steady increase in streamflow prior to Hurricane Ida, suggesting a start of a release to augment the reservoir's active storage capacity ahead of the hurricane. Twelve hours prior to the start of the rain, the early release rose to around four m^3/s and continued for

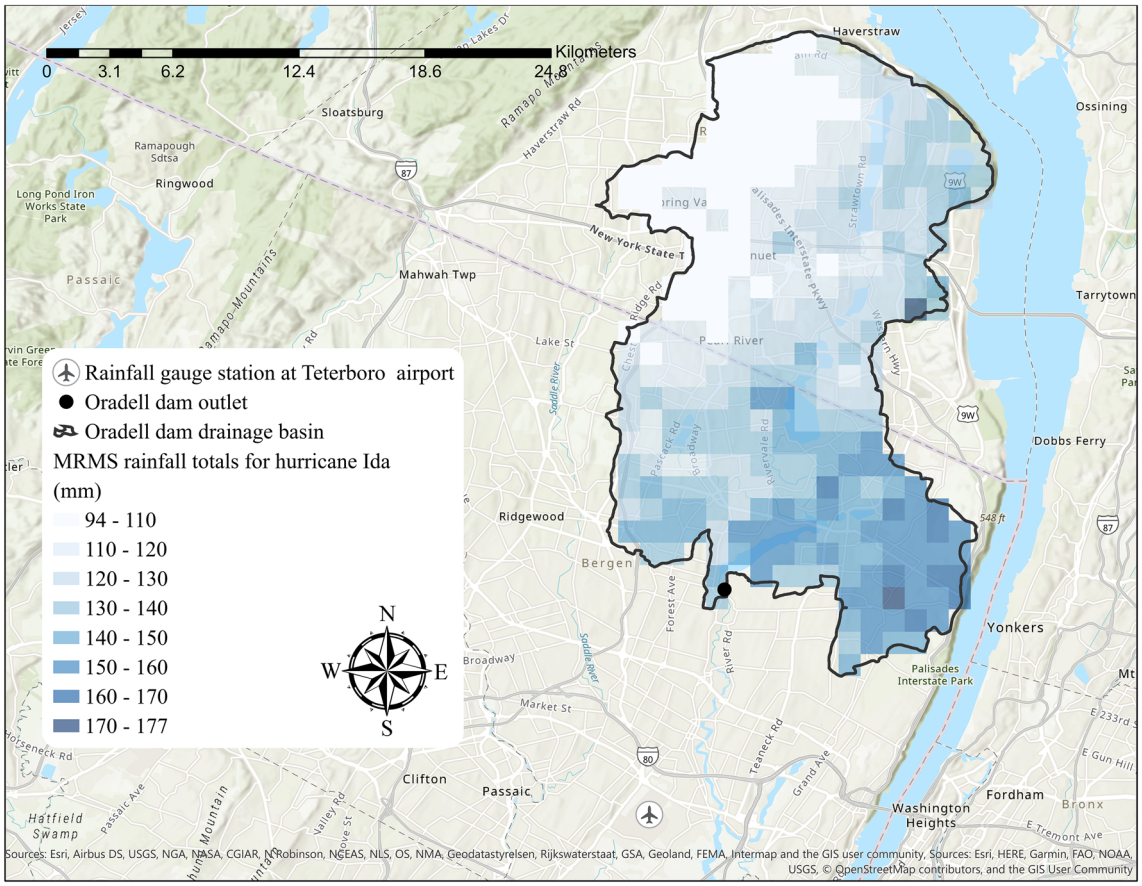

Fig. 7 Spatial distribution of MRMS precipitation during Hurricane Ida in the Hackensack Watershed

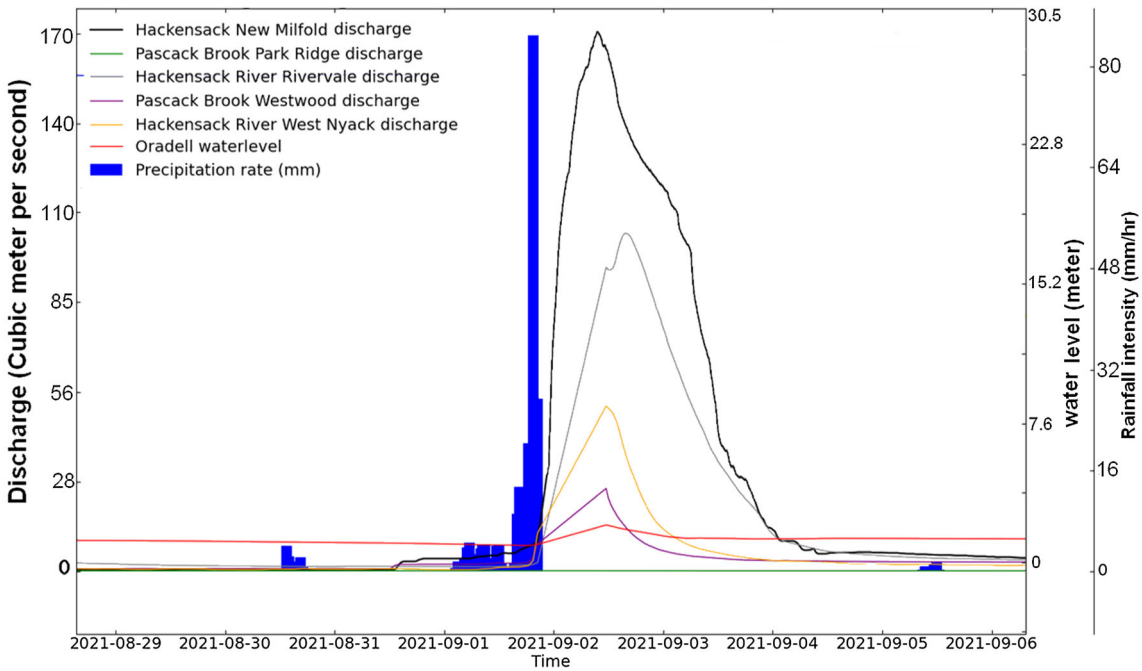

Fig. 8 Plot of upstream and downstream discharge and precipitation time series (MRMS), the water level in Oradell (before and after the rain), and predicted rain for Hurricane Ida

around 12 h which corresponds to a total volume of 172,800 m^3. The inflow to Oradell Reservoir from upstream stations over the same time was steady as well, around 1.41 m^3/s which corresponds to a total volume of 61,000 m^3. The incoming volume to the reservoir is smaller than the releases suggesting a decrease in water storage in the reservoir ahead of Hurricane Ida. As rainfall intensity increases from 18 to 85 mm/h on average over the study area, discharge from the reservoir and upstream stations progressively increases to control the water level in the reservoir and mitigate flood damage.

The peak discharge values at the same station occurred approximately 12 h after the peak precipitation values over the watershed (Fig. 8), which can explain the high performance of the model in the case of the 12-h lead time, which is comparable to the performance of the 48-h lead time models. The severity of Hurricane Ida and the concentration of the heavy rainfall around the outlet triggered a rapid response of the watershed, around 12 h. The testing period includes events that are less significant than Ida, with rainfall events that are not necessarily centered around the outlet but more distributed over the watershed. The response time of the watershed tends to be longer in such conditions, and therefore, the 48-h lead time was slightly better performing during the testing period. In both cases, the LSTM model seems to be driven by the watershed response time as it properly captures its change depending on the magnitude of the event.

Figure 9 compares the observed streamflow at New Milford to the simulated ones for each considered scenario. The early increase in the discharge on August 31st that corresponds to planned releases in preparation for the predicted event was partially captured by the LSTM models. All LSTM models showed a start of an increase in the discharge toward the end of the day, around midnight, which comes after the actual start in the observed data. However, it is prior to the start of the rainfall event suggesting that the models may have captured the proactive management of the reservoir. The steady release prior to Ida was lower than the peak flow of the extreme event but steady enough over more than 24 h to be correctly captured by the LSTM models. Nevertheless, all the developed models were able to capture the rapid increase in the discharge value as a result of the incoming runoff to the reservoir. Although all models performed well, the 48-h lead time model was the closest to predicting the actual discharge. This implies that the LSTM model was able to reproduce the reservoir operations and the releases prior to the event and the rainfall-runoff transformation to eventually simulate the total hydrograph at New Milford station.

The analysis of the simulated discharge value of the New Milford station during Hurricane Ida shows that most models performed well as NSE varied between 0.9 and 0.95 (Table 2). The best models that captured the streamflow variability during Hurricane Ida were the 12- and 48-h lead time LSTM models. Overall, the validation metrics for Ida simulations showed better performance than during the testing period (Fig. 10).

This implies that streamflow at New Milford station and releases from Oradell Reservoir is mostly driven by 48-h lead time, especially in the case of frequent events which are not as severe as Hurricane Ida. The findings suggest that same-day precipitation records do not significantly influence the reservoir releases from the Oradell reservoir and that its releases are conducted in anticipation of predicted precipitation 48-h in advance unless it is an extreme event where the 12-h lead time seems to be more influential, especially when most of the rainfall occurs near the outlet as it was the case during Hurricane Ida. This is common, especially in reservoirs used for flood protection,

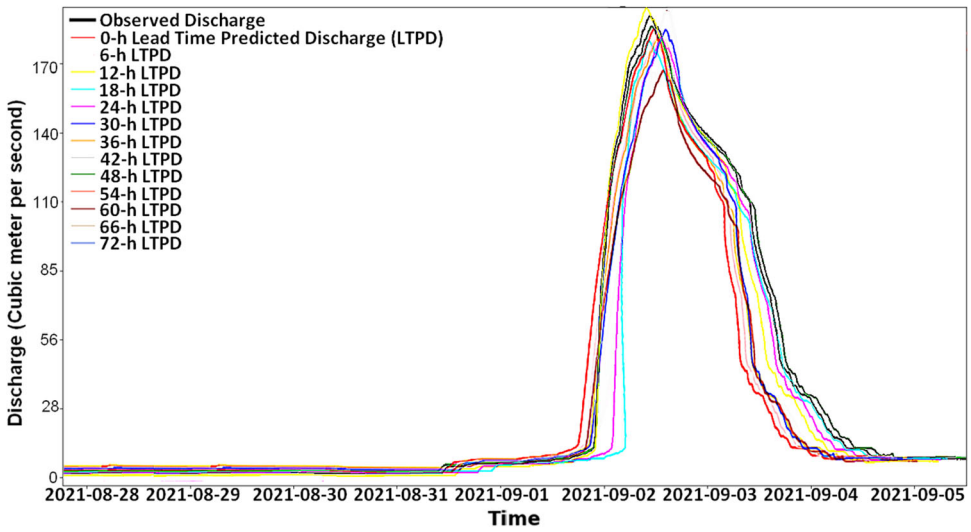

Fig. 9 Time series of simulated and observed discharge during Hurricane Ida

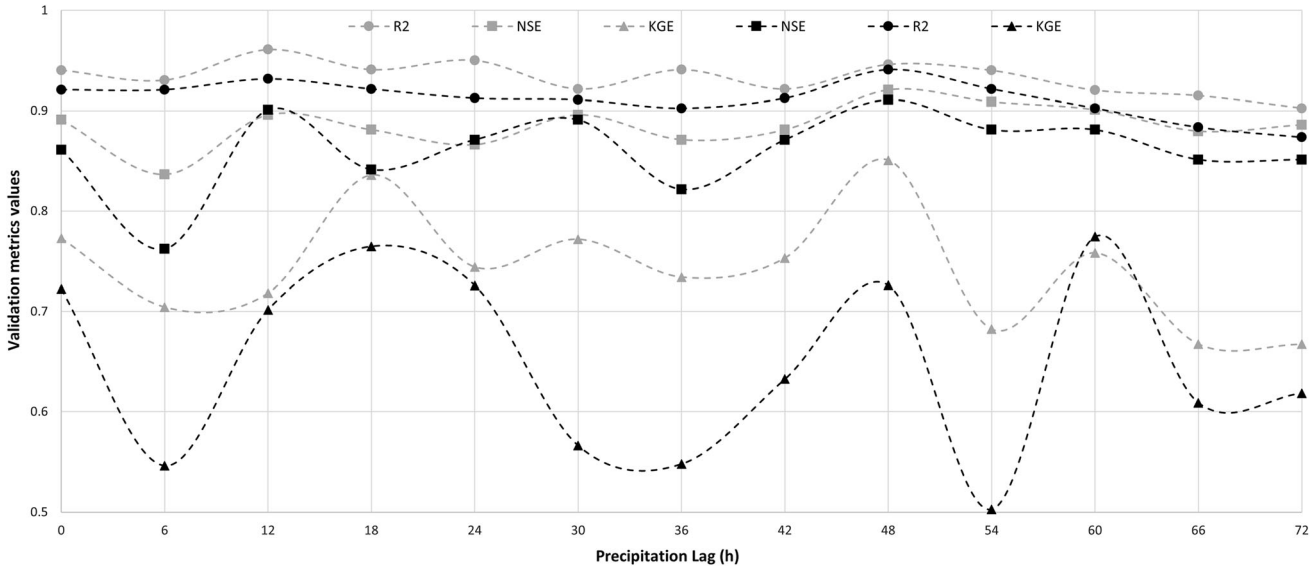

Fig. 10 Evaluation of the LSTM model performance with respect to precipitation lag: the dashed black lines and the dashed grey lines represent the test period validation metrics and hurricane Ida validation metrics, respectively

which are usually operated based on rainfall forecast and where releases are planned, depending on predicted events. The prediction of rainfall provides forecasters and managers with the amount of the expected rainfall as well as its spatial distribution over the watershed. These are two key pieces of information that reservoir managers use to plan their releases ahead of upcoming events. The lead time of the predicted rainfall that drives the decision of water release should be site-specific. In the case of the Oradell Reservoir, the 48-h one seems to have more influence overall on the releases compared to other lead times with comparable performance with the 12-h lead time during Hurricane Ida. Hence, the managers of Oradell reservoir seem to rely mostly on a 48-h precipitation forecast to guide their release strategy as the best agreement was obtained with this lead time that mimics a 48-h precipitation forecast. The determining lead times are site-specific and cannot be applied to other watersheds and reservoirs without proper training in the LSTM model. Nevertheless, the technique is expandable to other sites where discharge observations are available, but reservoir management rules are unknown.

4 Conclusions

This study uses a deep learning approach to predict streamflow and infer reservoir operating rules. An LSTM model was used to forecast discharge values at the New Milford USGS station downstream of the Oradell Reservoir in northern New Jersey. The model was trained on the discharge value of the upstream USGS stations and the MRMS precipitation data to predict the downstream discharge value at the New Milford station, right downstream of the Oradell reservoir. A set of precipitation lead times were introduced as a proxy for the forecast times that drive the decision to release streamflow from the reservoir. The streamflow simulation obtained from the LSTM model showed better performance, mostly with the 48-h lead time and the 12-h lead time, especially in the case of Hurricane Ida. The obtained results showed high performance with an R^2 of 0.94 and 0.93 at the 48- and 12-h lead times, respectively. The results were consistent with those obtained during Hurricane Ida, an extreme event that impacted the Northeast in 2021, which was used as a validation event. The obtained determination coefficients (R^2) in the case of Hurricane Ida were 0.95 and 0.96 for the 48- and 12-h lead times, respectively.

This higher performance with these specific lead times implies that discharge releases from the Oradell reservoir are mostly influenced and therefore driven by a 48-h precipitation forecast. The 12-h lead time seems to prevail in the case of extreme events. In addition, during the case of hurricane Ida where the heavy rainfall was mostly concentrated near the outlet, the 12-h lead time matched with the watershed lag time, also known as the response time. During Ida, most of the heavy rainfall was recorded in the southern part of the watershed, near the outlet, which fostered a shorter response time. This suggests that future work should account for the spatial pattern of rainfall over a watershed as an additional input variable that may influence the predicted discharge downstream. The proposed method can be integrated into decision support systems focused on managing water resources in

watersheds with unknown reservoir management rules. In future work, the study can be expanded geographically to integrate other reservoirs used in CONUS-wide models like the NOAA National Water Model [47] or apply the method to ungauged watersheds where streamflow observations and knowledge of reservoirs' operation rules are scarce.

Acknowledgements The authors acknowledge the financial support received from the Port Authority of New York and New Jersey.

Declaration

Conflict of interest No potential conflict of interest to declare.

References

1. Chen Y, Liu R, Barrett D et al (2015) A spatial assessment framework for evaluating flood risk under extreme climates. Sci Total Environ 538:512–523. https://doi.org/10.1016/j.scitotenv.2015.08.094
2. Sharifan RA, Roshan A, Aflatoni M et al (2010) Uncertainty and sensitivity analysis of SWMM model in computation of Manhole water depth and subcatchment peak flood. Procedia Soc Behav Sci 2:7739–7740. https://doi.org/10.1016/j.sbspro.2010.05.205
3. Wang K-H, Altunkaynak A (2012) Comparative case study of rainfall-runoff modeling between SWMM and fuzzy logic approach. J Hydrol Eng 17:283–291. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000419
4. Krebs G, Kokkonen T, Valtanen M et al (2013) A high resolution application of a stormwater management model (SWMM) using genetic parameter optimization. Urban Water J 10:394–410. https://doi.org/10.1080/1573062X.2012.739631
5. Sahoo GB, Ray C, de Carlo EH (2006) Calibration and validation of a physically distributed hydrological model, MIKE SHE, to predict streamflow at high frequency in a flashy mountainous Hawaii stream. J Hydrol 327:94–109. https://doi.org/10.1016/j.jhydrol.2005.11.012
6. Loi NK, Liem ND, Tu LH et al (2019) Automated procedure of real-time flood forecasting in Vu Gia–Thu Bon river basin, Vietnam by integrating SWAT and HEC-RAS models. J Water Clim Change 10:535–545. https://doi.org/10.2166/wcc.2018.015
7. Goodall J, Morsy M, Sadler J (2017) Real-time flood prediction using data-driven and hydrodynamic modeling tools. Model Manag Extrem Precip 10:535–545
8. Follum ML, Tavakoly AA, Niemann JD, Snow AD (2017) AutoRAPID: a model for prompt streamflow estimation and flood inundation mapping over regional to continental extents. JAWRA J Am Water Resour Assoc 53:280–299. https://doi.org/10.1111/1752-1688.12476
9. Wang X, Kinsland G, Poudel D, Fenech A (2019) Urban flood prediction under heavy precipitation. J Hydrol 577:123984. https://doi.org/10.1016/j.jhydrol.2019.123984
10. Teng J, Jakeman AJ, Vaze J et al (2017) Flood inundation modelling: a review of methods, recent advances and uncertainty analysis. Environ Model Softw 90:201–216. https://doi.org/10.1016/j.envsoft.2017.01.006
11. Solomon S, Plattner G-K, Knutti R, Friedlingstein P (2009) Irreversible climate change due to carbon dioxide emissions. Proc Natl Acad Sci 106:1704–1709. https://doi.org/10.1073/pnas.0812721106
12. Mirzaei M, Huang YF, El-Shafie A, Shatirah A (2015) Application of the generalized likelihood uncertainty estimation (GLUE) approach for assessing uncertainty in hydrological models: a review. Stoch Env Res Risk Assess 29:1265–1273. https://doi.org/10.1007/s00477-014-1000-6
13. Saleh F, Ramaswamy V, Wang Y et al (2017) A multi-scale ensemble-based framework for forecasting compound coastal-riverine flooding: the Hackensack-Passaic watershed and Newark Bay. Adv Water Resour 110:371–386. https://doi.org/10.1016/j.advwatres.2017.10.026
14. Nourani V, Komasi M (2013) A geomorphology-based ANFIS model for multi-station modeling of rainfall–runoff process. J Hydrol 490:41–55. https://doi.org/10.1016/j.jhydrol.2013.03.024
15. Sudheer Ch, Maheswaran R, Panigrahi BK, Mathur S (2014) A hybrid SVM-PSO model for forecasting monthly streamflow. Neural Comput Appl 24:1381–1389. https://doi.org/10.1007/s00521-013-1341-y
16. Rasouli K, Hsieh WW, Cannon AJ (2012) Daily streamflow forecasting by machine learning methods with weather and climate inputs. J Hydrol 414–415:284–293. https://doi.org/10.1016/j.jhydrol.2011.10.039
17. Yaseen ZM, Jaafar O, Deo RC et al (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
18. Adnan RM, Liang Z, Trajkovic S et al (2019) Daily streamflow prediction using optimally pruned extreme learning machine. J Hydrol 577:123981. https://doi.org/10.1016/j.jhydrol.2019.123981
19. Kisi O, Cimen M (2011) A wavelet-support vector machine conjunction model for monthly streamflow forecasting. J Hydrol 399:132–140. https://doi.org/10.1016/j.jhydrol.2010.12.041
20. Fang W, Huang S, Ren K et al (2019) Examining the applicability of different sampling techniques in the development of decomposition-based streamflow forecasting models. J Hydrol 568:534–550. https://doi.org/10.1016/j.jhydrol.2018.11.020
21. Rathinasamy M, Adamowski J, Khosa R (2013) Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method. J Hydrol 507:186–200. https://doi.org/10.1016/j.jhydrol.2013.09.025
22. Prasad R, Deo RC, Li Y, Maraseni T (2017) Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. Atmos Res 197:42–63. https://doi.org/10.1016/j.atmosres.2017.06.014
23. Alvisi S, Franchini M (2011) Fuzzy neural networks for water level and discharge forecasting with uncertainty. Environ Model Softw 26:523–537. https://doi.org/10.1016/j.envsoft.2010.10.016
24. Allawi MF, Jaafar O, Mohamad Hamzah F et al (2018) Review on applications of artificial intelligence methods for dam and reservoir-hydro-environment models. Environ Sci Pollut Res 25:13446–13469. https://doi.org/10.1007/s11356-018-1867-8
25. Raman H, Chandramouli V (1996) Deriving a general operating policy for reservoirs using neural network. J Water Resour Plan Manag 122:342–347. https://doi.org/10.1061/(ASCE)0733-9496(1996)122:5(342)
26. Deka P, Chandramouli V (2003) A fuzzy neural network model for deriving the river stage—discharge relationship. Hydrol Sci J 48:197–209. https://doi.org/10.1623/hysj.48.2.197.44697
27. Chang F-J, Chang Y-T (2006) Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. Adv Water Resour 29:1–10. https://doi.org/10.1016/j.advwatres.2005.04.015
28. Nourani V, Hosseini Baghanam A, Adamowski J, Kisi O (2014) Applications of hybrid wavelet–Artificial Intelligence models in hydrology: a review. J Hydrol 514:358–377. https://doi.org/10.1016/j.jhydrol.2014.03.057

29. Shiri J, Shamshirband S, Kisi O et al (2016) Prediction of water-level in the Urmia Lake using the extreme learning machine approach. Water Resour Manage 30:5217–5229. https://doi.org/10.1007/s11269-016-1480-x
30. Emamgholizadeh S, Moslemi K, Karami G (2014) Prediction the Groundwater Level of Bastam Plain (Iran) by Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS). Water Resour Manage 28:5433–5446. https://doi.org/10.1007/s11269-014-0810-0
31. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
32. Hu J, Zheng W (2019) Transformation-gated LSTM: efficient capture of short-term mutation dependencies for multivariate time series prediction tasks. In: 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, pp 1–8
33. Mouatadid S, Adamowski JF, Tiwari MK, Quilty JM (2019) Coupling the maximum overlap discrete wavelet transform and long short-term memory networks for irrigation flow forecasting. Agric Water Manag 219:72–85. https://doi.org/10.1016/j.agwat.2019.03.045
34. Zhang D, Lin J, Peng Q et al (2018) Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. J Hydrol 565:720–736. https://doi.org/10.1016/j.jhydrol.2018.08.050
35. Kratzert F, Klotz D, Brenner C et al (2018) Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrol Earth Syst Sci 22:6005–6022. https://doi.org/10.5194/hess-22-6005-2018
36. Ni L, Wang D, Wu J et al (2020) Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. J Hydrol 586:124901. https://doi.org/10.1016/j.jhydrol.2020.124901
37. Hu C, Wu Q, Li H et al (2018) Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water (Basel) 10:1543. https://doi.org/10.3390/w10111543
38. Mignot E, Li X, Dewals B (2019) Experimental modelling of urban flooding: a review. J Hydrol 568:334–342. https://doi.org/10.1016/j.jhydrol.2018.11.001
39. Yuan X, Chen C, Lei X et al (2018) Monthly runoff forecasting based on LSTM–ALO model. Stoch Env Res Risk Assess 32:2199–2212. https://doi.org/10.1007/s00477-018-1560-y
40. U.S. Department of the Interior USGS (2022) USGS Water Data for USA https://waterdata.usgs.gov/nwis
41. PANYNJ (2013) Port Authority of New York and New Jersey Teterboro Airport 2013 Sustainability Report
42. Gourley JJ, Flamig ZL, Vergara H et al (2017) The FLASH Project: improving the tools for flash flood monitoring and prediction across the United States. Bull Am Meteorol Soc 98:361–372. https://doi.org/10.1175/BAMS-D-15-00247.1
43. Jian Zhang JG (2018) Multi-Radar Multi-Sensor Precipitation Reanalysis (Version 1.0). In: Open Commons Consortium Environmental Data Commons. https://doi.org/10.25638/EDC.PRECIP.0001
44. Sagheer A, Kotb M (2019) Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing 323:203–213. https://doi.org/10.1016/j.neucom.2018.09.082
45. Greff K, Srivastava RK, Koutnik J et al (2017) LSTM: a search space Odyssey. IEEE Trans Neural Netw Learn Syst 28:2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924
46. Mangalathu S, Hwang S-H, Jeon J-S (2020) Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Eng Struct 219:110927. https://doi.org/10.1016/j.engstruct.2020.110927
47. Office of Water Prediction (2022) The National Water Model. https://water.noaa.gov/about/nwm. Accessed 17 Jan 2022