


Article

Water Level Forecasting Using Deep Learning Time-Series Analysis: A Case Study of Red River of the North

Vida Atashi ^{1,*}, Hamed Taheri Gorji ², Seyed Mojtaba Shahabi ³ , Ramtin Kardan ⁴ and Yeo Howe Lim ¹

¹ Department of Civil Engineering, University of North Dakota, Grand Forks, ND 58202, USA; yeo.lim@und.edu

² Department of Biomedical Engineering, University of North Dakota, Grand Forks, ND 58202, USA; hamed.taherigorji@und.edu

³ School of Electrical Engineering & Computer Science, University of North Dakota, Grand Forks, ND 58202, USA; seyedmojtaba.shahabi@und.edu

⁴ Department of Mechanical Engineering, University of North Dakota, Grand Forks, ND 58202, USA; ramtin.kardan@und.edu

* Correspondence: vida.atashi@und.edu; Tel.: +1-7018852449

Abstract: The Red River of the North is vulnerable to floods, which have caused significant damage and economic loss to inhabitants. A better capability in flood-event prediction is essential to decision-makers for planning flood-loss-reduction strategies. Over the last decades, classical statistical methods and Machine Learning (ML) algorithms have greatly contributed to the growth of data-driven forecasting systems that provide cost-effective solutions and improved performance in simulating the complex physical processes of floods using mathematical expressions. To make improvements to flood prediction for the Red River of the North, this paper presents effective approaches that make use of a classical statistical method, a classical ML algorithm, and a state-of-the-art Deep Learning method. Respectively, the methods are seasonal autoregressive integrated moving average (SARIMA), Random Forest (RF), and Long Short-Term Memory (LSTM). We used hourly level records from three U.S. Geological Survey (USGS), at Pembina, Drayton, and Grand Forks stations with twelve years of data (2007–2019), to evaluate the water level at six hours, twelve hours, one day, three days, and one week in advance. Pembina, at the downstream location, has a water level gauge but not a flow-gauging station, unlike the others. The floodwater-level-prediction results show that the LSTM method outperforms the SARIMA and RF methods. For the one-week-ahead prediction, the RMSE values for Pembina, Drayton, and Grand Forks are 0.190, 0.151, and 0.107, respectively. These results demonstrate the high precision of the Deep Learning algorithm as a reliable choice for flood-water-level prediction.

Keywords: flood prediction; water level; Red River; SARIMA; RF; LSTM



Citation: Atashi, V.; Gorji, H.T.; Shahabi, S.M.; Kardan, R.; Lim, Y.H. Water Level Forecasting Using Deep Learning Time-Series Analysis: A Case Study of Red River of the North. *Water* **2022**, *14*, 1971. <https://doi.org/10.3390/w14121971>

Academic Editors: Zhijia Li and Cheng Yao

Received: 23 April 2022

Accepted: 17 June 2022

Published: 20 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forecasting the water levels in rivers and lakes is critical for flood warning and water-resource management.

Since water-level data from hydrological stations typically have a time series structure, researchers typically employ time-series hydrological-prediction models to forecast future data. Hidden information can be revealed by using past data to predict future water levels (future behavior), which is important for mitigating flood effects, reducing or preventing disasters, and managing water resources.

1.1. History of Flood in Red River of the North

Red River discharge varies annually and seasonally, and the water demand of the Red River basin may rise in the future due to a variety of factors, including economic development, population growth, and climate change [1]. Patterns in seasonal and annual

streamflow in the basin reflect variability in precipitation. Floods happen in the Red River when the water level increases over the tops of riverbanks, due to significant precipitation over the same area for long periods, in the forms of persistent thunderstorms, rain, or snow combined with spring snow melt and ice jam. Due to a long and severe winter for snow accumulation, warmer temperatures in the spring, and flat topography with weak permeability soil, the mid-latitude regions of North America are highly vulnerable to spring-melt floods [2–4]. Spring-melt floods are frequent in the Red River as it heads north [5,6]. During the spring thaw, the southern part of the Red River basin melts first, and the river becomes hydrologically active; meanwhile, the northern part of the basin is often frozen. Along with the flat and homogenous topography, the river activity forms a slow, meandering river, which causes an overflow in the Red River of the North on the northern side, resulting in floods [5,7]. Surface runoff from snowmelt during significant floods leads the Red River to overflow its shallow banks, flooding the whole valley and causing immense damage. Research by Hirsch and Ryberg (2012) and Rice et al. (2015) indicate that the frequency of floods in the Red River basin is increasing dramatically [8,9]. Early flood forecasting can help provide communities with early warnings about protecting homes and lands as well as mitigating the impact of floods. With this introduction, there is an increasing need to improve the characterization and identification of precursors, which affect the hydrological conditions that cause spring-snowmelt floods and improve predictions to reduce Red River flood damage.

1.2. Method Used Previously for Flood Prediction

In general, there are three methods for forecasting streamflow. The first approach mainly depends on physically based models [10] that have long been used to forecast hydrological events, including storms [11,12], runoff or rainfall [13,14], shallow streamflow [15], hydraulic models [16,17], and more cases of global circulation [18], encompassing the interaction between atmosphere, water, and floods [19]. Although physical models are capable of forecasting a broad range of flooding situations, they typically require a variety of hydro-geomorphological-monitoring datasets, which necessitates costly computing and prevents short-term prediction [20]. Moreover, the construction of physically based models frequently demands in-depth knowledge and expertise in hydrological factors, which has been noted as challenging [21]. Furthermore, many types of research demonstrate that there is a gap in the short-term prediction capability of physical models [19].

In the second approach, mathematical models are used to model the streamflow hydrodynamic. Since this approach is based on original hydrological and hydraulic principles, this alternative is broadly utilized in nations across the world. Flood-modeling studies have utilized physically based hydrologic models, such as the Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) [22], the soil and water assessment tool (SWAT) [23], IHACRES [24], and the HSPF model [25], which all have been engaged in flood-modeling studies. However, using these models necessitates substantial field observations as well as trial and error parameterization techniques [26]. They still only supply at-site flood-risk estimates based on local streamflow data obtained at gauging hydrometric stations, making them inappropriate for regional-flood assessment [27,28].

The last approach is data-driven and is based on the statistical relationship between input and output data for near-future predictions. The Machine Learning (ML) model, which has been applied in flood forecasting since the 1990s, is one of the most popular frameworks utilized in the data-driven method. ML models can offer a powerful solution for flooding prediction without explicitly knowing such nonlinear dynamic processes, in contrast to a physically based numerical model [29].

Numerous research has been conducted to predict the water levels in rivers, lakes, and other water bodies worldwide using different time-series models. The Autoregressive Integrated Moving Average (ARIMA) model is widely used for river discharge and flood forecasting [30–36]. Yürekli et al., presented a monthly streamflow forecasting method for three gauging stations in the north Anatolia fault line and evaluated the residuals

of the ARIMA model [30]. The authors state that a comparison of monthly mean and standard deviation for observed and anticipated data using the ARIMA model reveals that the anticipated values maintained the main statistical features of the observed data. By comparing the observed and anticipated monthly data sequences using linear regression, they discovered a statistically significant linear relationship between the observed and anticipated monthly data. In another study by [31], data from two Schuylkill River stations in Berne and Philadelphia (in the United States) were collected over six years. The author demonstrated that daily data have no seasonality; therefore, there was no seasonality in the proposed ARIMA formulation. Even though both stations are located along the same river, the proposed ARIMA models provided for each station differed due to the differing watershed coverage. Exponential smoothing was employed by [37] to study and predict the water-level trends in the Mtera dam in Tanzania. They discovered that the water level in the Mtera dam has been declining over time, and the highest and lowest water levels were both showing a declining trend in recent years. Additionally, estimates for the next five years based on the exponential smoothing of time-series data revealed that the water level would be below the lowest water level required for energy production in the spring of 2023. The authors evaluate the efficiency and the accuracy of several models for predicting Tanshui River water levels in Taiwan during 50 historical typhoon events that occurred over 11 years between 1996 and 2007. The authors compared three eager models, including artificial neural network (ANN), linear regression (REG), and support vector regression (SVR), with two lazy models, including locally weighted regression (LWR) and the k-nearest neighbor (kNN). According to the results, ANN and SVR outperformed REG among eager-learning models. However, the authors state that although ANN, SVR, and REG were considered eager-learning models, their prediction capabilities differed due to different learning optimizers. In their results, among lazy-learning models, LWR outperformed kNN, and both lazy models showed more accurate predictions than the REG eager model.

To our knowledge, no previous studies have explicitly applied a classical statistical method, a classical ML algorithm, and a state-of-the-art Deep Learning method to achieve improvements to flood prediction in the Red River of the North. The goal of our study is to apply three models: SARIMA (a conventional statistical model), RF (a classical ML algorithm), and LSTM (a Deep Learning method) to map flood susceptibility and distinguish flood-hazard regions in the Red River of the North. The findings of this study will assist regional and local authorities as well as policymakers in mitigating flood risks and developing appropriate mitigation measures to minimize potential damages. The observed water level of the Red River of the North in three United States Geological Survey (USGS) stations, at Pembina, Drayton, and Grand Forks, sampled hourly from 2007 to 2019, are used to evaluate the water level for six hours, twelve hours, one day, three days, and one week in advance at the Red River the North. Pembina is the downstream forecasting point, but it only has a water-level station. Both Drayton and Grand Forks have a full discharge-measurement station that provides water-level and discharge series.

2. Materials and Methods

2.1. Study Area

The Red River basin is an international, multi-jurisdictional watershed of 116,550 square kilometers (45,000 square miles), with 80% of the basin in the United States and 20% in Manitoba, Canada. With a drainage area of 45,000 mi² (104,100 km²), it is a unique basin that flows from the south of the region northward into Canada through Pembina (Figure 1a) [38]. The basin itself is approximately 60 miles (97 km) at its widest point and 315 miles (507 km) in length. The climate can be characterized as semi-arid, with cold winters and dry summers. With a length of 545 river miles, the highly sinuous topography of the north, low-sloping canal of the northern Red River stretches from Wahpeton, North Dakota to Lake Winnipeg, Manitoba and marks the border between North Dakota and Minnesota (Figure 1b) [39]. Most streamflow occurs in the spring and early summer in

a typical year due to snowmelt, rainfall on the snowpack, or severe rain on saturated soil. Flooding is more common in the spring and early summer, and it is more severe during wet seasons [39]. Furthermore, the flat topography of the basin, along with the climatic conditions stated above, often leads to major floods in the Red River and its tributaries.

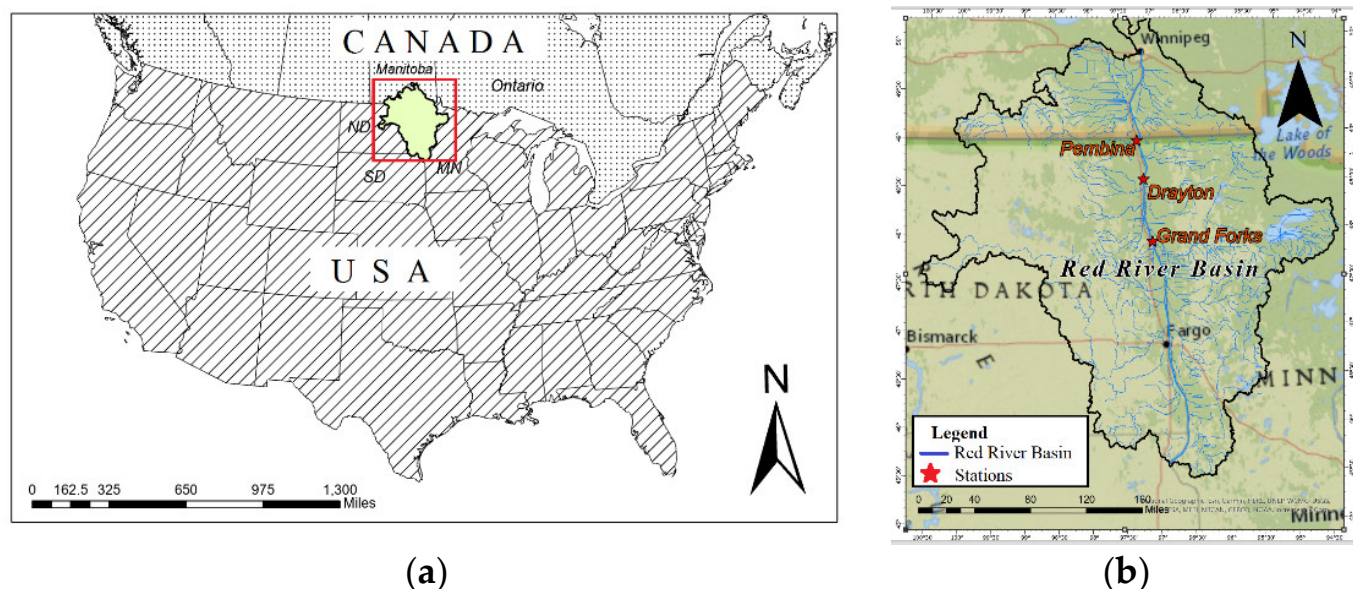


Figure 1. (a) Location of Red River basin; (b) location of USGS stations on Red River in Pembina, Drayton, and Grand Forks.

2.2. Data Representation and Pre-Processing

The Red River of the North has been chosen for four reasons. First, the river network has a limited number of minor flow-control structures, which is advantageous when using an ML technique to predict the flood. Second, the Red River presents a challenging condition for using satellite altimetry to estimate the stage. Despite its vast catchment area, the Red River runs along a main stem channel that is only around 100 m wide at the bankfull stage. Third, USGS gauging stations are adequately established along the main tributaries, offering field-based estimates of river flow and river stage for the verification of the modeling system. Finally, although the typical occurrence of devastating property losses due to floods is in years with substantial snow accumulations, the river basin, particularly the basin-scale hydrologic response to climatic variability, has not been extensively modeled. It is worth mentioning that among our three selected stations, the Pembina station on Red River, which is located downstream, does not have any data for river-flow discharge.

The characteristics of all three datasets are summarized in Table 1. Figures 2 and 3 present the monthly and annual water levels of the three selected stations, respectively. All datasets used in this paper have a monthly and annual component (see Figures 2 and 3), which make the data non-stationary.

Table 1. Characteristics of the water-level time series at three hydrology stations of the Red River.

Station No.	Station Name	Period	No. of Samples	Frequency
1	Pembina	2007–2019	104,616	Hourly
2	Drayton	2007–2019	100,140	Hourly
3	Grand Forks	2007–2019	105,117	Hourly

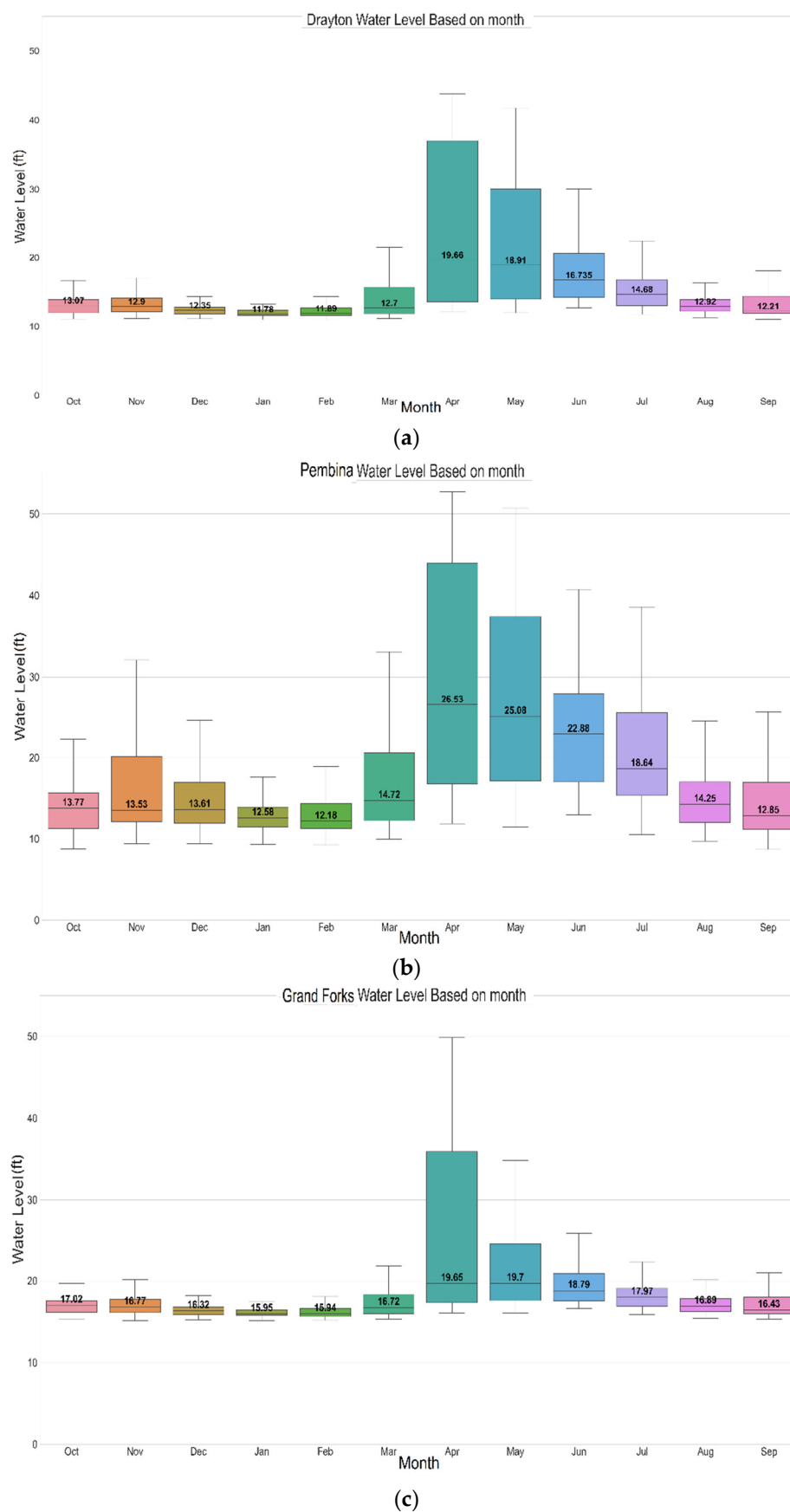
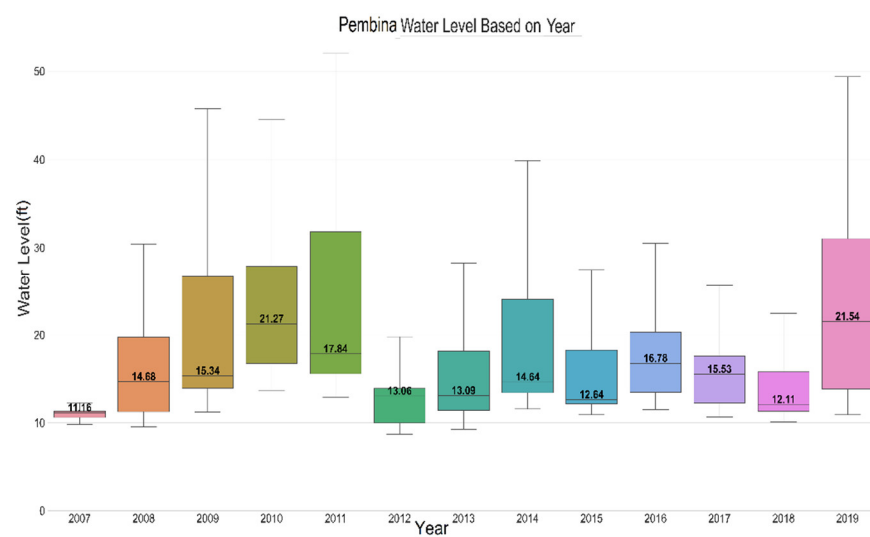
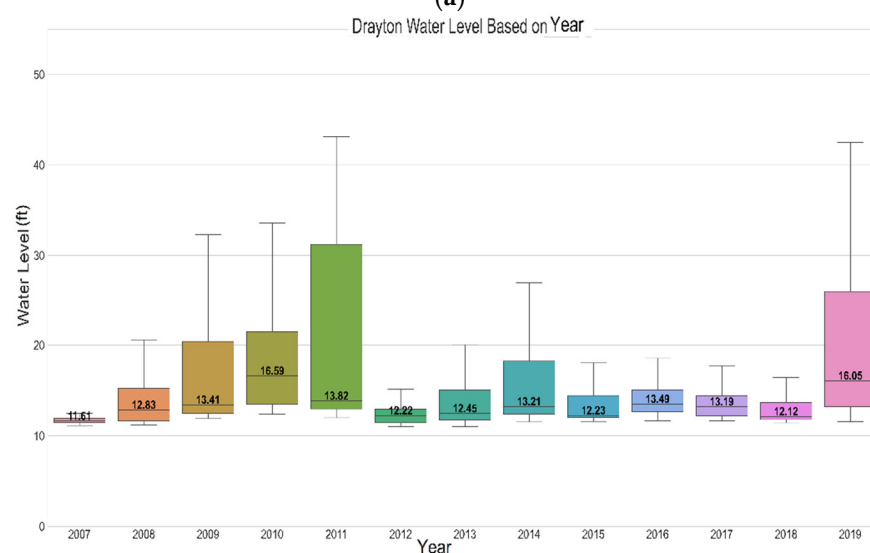


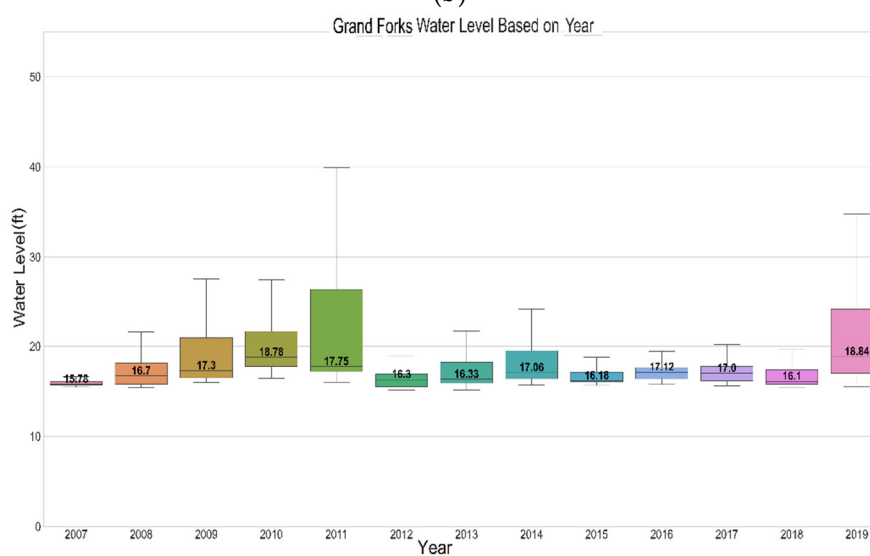
Figure 2. Monthly water level at three hydrology stations of Red River of the North: (a) Pembina, (b) Drayton, and (c) Grand Forks stations.



(a)



(b)



(c)

Figure 3. Box and whisker plot of water-level data at three hydrology stations of Red River of the North: (a) Pembina, (b) Drayton, and (c) Grand Forks stations.

The water level of these three datasets was collected from USGS hourly gauge-height record. The samples were adopted from 1 November 2007 to 31 December 2019. During data preprocessing, if the number of consecutive missing values were less than eight hours, we used the linear-interpolation technique to fill the missing data. In the case of the more than eight hours of missing values, we removed the period from our dataset.

The Pembina River, a tributary of the Red River of the North, is the major source of water in south-central Manitoba. The Pembina River flows southeast from the Turtle Mountains' highlands, beginning at its highest point (elevation 2000 feet). It joins the Red River from the west just south of Pembina, North Dakota, approximately 2 miles (3 km) south of the US—Canadian border. At Pembina, the height of the water flowing down the Red River is recorded by a stream gauge. The sensor, one of around 8000 maintained by the USGS, acts as a sentinel for communities along the river that were devastated by floods in 2009, 2010, and 2011. The Pembina gauge was targeted mainly for flood prediction because of two main reasons: first, this station is the last station on the Red River before it flows into Canada, and second, two upstream stations, Drayton and Grand Forks, have discharge information with the USGS, but Pembina station, as the downstream station, does not have any discharge information.

Based on Figure 2, April is the month with the highest streamflow at Pembina station, with an average water level of 26.53 feet. The maximum water level recorded at this station was 52.71 feet on 15 April 2009. The streamflow records for the second station, Drayton, have been continuous since 1942. Since 1970, specific-conductance measurements have been made at both stations, Drayton and Emerson, whenever discharge measurements were obtained or about once every month. This long-term data provided the information so that trends in streamflow and water quality could be examined. On 21 January 1986, streamflow measurements under ice conditions were obtained by the USGS crew on the Red River at Drayton [40]. Figure 2 shows that April is the month with the highest flow found at the Drayton station, which has an average water level of 19.66 feet. The maximum water level at the time of this study was recorded on 6 April 2009, with an average water level of 43.82 feet. Figure 2 also shows that May has the second-highest streamflow for two stations, Pembina and Drayton, with an average water level of 25.08 feet and 1891 feet, respectively.

The upstream gauge station on the Red River of the North at Grand Forks was established in 1882 by the U.S. Engineers (currently the U.S. Army Corps of Engineers). Charles M. Hall, a geology professor at North Dakota Agricultural College, installed an additional station above the original stream gauge on 26 May 1901. Hall's primary objective for the stream gauge was to investigate the possibility of storing Red River floodwaters for hydropower, irrigation, and domestic supply needs [41]. Today, this stream gauge has a continuous record of stream gauge height, discharge, stream velocity, and water quality parameters, as well as real-time web data. Figure 2 shows that May is the month with the highest streamflow found at the Grand Forks station, which has an average water level of 19.70 feet. The maximum water level at the time of this study was recorded on 6 April 2009, with an average water level of 49.84 feet.

Frequent flooding has been an issue for the Red River of the North at Grand Forks, ND, most notably the major floods of 1882, 1897, 1950, 1996, 1997, 2006, 2009, and 2011, and that is why Grand Forks stream gauge data are essential for the flood protection of the cities of Grand Forks, ND and East Grand Forks, MN.

Figure 3 shows the box and whisker plot of the water-level data annually at three hydrology stations of the Red River of the North. The maximum average of annual water-level data for all three stations occurred in 2019, which was 21.54 feet, 16.05 feet, and 18.84 feet for Pembina, Drayton, and Grand Forks stations, respectively.

Motivated by the success of the Autoregressive Integrated Moving Average (ARIMA) model [32,42,43] we used a seasonal statistical approach called the SARIMA method to capture the components of the time series separately. This method is tested on the real datasets of the Red River for hourly water level forecasting. Linear statistical models, such as SARIMA, might not be perfect at modeling the nonlinear relationships in the time series,

but it is sufficient for modeling the linear component [44]. Meanwhile, non-parametric statistical ML models, such as long short-term memory (LSTM), can model any nonlinear components (universal approximators). Furthermore, for the last method, RF was selected due to its popular use as an ML algorithm in hydrology applications [45–47]. All these three selected methods are discussed in the following section.

2.3. Seasonal Autoregressive Integrated Moving Average (SARIMA)

Seasonal Autoregressive Integrated Moving Average (SARIMA) extends the Autoregressive Integrated Moving Average (ARIMA) and is often known as Seasonal ARIMA. ARIMA combines the differencing with (Autoregressive) AR and (Moving Average) MA. In other words, in ARIMA, the “AR” indicates the relationship between a variable in time-series data and its own lagged values. The “I” represents differencing an observation’s value from the previous values to deliver stationary time-series data. The “MA” denotes the linear combination of observation and the errors of previous observations. The ARIMA model is also named non-seasonal ARIMA, and it is not suitable when time-series data include seasonal components. Hence, an extended version of ARIMA was proposed by adding the seasonal terms called Seasonal ARIMA or, in short, SARIMA. The ARIMA (p, d, q) can be represented mathematically by the following formulas:

$$\Delta^d y(t) = c + \sum_{j=1}^p \alpha_j \times y(t-j) + \epsilon(t) + \sum_{j=1}^q \beta_j \times \epsilon(t-j) \quad (1)$$

where Δ is $(1 - B)$, in which B denotes the “backward” operator and $B_y(t) = y(t-1)$, $y(t)$ shows data samples at time t , c represents the symbol for the constant value, $\alpha_1, \dots, \alpha_p$ are defined as auto-regressive parameters, the white noise at time t is defined as $\epsilon(t)$, and β_1, \dots, β_q are the moving average coefficients [32].

2.4. Random Forest

The Random Forest (RF) model is an ensemble supervised ML-algorithm technique for multiple decorrelated decision trees. We define a decision tree as a random model that relates output to elucidative variables or attributes. As a result, an individual decision tree has a set of states, which are organized and consecutively devoted to a dataset. We can grow them from stochastic resampled training batches selected from the original data to orthogonalize the trees. Numerous decision trees deliver independent numerical forecasts of the research target for regression applications, contrary to class labels for classification. Eventually, the outcome fits the mean forecast of individual trees.

The RF is a straightforward yet proper choice to tackle real-world water-science problems [48,49]. The RF requires users to determine the number of trees and the feature number of each node. Moreover, the RF model is not sensitive to these two factors and does not require fine-tuning parameters on a new dataset [50]. Additionally, the RF does not overfit when more independent and diverse trees are added. These make the usability of the RF more convenient. The RF was selected due to its simplicity; tuning a few parameters can result evaluates accuracy more than other ML models [51]. In this research, we evaluate Python’s scikit-learn package. The systemionality of the RF algorithm-values work is briefly explained as follows: the system selects a set of independent values to make an impact on each tree response, which is a subset of the predictor values of the initial dataset. The optimal subset-predictor value is calculated from $\log_2^{(M+1)}$, where M is the input. Now, we can calculate the mean-square error (RMSE) for an RF from

$$\varepsilon = (v_{observed} - v_{response})^2, \quad (2)$$

where ε , $v_{observed}$, and $v_{response}$ are mse, variables from observed, and result, respectively. Moreover, we can calculate the trees’ average prediction.

$$S = \frac{1}{t} \sum_t t^{th} v_{response}, \quad (3)$$

where S and t are RF prediction and the number of trees in the forest, respectively.

In classification, after defining a set of random trees and prediction, the algorithm compares the number of excess votes to other classes' average votes. Although a predictor set is randomly chosen for each tree from the equal distribution in the regression algorithm, each tree can add a numerical value response to form the RFs.

2.5. Long Short-Term Memory (LSTM)

For this research, we employed another Deep Learning method, named the long short-term memory (LSTM) network, a similar method to the recurrent neural network (RNN). The notion behind RNNs is to employ input data arbitrarily over extended sequences. It repeats the exact task to all elements in the series, and the results rely on the prior analysis. To be more specific, RNN includes a memory cell that grabs data until the training data sequence is completed. The RNNs are the better choice for the nonlinear time-series problems [52]. However, there are gradient issues to train long-time lags, which is required to predict time series or hydrology [52]. LSTM is developed to build a robust many-to-one model for hydrological-time series similar to RNN memory cells' structure of the input, self-recurrent connection, forget, and output gates [53]. Let's say the i_t , o_t , f_t are input, output, and forget gate at the time of t .

Figure 4 illustrates the (LSTM) adopted from [54], where x_t and h_t show the input and state at time t . Similarly, we have h and x at time $t - 1$ and $t + 1$, etc. C_t and h_t are defined as long-term and short-term (hidden) memory in this cell. According to the diagram, the chain of action happens in the network and lets the network learn long-term. The following equation will demonstrate the calculation of h_t and C_t at the t_{th} step in this process.

$$f_t = \sigma(U_f x_t + W_f h_{t-1} + b_f) \quad (4)$$

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i) \quad (5)$$

$$o_t = \sigma(U_o x_t + W_o h_{t-1} + b_o) \quad (6)$$

$$\dot{C}_t = \tanh(U_c x_t + W_c h_{t-1} + b_c) \quad (7)$$

$$C_t = f_t C_{t-1} + i_t \dot{C}_t \quad (8)$$

$$h_t = o_t \tanh(C_t) \quad (9)$$

where U_i and W_i are matrices for weight; b_i is the bias; σ is a sigmoid activation function; and \dot{C}_t is the candidate for the cell-state value.

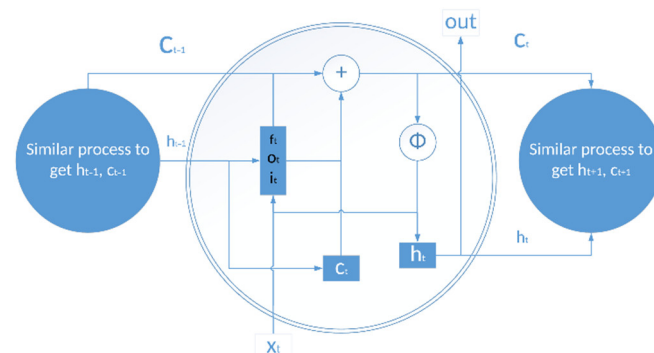


Figure 4. Memory block with the memory cell C_t .

In this work, the assembly of the time-delay model is used “Keras: The Python Deep Learning Library”. Similar to previous methods, we divided the dataset into training and testing subsets. We partitioned 70% of the data as the training set, 15% as the validation set, and 15% as the testing set. The LSTM-RNN has one layer for each input, output, and

LSTM with memory blocks. Based on two criteria, we assessed the model's accuracy: (i) the root mean square error (RMSE) and (ii) the ENS (Nash-Sutcliffe efficiency coefficient). Using these parameters is common in hydrological fields to assess the correlation between predicted and observed outcomes. The calculation formula is shown as follows:

$$\text{MSE} = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}}, \quad (10)$$

where O_i , P_i , and N are observation at time i , prediction at the time i , and several observations, respectively.

3. Results and Discussion

Forecasting time series accurately, particularly water levels for early flood warnings, is an essential but complicated process. As a classical statistical method, a classical ML algorithm, and a state-of-the-art Deep Learning method, respectively, the methods are seasonal autoregressive integrated moving average (SARIMA), Random Forest (RF), and Long Short-Term Memory (LSTM), which are widely used and effective forecasting models that have been proposed and tested on hydrological time series. Figures 2 and 3 present the monthly and annual data of these three selected stations. We evaluated and compared all tested ML methods by dividing the collected data into two parts for training and testing. The samples were taken with different frequencies from 1 January 2007 to 3 June 2017 for Pembina station, from 1 January 2007 to 7 February 2017 for Drayton station, and from 1 January 2007 to 5 August 2017 for Grand Forks station. As mentioned previously, the studied data involve 70% of the data as a training set, 15% as validation, and 15% as a testing set. All models were trained on the training datasets and then the trained models were used to forecast at a different time on the testing sets.

After applying the algorithms described above to three different sampling stations, the models were extracted for further evaluation and tabulated in Table 2. The table gives the details on the average forecast results of all tested methods at five different time intervals: six hours, twelve hours, one day, three days, and one week, for the Pembina, Drayton, and Grand Forks datasets. Low values of RMSE indicate a higher forecast accuracy of the chosen models. The best results for each forecasting horizon are highlighted in bold. By detecting the structures of the SARIMA, RF, and LSTM models, it was verified that the LSTM is more accurate than the two other models. The reason is that the LSTM model possesses a lower RMSE than the RF and SARIMA models for predicting the water-level data for the Red River of the North. Comparing the LSTM to the RF and SARIMA models in the Pembina station, the RMSE values are lower by 77.22% and 78.70%, respectively. Furthermore, there are 26.31% and 31.71% reductions in RMSE between the RF and SARIMA models at Drayton station, respectively, when using LSTM. Finally, the RMSE values for the Grand Forks station for LSTM are 83.70% lower than the RF model and 96.39% lower than the SARIMA model.

Figures 5–7 present the visual comparisons of all methods for forecasting one week of water level at Pembina, Drayton, and Grand Forks using a classical statistical method, SARIMA, a classical ML algorithm, RF, and a Deep Learning method, LSTM. The green line indicates the observed data that were used as the test data, and the red line indicates prediction data that is the output of our models.

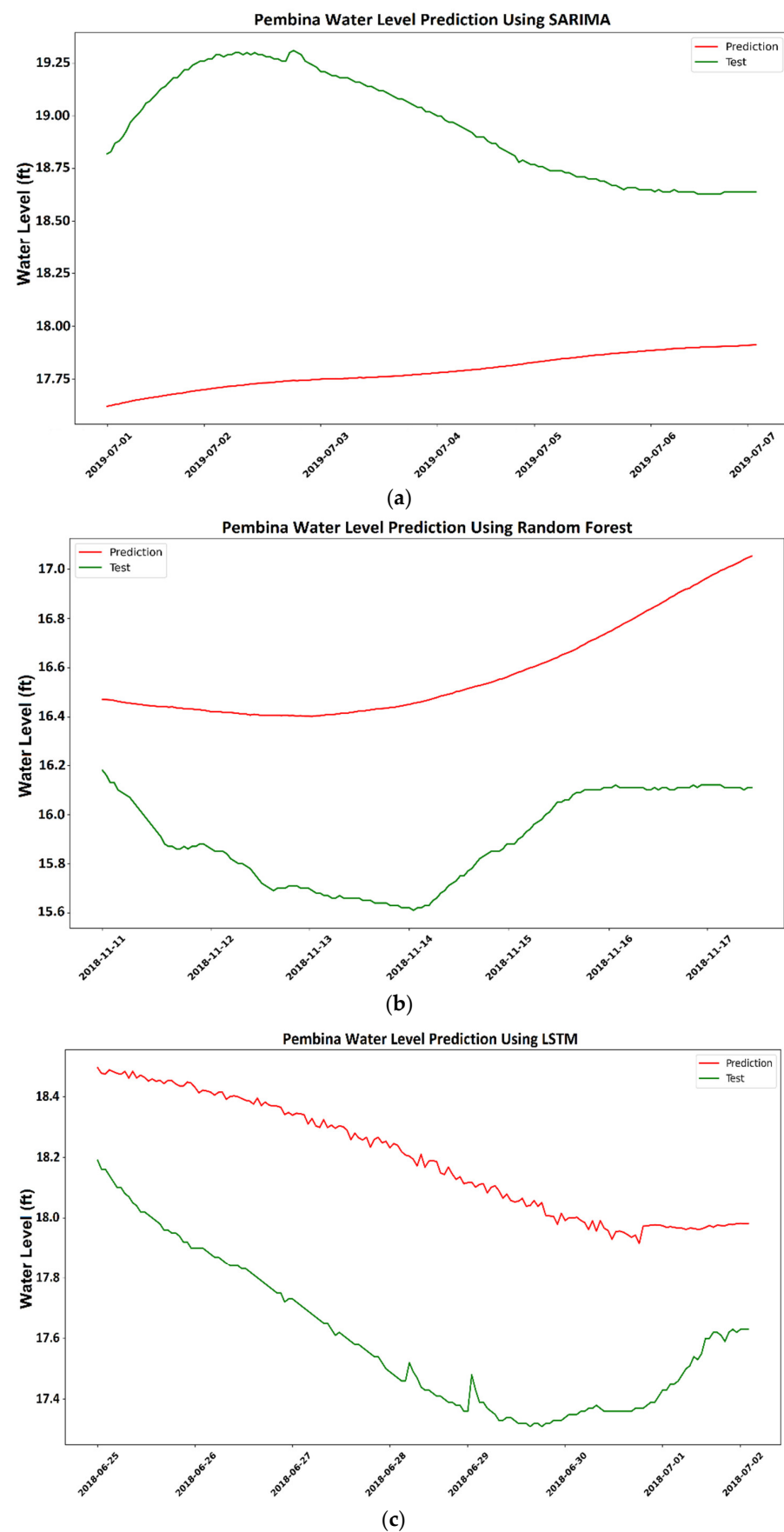


Figure 5. Visual comparison of one-week-ahead predicted values using (a) SARIMA, (b) RF, and (c) LSTM forecasting methods with true values on the Pembina series.

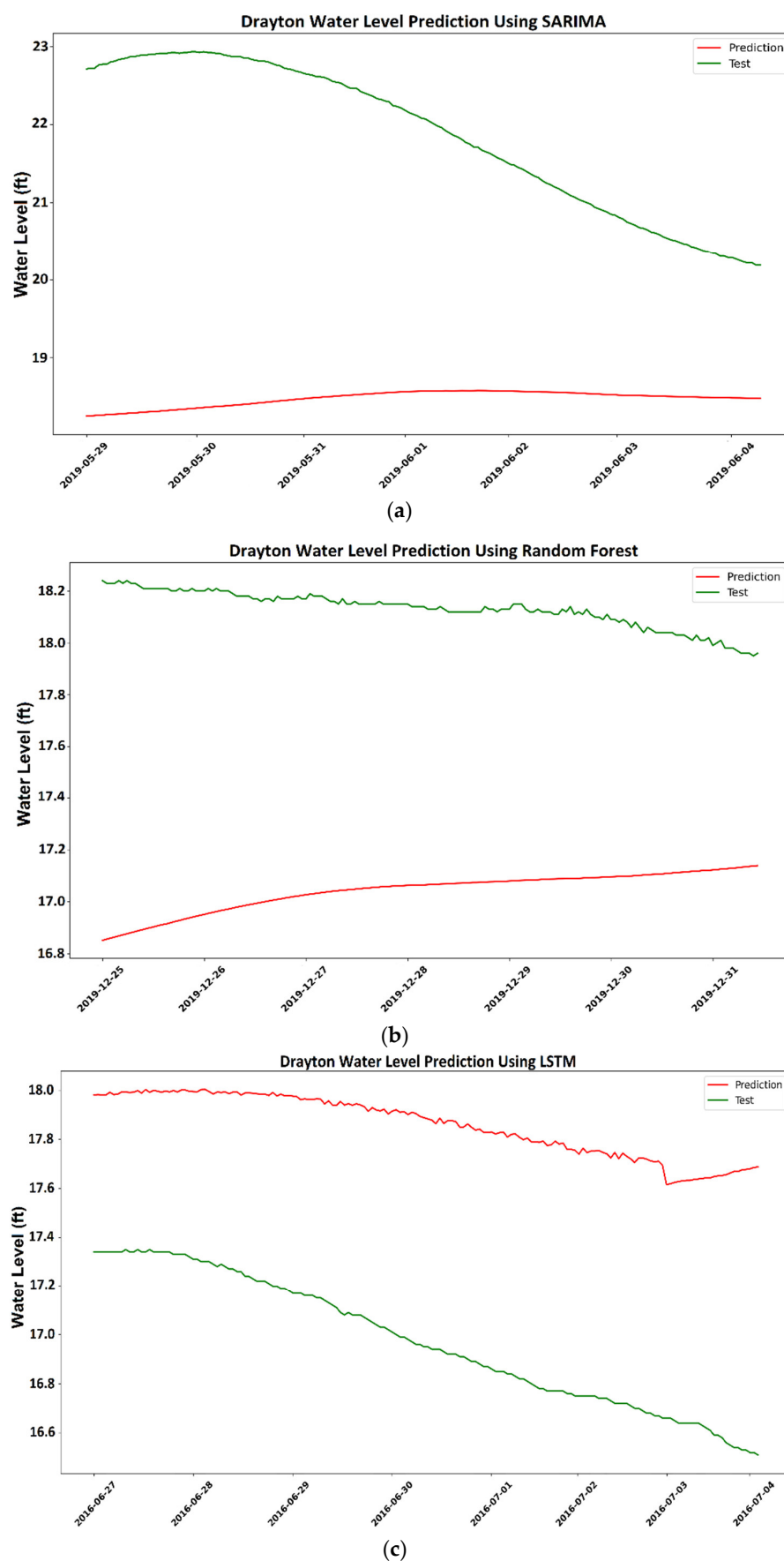


Figure 6. Visual comparison of one-week-ahead predicted values using (a) SARIMA, (b) RF, and (c) LSTM forecasting methods with true values on the Drayton series.

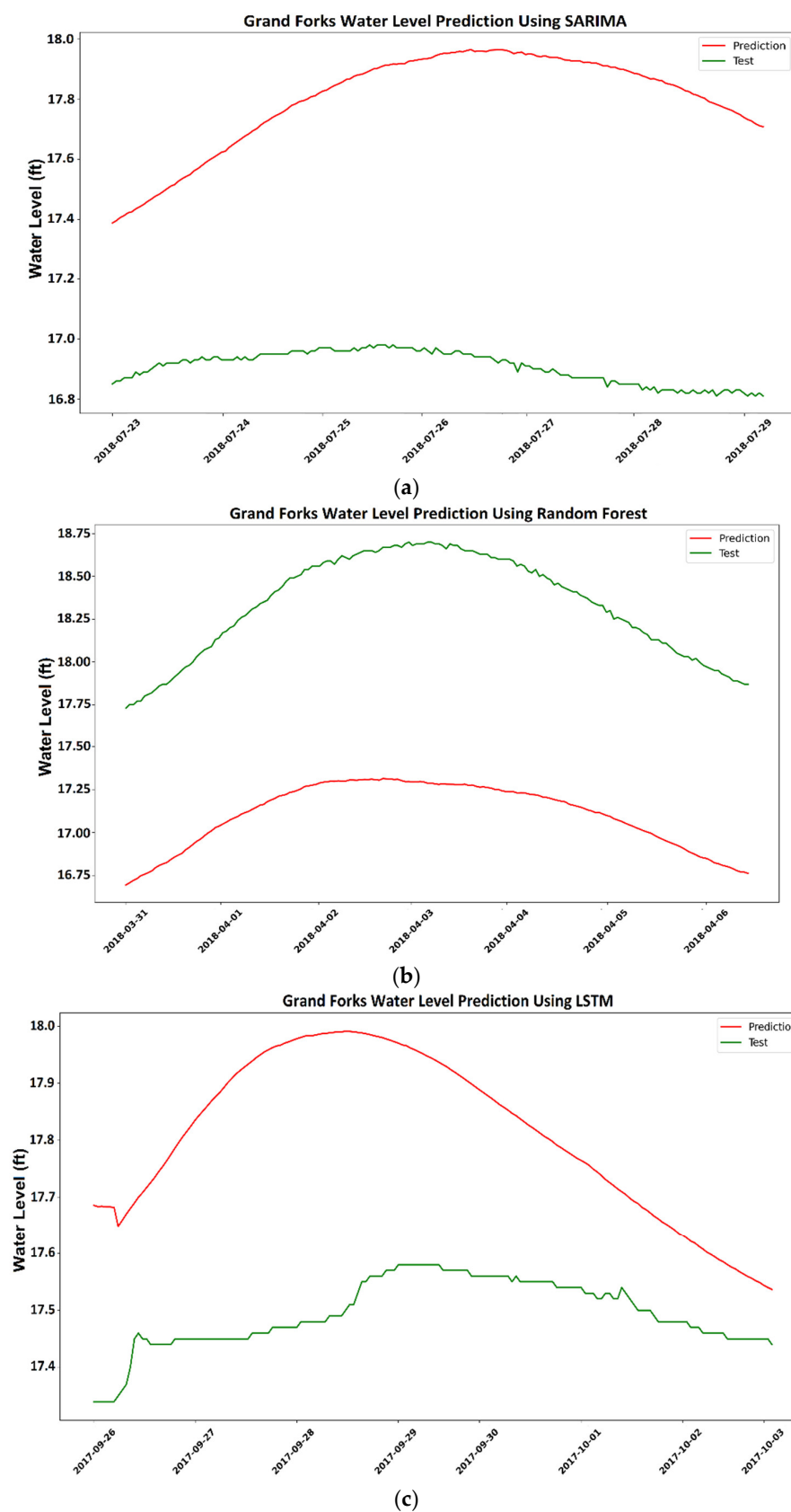


Figure 7. Visual comparison of one-week-ahead predicted values using (a) SARIMA, (b) RF, and (c) LSTM forecasting methods with true values on Grand Forks series.

Table 2. Evaluation of the performance of SARIMA, RF, and LSTM models at three USGS stations root mean square error (RMSE between the predicted and observed water-level data in the testing phase).

Pembina	6 h RMSE	12 h RMSE	1 Day RMSE	3 Days RMSE	1 Week RMSE
SARIMA	0.108	0.204	0.505	1.860	2.268
RF	0.101	0.160	0.269	0.865	2.287
LSTM	0.023	0.031	0.039	0.076	0.190
Drayton	6 h RMSE	12 h RMSE	1 day RMSE	3 days RMSE	1 week RMSE
SARIMA	0.041	0.074	0.152	0.535	1.491
RF	0.038	0.096	0.184	0.707	1.819
LSTM	0.028	0.035	0.041	0.065	0.151
Grand Forks	6 h RMSE	12 h RMSE	1 day RMSE	3 days RMSE	1 week RMSE
SARIMA	0.609	0.655	0.754	1.198	2.027
RF	0.135	0.246	1.059	1.632	2.673
LSTM	0.022	0.028	0.051	0.086	0.107

Figure 5 shows the results of forecasting the water level in a randomly chosen period at Pembina station one week ahead using SARIMA (from 1 July 2019 to 8 July 2019, Figure 5a), RF (from 10 November 2018 to 18 November 2018, Figure 5b), and LSTM (from 25 June 2018 to 2 July 2018, Figure 5c). When forecasting one week in advance, the LSTM yields the best results, as it could capture well the trend of the actual data. The results show that the LSTM performed better than the RF and SARIMA to predict the water level, with an average difference of 0.583 ± 0.21 feet between the tested and predicted water levels for three stations. The mean difference between the tested and predicted water levels for RF and SARIMA are 0.983 ± 0.64 feet and 1.848 ± 0.97 feet, respectively. The other two methods do not work as well as LSTM for the Pembina station. Figure 6 demonstrates the results of forecasting the water level in a randomly chosen period at Drayton station one-week ahead using SARIMA (from 28 May 2019 to 4 June 2019, Figure 6a), RF (from 25 December 2019 to 31 December 2019, Figure 6b), and LSTM (from 27 June 2016 to 4 July 2016, Figure 6c). Figure 7 shows a similar result to the case of Drayton station in that LSTM could quite accurately forecast the peak one week ahead. It still captures rather well the trend of the data in one-week ahead forecasts, but the errors are high. Meanwhile, all other methods failed to forecast and could not capture the data trend.

For Grand Forks data with hourly sampling, in a randomly chosen period for one-week-ahead prediction using SARIMA (from 23 July 2018 to 30 July 2018, Figure 7a), RF (from 31 March 2018 to 4 July 2018, Figure 7b), and LSTM (from 8 August 2019 to 15 August 2018, Figure 7c), Figure 7 demonstrates once again that the LSTM approach is superior to the SARIMA and RF methods. When predicting water levels one week ahead, LSTM produces the closest values to the real ones (Figure 7c). When forecasting water levels one week in advance, SARIMA and RF originate good results as in the case of one week, but LSTM produces predicted values, which are more similar to the true ones than other methods (Figure 7c). Although RF is second behind LSTM, the gaps between the forecast errors of the two methods are rather wide.

Figures 5, 6 and 7c demonstrate that for all water levels in all three stations, the LSTM method forecast was slightly overestimated. As can be seen in Figures 5, 6 and 7a, SARIMA underestimated the water level for Pembina and Drayton stations but overestimated the water level for Grand Forks station. Finally, the RF method overestimates the water level for Pembina station but underestimates the water level for Drayton and Grand Forks stations (Figures 5, 6 and 7b).

Although Figures 5–7 indicate the capacity of the model to estimate the water level in two weeks, the short duration of the sampling data may not be a suitable representation

of the models' capturing the flood peak. To present the accuracy of our model with different water-level datasets as driving inputs in capturing the flood peaks and times, we have considered one extreme three-month period that occurred in 2016, from May 16 to August 14. The major reason we offer this plot is that the reader cannot see how our model is excellent based on the statistics above. For this purpose, we have considered the maximum water-level events in the year 2016 and forecasted these events one week ahead. Figure 8 presents a comparison between observed and predicted data in Grand Forks station. The green line indicates the observed data that were used as test data, and the red line indicates the prediction data, which is the output of our model. The results indicate that the peak-flow scenarios in the field for May to August 2016 are well-captured by the trained LSTM.

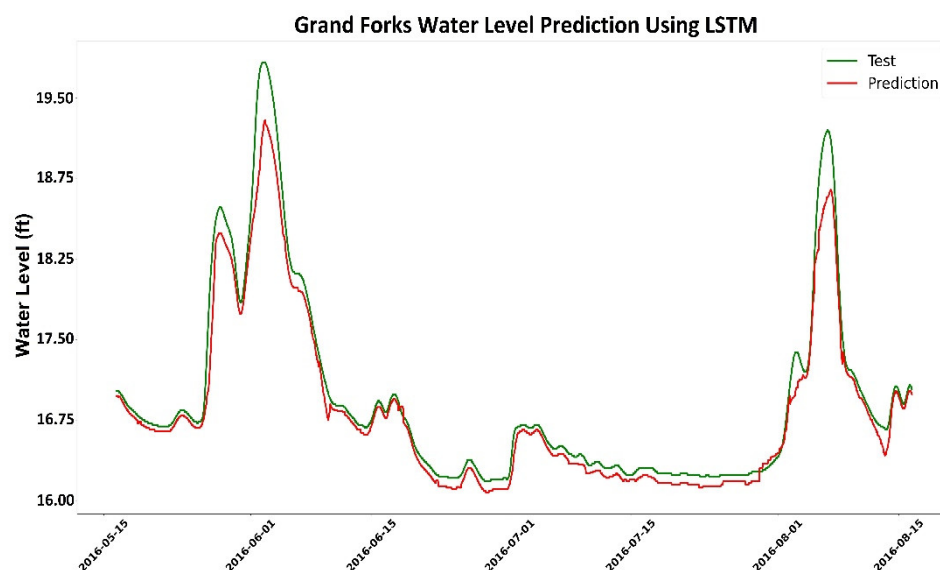


Figure 8. Visual comparison of 3 months of predicted values using LSTM forecasting method with true values on Grand Forks series.

4. Conclusions

Forecasting time series accurately, especially water levels for flood-warning systems, is an important but challenging task. The water-level forecasts at the Red River flow-gauging stations, specifically for downstream stations without any discharge information available, such as Pembina in this study, play a vital role in the early flood-warning system. In this paper, we have examined a classical statistical method, SARIMA; a classical ML algorithm, RF; and a Deep Learning method, LSTM. As shown in our comparison of the models for Pembina, Drayton, and Grand Forks stations, the LSTM method achieved better results and more accurate prediction performance than the SARIMA and RF methods. SARIMA is effective at modeling linear data, whereas the other statistical machine-learning models are superior at modeling nonlinear data. A water-stage time series, on the other hand, frequently has both linear and nonlinear correlation structures. Results show that for one-week-ahead prediction, the RMSE values for models fit to the series found at Pembina, Drayton, and Grand Forks are 0.190, 0.151, 0.107, respectively. These results demonstrate the high precision of the Deep Learning algorithm as a reliable choice for flood prediction. Experimental results on Pembina, Drayton, and Grand Forks stations show a better performance with the LSTM model in all prediction times. RMSE values for LSTM are lower by 77.22% in comparison with the RF model and lower by 78.70% in comparison with the SARIMA model. There are 26.31% and 31.71% reductions in RMSE between the RF and SARIMA models at Drayton station, respectively, when using LSTM. For the Grand Forks station, the RMSE values for LSTM are lower by 83.70% compared to the RF model and lower by 96.39% compared to the SARIMA model.

Author Contributions: Conceptualization, V.A., H.T.G. and S.M.S.; methodology, R.K.; software, H.T.G.; validation, H.T.G.; formal analysis, V.A.; investigation, V.A.; resources, S.M.S.; data curation, H.T.G.; writing—original draft preparation, V.A.; writing—review and editing, Y.H.L.; visualization, H.T.G.; supervision, V.A. and Y.H.L.; project administration, V.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data utilized in this study are available from the following source: <https://waterdata.usgs.gov/nwis/rt> (accessed on 10 November 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

- De Loë, R. *Sharing the Waters of the Red River Basin: A Review of Options for Transboundary Water Governance*; Prepared for International Red River Board, International Joint Commission; Rob de Loë Consulting Services: Guelph, ON, Canada, 2009; Available online: <http://www.ijc.org/files/publications/Sharing%20the%20Waters%20of%20the%20Red%20River%20Basin.pdf> (accessed on 22 January 2022).
- International Joint Commission. *Living with the Red: A Report to the Governments of Canada and the United States on Reducing Flood Impacts in the Red River Basin*; International Joint Commission: Washington, DC, USA, 2000.
- Li, L.; Simonovic, S.P. System dynamics model for predicting floods from snowmelt in North American prairie watersheds. *Hydrol. Process.* **2002**, *16*, 2645–2666. [\[CrossRef\]](#)
- Rannie, W. The 1997 flood event in the Red River basin: Causes, assessment and damages. *Can. Water Resour. J./Rev. Can. Des. Ressour. Hydr.* **2016**, *41*, 45–55. [\[CrossRef\]](#)
- Ryberg, K.R.; Macek-Rowland, K.M.; Banse, T.A.; Wiche, G.J. *A History of Flooding in the Red River Basin (No. 55)*; US Geological Survey: Reston, VA, USA, 2007.
- Tuttle, S.E.; Cho, E.; Restrepo, P.J.; Jia, X.; Vuyovich, C.M.; Cosh, M.H.; Jacobs, J.M. Remote Sensing of Drivers of Spring Snowmelt Flooding in the North Central U.S. In *Remote Sensing of Hydrological Extremes*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 21–45.
- Wang, S.; Russell, H.A.J. Forecasting Snowmelt-Induced Flooding Using GRACE Satellite Data: A Case Study for the Red River Watershed. *Can. J. Remote Sens.* **2016**, *42*, 203–213. [\[CrossRef\]](#)
- Rice, J.S.; Emanuel, R.E.; Vose, J.M.; Nelson, S.A. Continental US streamflow trends from 1940 to 2009 and their relationships with watershed spatial characteristics. *Water Resour. Res.* **2015**, *51*, 6262–6275. [\[CrossRef\]](#)
- Hirsch, R.M.; Ryberg, K.R. Has the magnitude of floods across the USA changed with global CO₂ levels? *Hydrol. Sci. J.* **2012**, *57*, 1–9. [\[CrossRef\]](#)
- Zhao, M.; Hendon, H.H. Representation and prediction of the Indian Ocean dipole in the POAMA seasonal forecast model. *Q. J. R. Meteorol. Soc.* **2009**, *135*, 337–352. [\[CrossRef\]](#)
- Borah, D.K. Hydrologic procedures of storm event watershed models: A comprehensive review and comparison. *Hydrol. Process.* **2011**, *25*, 3472–3489. [\[CrossRef\]](#)
- Costabile, P.; Costanzo, C.; Macchione, F. A storm event watershed model for surface runoff based on 2D fully dynamic wave equations. *Hydrol. Process.* **2013**, *27*, 554–569. [\[CrossRef\]](#)
- Cea, L.; Garrido, M.; Puertas, J. Experimental validation of two-dimensional depth-averaged models for forecasting rain-fall-runoff from precipitation data in urban areas. *J. Hydrol.* **2010**, *382*, 88–102. [\[CrossRef\]](#)
- Fernández-Pato, J.; Caviedes-Voullième, D.; García-Navarro, P. Rainfall/runoff simulation with 2D full shallow water equations: Sensitivity analysis and calibration of infiltration parameters. *J. Hydrol.* **2016**, *536*, 496–513. [\[CrossRef\]](#)
- Caviedes-Voullième, D.; García-Navarro, P.; Murillo, J. Influence of mesh structure on 2D full shallow water equations and SCS Curve Number simulation of rainfall/runoff events. *J. Hydrol.* **2012**, *448*, 39–59. [\[CrossRef\]](#)
- Costabile, P.; Costanzo, C.; Macchione, F. Comparative analysis of overland flow models using finite volume schemes. *J. Hydroinform.* **2012**, *14*, 122–135. [\[CrossRef\]](#)
- Xia, X.; Liang, Q.; Ming, X.; Hou, J. An efficient and stable hydrodynamic model with novel source term discretization schemes for overland flow and flood simulations. *Water Resour. Res.* **2017**, *53*, 3730–3759. [\[CrossRef\]](#)
- Liang, X.; Lettenmaier, D.P.; Wood, E.F.; Burges, S.J. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res. Earth Surf.* **1994**, *99*, 14415–14428. [\[CrossRef\]](#)
- Costabile, P.; Macchione, F. Enhancing river model set-up for 2-D dynamic flood modelling. *Environ. Model. Softw.* **2015**, *67*, 89–107. [\[CrossRef\]](#)
- Nayak, P.C.; Sudheer, K.P.; Rangan, D.M.; Ramasastri, K.S. Short-term flood forecasting with a neurofuzzy model. *Water Resour. Res.* **2005**, *41*, W04004. [\[CrossRef\]](#)

21. Kim, B.; Sanders, B.F.; Famiglietti, J.S.; Guinot, V. Urban flood modeling with porous shallow-water equations: A case study of model errors in the presence of anisotropic porosity. *J. Hydrol.* **2015**, *523*, 680–692. [\[CrossRef\]](#)
22. Feldman, A. *Hydrologic Modeling System HEC-HMS Technical Reference Manual: US Army Corps of Engineers*; Hydrologic Engineering Center: Davis, CA, USA, 2000.
23. Arnold, J.G.; Srinivasan, R.; Muttiah, R.S.; Williams, J.R. Large area hydrologic modeling and assessment part I: Model development. *JAWRA J. Am. Water Resour. Assoc.* **1998**, *34*, 73–89. [\[CrossRef\]](#)
24. Croke, B.F.; Andrews, F.; Jakeman, A.J.; Cuddy, S.; Luddy, A. Redesign of the IHACRES rainfall-runoff model. In Proceedings of the 29th Hydrology and Water Resources Symposium, Canberra, Australia, 20–23 February 2005.
25. Bicknell, B.R.; Imhoff, J.C.; Kittle, J.L., Jr.; Donigan, A.S., Jr.; Johanson, R.C. *Hydrological Simulation Program-FORTRAN. User's Manual for Release*; US EPA: Washington, DC, USA, 1996.
26. Fenicia, F.; Savenije, H.H.G.; Matgen, P.; Pfister, L. Understanding catchment behavior through stepwise model concept improvement. *Water Resour. Res.* **2008**, *44*, W01402. [\[CrossRef\]](#)
27. Li, X.H.; Zhang, Q.; Shao, M.; Li, Y.L. A Comparison of Parameter Estimation for Distributed Hydrological Modelling Using Automatic and Manual Methods. *Adv. Mater. Res.* **2012**, *356–360*, 2372–2375.
28. Bui, D.T.; Pradhan, B.; Nampak, H.; Bui, Q.-T.; Tran, Q.-A.; Nguyen, Q.-P. Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS. *J. Hydrol.* **2016**, *540*, 317–330. [\[CrossRef\]](#)
29. Mosavi, A.; Ozturk, P.; Chau, K.-W. Flood Prediction Using Machine Learning Models: Literature Review. *Water* **2018**, *10*, 1536. [\[CrossRef\]](#)
30. Yürekli, K.; Kurunç, A.; Öztürk, F. Testing the residuals of an ARIMA model on the Cekerek Stream Watershed in Turkey. *Turk. J. Eng. Environ. Sci.* **2005**, *29*, 61–74.
31. Ghimire, B.N. Application of ARIMA Model for River Discharges Analysis. *J. Nepal Phys. Soc.* **2017**, *4*, 27–32. [\[CrossRef\]](#)
32. Phan, T.-T.-H.; Nguyen, X.H. Combining statistical machine learning models with ARIMA for water level forecasting: The case of the Red river. *Adv. Water Resour.* **2020**, *142*, 103656. [\[CrossRef\]](#)
33. Kassem, A.A.; Raheem, A.M.; Khidir, K.M. Daily Streamflow Prediction for Khazir River Basin Using ARIMA and ANN Models. *ZANCO J. PURE Appl. Sci.* **2020**, *32*, 30–39. [\[CrossRef\]](#)
34. Singh, H.; Ray, M.R. Synthetic stream flow generation of River Gomti using ARIMA model. In *Advances in Civil Engineering and Infrastructural Development*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 255–263.
35. Elganiny, M.A.; Eldwer, A.E. Enhancing the Forecasting of Monthly Streamflow in the Main Key Stations of the River Nile Basin. *Water Resour.* **2018**, *45*, 660–671. [\[CrossRef\]](#)
36. Fernández, C.; Vega, J.A.; Fonturbel, T.; Jiménez, E. Streamflow drought time series forecasting: A case study in a small watershed in North West Spain. *Stoch. Hydrol. Hydraul.* **2009**, *23*, 1063–1070. [\[CrossRef\]](#)
37. Mgandu, F.A.; Mkandawile, M.; Rashid, M. Trend Analysis and Forecasting of Water Level in Mtera Dam Using Exponential Smoothing. *Int. J. Math. Sci. Comput.* **2020**, *4*, 26–34.
38. Lim, Y.H.; Voeller, D.L. Regional flood estimations in Red River using L-moment-based index-flood and bulletin 17B procedures. *J. Hydrol. Eng.* **2009**, *14*, 1002–1016. [\[CrossRef\]](#)
39. Board, R.R.B. *Inventory Team Report: Hydrology*; Red River Basin Board: Moorhead, MN, USA, 2000.
40. Pelletier, P.M. Uncertainties in streamflow measurement under winter ice conditions a case study: The Red River at Emerson, Manitoba, Canada. *Water Resour. Res.* **1989**, *25*, 1857–1867. [\[CrossRef\]](#)
41. Dakota Water Science Center. Red River of the North at Grand Forks, North Dakota—129 Years. Available online: https://www.usgs.gov/centers/dakota-water-science-center/red-river-north-grand-forks-north-dakota-129-years?qt-science_center_objects=0 (accessed on 22 January 2022).
42. Yu, Z.; Lei, G.; Jiang, Z.; Liu, F. ARIMA modelling and forecasting of water level in the middle reach of the Yangtze River. In Proceedings of the 2017 4th International Conference on Transportation Information and Safety (ICTIS), Edmonton, AB, Canada, 8–10 August 2017; pp. 172–177. [\[CrossRef\]](#)
43. Wang, W.-C.; Chau, K.-W.; Xu, D.-M.; Chen, X.-Y. Improving Forecasting Accuracy of Annual Runoff Time Series Using ARIMA Based on EEMD Decomposition. *Water Resour. Manag.* **2015**, *29*, 2655–2675. [\[CrossRef\]](#)
44. Azad, A.S.; Sokkalingam, R.; Daud, H.; Adhikary, S.K.; Khurshid, H.; Mazlan, S.N.A.; Rabbani, M.B.A. Water Level Prediction through Hybrid SARIMA and ANN Models Based on Time Series Analysis: Red Hills Reservoir Case Study. *Sustainability* **2022**, *14*, 1843. [\[CrossRef\]](#)
45. Yang, T.; Gao, X.; Sorooshian, S.; Li, X. Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme. *Water Resour. Res.* **2016**, *52*, 1626–1651. [\[CrossRef\]](#)
46. Wang, Z.; Lai, C.; Chen, X.; Yang, B.; Zhao, S.; Bai, X. Flood hazard risk assessment model based on random forest. *J. Hydrol.* **2015**, *527*, 1130–1141. [\[CrossRef\]](#)
47. Loos, M.; Elsenbeer, H. Topographic controls on overland flow generation in a forest—An ensemble tree approach. *J. Hydrol.* **2011**, *409*, 94–103. [\[CrossRef\]](#)
48. Biau, G.; Scornet, E. A random forest guided tour. *TEST* **2016**, *25*, 197–227. [\[CrossRef\]](#)
49. Tyrallis, H.; Papacharalampous, G.; Langousis, A. A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water* **2019**, *11*, 910. [\[CrossRef\]](#)

-
50. Lin, L.; Wang, F.; Xie, X.; Zhong, S. Random forests-based extreme learning machine ensemble for multi-regime time series prediction. *Expert Syst. Appl.* **2017**, *83*, 164–176. [[CrossRef](#)]
 51. Sahoo, B.B.; Jha, R.; Singh, A.; Kumar, D. Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophys.* **2019**, *67*, 1471–1481. [[CrossRef](#)]
 52. Akar, Ö.; Güngör, O. Classification of multispectral images using Random Forest algorithm. *J. Geodesy Geoinf.* **2012**, *1*, 105–112. [[CrossRef](#)]
 53. Abidogun, O.A. Data Mining, Fraud Detection and Mobile Telecommunications: Call Pattern Analysis with Unsupervised Neural Networks. Master's Thesis, University of the Western Cape, Cape Town, South Africa, 2005.
 54. Le, X.-H.; Ho, H.V.; Lee, G. River streamflow prediction using a deep neural network: A case study on the Red River, Vietnam. *Korean J. Agric. Sci.* **2019**, *46*, 843–856.