




Article

Application of Machine Learning and Process-Based Models for Rainfall-Runoff Simulation in DuPage River Basin, Illinois

Amrit Bhusal , Utsav Parajuli, Sushmita Regmi  and Ajay Kalra * 

School of Civil, Environmental, and Infrastructure Engineering, Southern Illinois University, 1230 Lincoln Drive, Carbondale, IL 62901-6603, USA; amrit.bhusal@siu.edu (A.B.); utsav.parajuli@siu.edu (U.P.); sushmita.regmi@siu.edu (S.R.)

* Correspondence: kalraa@siu.edu; Tel.: +1-(618)-453-7008

Abstract: Rainfall-runoff simulation is vital for planning and controlling flood control events. Hydrology modeling using Hydrological Engineering Center—Hydrologic Modeling System (HEC-HMS) is accepted globally for event-based or continuous simulation of the rainfall-runoff operation. Similarly, machine learning is a fast-growing discipline that offers numerous alternatives suitable for hydrology research's high demands and limitations. Conventional and process-based models such as HEC-HMS are typically created at specific spatiotemporal scales and do not easily fit the diversified and complex input parameters. Therefore, in this research, the effectiveness of Random Forest, a machine learning model, was compared with HEC-HMS for the rainfall-runoff process. Furthermore, we also performed a hydraulic simulation in Hydrological Engineering Center—Geospatial River Analysis System (HEC-RAS) using the input discharge obtained from the Random Forest model. The reliability of the Random Forest model and the HEC-HMS model was evaluated using different statistical indexes. The coefficient of determination (R^2), standard deviation ratio (RSR), and normalized root mean square error (NRMSE) were 0.94, 0.23, and 0.17 for the training data and 0.72, 0.56, and 0.26 for the testing data, respectively, for the Random Forest model. Similarly, the R^2 , RSR, and NRMSE were 0.99, 0.16, and 0.06 for the calibration period and 0.96, 0.35, and 0.10 for the validation period, respectively, for the HEC-HMS model. The Random Forest model slightly underestimated peak discharge values, whereas the HEC-HMS model slightly overestimated the peak discharge value. Statistical index values illustrated the good performance of the Random Forest and HEC-HMS models, which revealed the suitability of both models for hydrology analysis. In addition, the flood depth generated by HEC-RAS using the Random Forest predicted discharge underestimated the flood depth during the peak flooding event. This result proves that HEC-HMS could compensate Random Forest for the peak discharge and flood depth during extreme events. In conclusion, the integrated machine learning and physical-based model can provide more confidence in rainfall-runoff and flood depth prediction.

Keywords: rainfall-runoff; HEC-HMS; HEC-RAS; random forest; flood; forecast



Citation: Bhusal, A.; Parajuli, U.; Regmi, S.; Kalra, A. Application of Machine Learning and Process-Based Models for Rainfall-Runoff Simulation in DuPage River Basin, Illinois. *Hydrology* **2022**, *9*, 117. <https://doi.org/10.3390/hydrology9070117>

Academic Editors: Davide Luciano De Luca and Andrea Petroselli

Received: 30 May 2022

Accepted: 24 June 2022

Published: 27 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

Floods are some of the most common and costly natural catastrophes in the world [1–3]. The magnitude and frequency of extreme flooding events have increased considerably worldwide over the previous few decades [4]. Climate change, urbanization, and other anthropogenic activities are causing a flood risk globally [5–7]. Water-related natural hazards, such as floods, droughts, and landslides, have become the new normal due to the uncertainty in rainfall patterns and magnitudes caused by climate change and urbanization [8]. Flooding is projected to become more common in the coming years as the frequency of extreme precipitation events increases [9–11].

Flood severity has increased, resulting in a large number of flood fatalities, massive economic losses, and social consequences [12]. Given the negative consequences of flooding,

developing floodplain management plans to avoid and mitigate flood damage is critical [13]. The estimation of Intensity–Duration–Frequency (IDF) curves and the monitoring of rainfall intensity are also critical factors in precisely calculating the flood hydrograph and the peak discharges [14,15]. The flood risk assessment depends on a precise estimation of peak runoff, calculated by rainfall-runoff simulation [16]. Accurate rainfall-runoff simulation is a prominent topic in hydrology research [17]. Precise rainfall-runoff modeling is essential for planning and applying flood control strategies in vulnerable areas to reduce the dangers to human life and infrastructure during high-precipitation events. Different hydrology models have been used in the past to perform a rainfall-runoff simulation in a watershed. The Hydrologic Modeling System (HMS), designed by the Hydrologic Engineering Center (HEC) of the United States Army Corps of Engineers, is a popular rainfall-runoff analysis tool worldwide [18].

Process-based physical models are typically employed to calculate runoff in a particular catchment area. By integrating regional variability in the watershed, a physical-based model such as HEC-HMS can compute an actual hydrology system [19]. Hydrology modeling using the HEC-HMS model can be used to investigate urban floods, flood frequency, flood warning systems, and the effectiveness of spillways and detention ponds over a watershed [20]. The HEC-HMS model is made up of four essential components. An analytical method is first applied to compute direct discharge and reach routing. Secondly, a basin model with interactive components is employed for depicting hydrology aspects within a catchment. Third, data are entered, edited, managed, and stored via a system. Fourth, the simulation results are reported and illustrated using a functional system [21]. Finally, the calibration procedure, which compares simulated results to observed data, can help to enhance the model's precision and predictability. With the regional and temporal variety of catchment features, rainfall patterns, and the number of variables applied in modeling physical processes, the connection between precipitation and discharge using HEC-HMS is challenging [22]. A physical-based model such as HEC-HMS necessitates a large amount of data, such as land use and land cover data, soil group data, and infiltration data, and a significant amount of time to calibrate to ensure the correctness of the model [23]. Furthermore, there are drawbacks to using a physical-based hydrology model, owing to the difficulties in completely understanding the complicated, nonlinear, and inter-related hydrology [24,25]. A hydrology model that uses HEC-HMS may be unsuitable for a larger watershed with scarce data. Therefore, as a complement to the physical model, recently, the application of machine learning and data-driven models has been used across hydrology domains [26,27].

Machine learning (ML) is a kind of artificial intelligence that can make an accurate prediction by training and testing datasets. ML provides a solution to a real-world problem by studying previously observed data and has been effective in generating accurate results [28]. ML provides adequate computation power [29,30] and is used in a wide variety of research and applications in hydrology. Some examples of ML applications in the hydrology domain are rainfall-runoff prediction [31–33], flood forecasting [34–36], sedimentation studies [37–39], water quality prediction [40–43], groundwater prediction [44,45], river temperature prediction [46–49], and rainfall estimation [50,51]. In recent years, ML algorithms have significantly improved and are also widely used for rainfall-runoff simulation [52,53] thanks to the rapid advancement of computer technology. Recently, many researchers have performed rainfall-runoff predictions using different machine learning and data-driven models. Some examples of these models are long short-term memory [54,55], artificial neural networks [56,57], support vector machines [58,59], and the Random Forest model [16,60]. Random Forest is a popular machine learning tool, and Breiman first developed it in 2001 [61]. Random Forest has recently acquired popularity as a powerful predictive modeling tool, and many researchers are using it in their fields as a potential method [62]. It is a classification and regression tree-based ensemble learning algorithm [61]. A bootstrap sample is used to train each tree, and optimal variables at each split are chosen

from a random subset of all variables. Random Forest offers the highest accuracy of any contemporary method and works quickly on large datasets [63].

Previous studies showed that Random Forest's performance surpassed other machine learning and data-driven tools such as artificial neural networks, regression models, and support vector machines in multiple comparative studies in hydrology [63–67]. However, Random Forest is the least used for hydrology analysis among the data-driven and machine learning models [68]. Among the few applications of Random Forest, most of these studies focused on flood risk hazards [16,69] and mapping [70]. Therefore, this study evaluated the effectiveness of the Random Forest model for rainfall-runoff simulation. In addition, the main objective of this research is to determine the suitability of the Random Forest model for rainfall-runoff simulation in a scarce-data region. Therefore, this research also used a satellite precipitation product as an input variable for rainfall-runoff simulation and determined its appropriateness in hydrology research. Furthermore, this study assessed the appropriateness of using Random Forest generated discharge for hydraulic modeling using the Hydrologic Analysis Center's River Analysis Model (HEC-RAS).

HEC-RAS is the most widely accepted model [71] for analyzing channel flow and floodplain characterization [72]. Users can compute one-dimensional steady and unsteady flow, two-dimensional unsteady flow, sediment transport, and water quality models by using HEC-RAS [72]. Regularizing geometric data and identifying and analyzing hydraulic structures, such as weirs, culverts, reservoirs, pump stations, bridges, levees, and gates, blockage and ineffective regions, land use, the Manning roughness coefficient, streambed slopes, and ice cover are achievable with HEC-RAS [73]. The model employs geometric data and geometric and hydraulic computer algorithms to model natural and artificial streams. HEC-RAS requires fundamental inputs such as river discharge, channel geometry, bank lines, flow paths, and channel resistance. The discharge generated by Random Forest was employed as an input parameter in this study. While the HEC-RAS model has a wide variety of capabilities, the current research considered its capability to execute 1D river flow and calculate the flood depth at the most downstream section of the study reach.

The integration of different models in the sectors of hydrology and hydraulic domains is gaining global attention and is crucial for flood risk management techniques [74]. The novelty of this research is to assess the effectiveness of the Random Forest model for rainfall-runoff simulation using satellite precipitation products in a data-scarce region. This research work also evaluated the integration of machine learning and a HEC-RAS model for calculating water depth at the proposed study location during the study period. The following is an outline of this paper. Section 2 describes the study area, data preparation, and a physical-based and Random Forest model. Section 3 presents the results of this research, Section 4 provides a discussion of the results, and Section 5 provides the major conclusions from the current analysis.

2. Data and Methods

This section describes the methodology used for hydrology and hydraulic analysis in this research. Random Forest, HEC-HMS, and HEC-RAS are the three models used in this study. HEC-HMS and the Random Forest model were applied for hydrology analysis, and HEC-RAS was used for the hydraulic analysis. The complete workflow of the methodology used in this research work is shown in Figure 1. First, this study started with extracting and preprocessing the data on basin characteristics, such as digital elevation model (DEM), land use and land cover (LULC), and soil group data, and meteorological data, such as daily precipitation and discharge data. The integrated use of Arc-Hydro, HEC-GeoHMS, and HEC-HMS was employed for hydrology analysis in the upstream catchment area. Similarly, Random Forest, a machine learning algorithm, was used to predict the runoff for the training and testing period. After the preparation of the hydrology model, a comparison was performed between the machine learning model (Random Forest Regression) and the physical model (HEC-HMS) using the different statistical indexes. Finally, the runoff obtained from the machine learning model was used as an input variable in the HEC-RAS

model to calculate the water depth at the downstream location. In conclusion, the modeling approach determined the effectiveness of Random Forest Regression for hydrology and the integrated Random Forest and HEC-RAS model for hydraulic analysis.

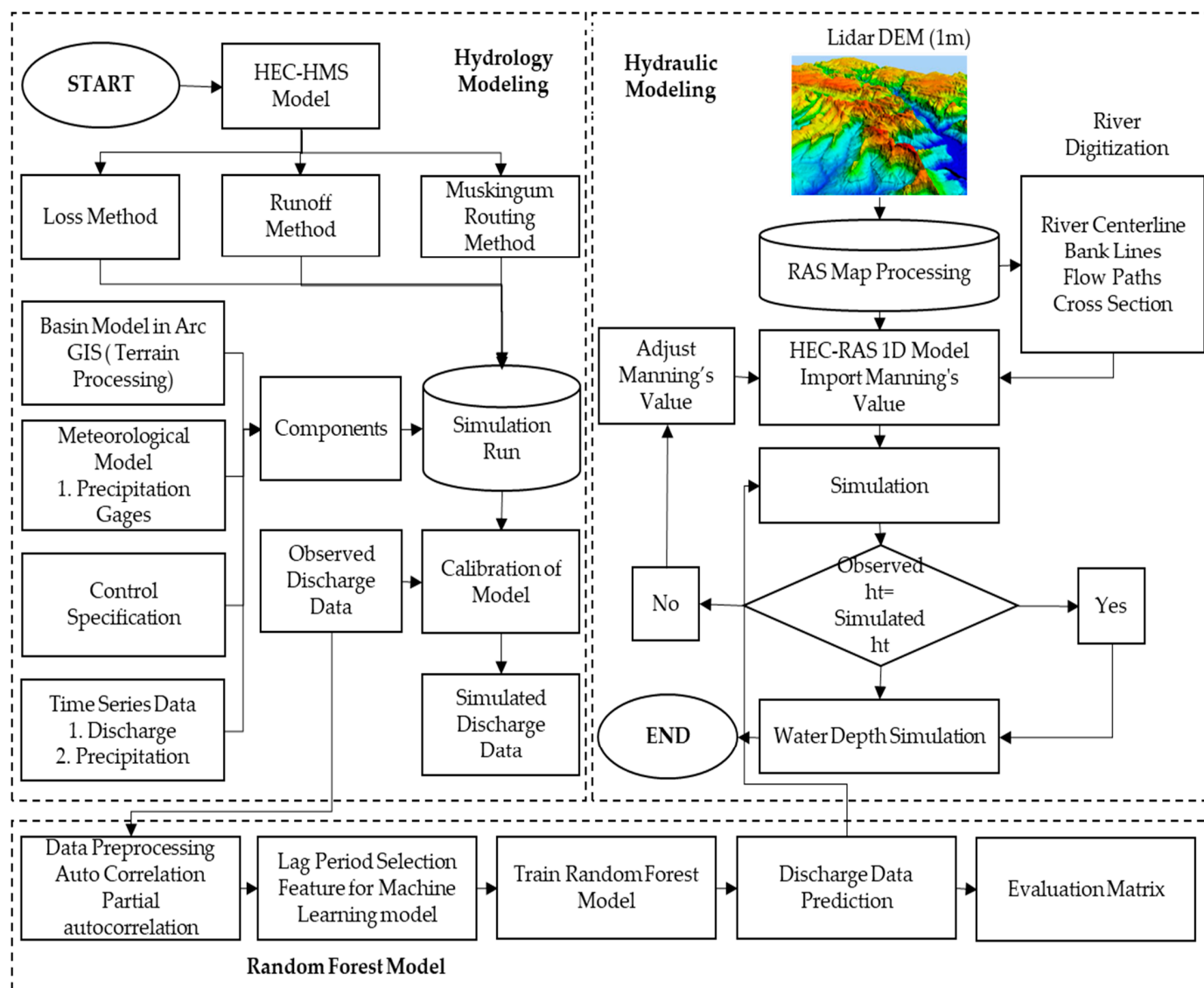


Figure 1. Figure portraying the flowchart of hydrology analysis using Random Forest and HEC-HMS and hydraulic analysis using HEC-RAS.

2.1. Study Area

This research used the East Branch DuPage watershed as a study area. Over the last twenty years, the study area has observed significant urbanization [75]. The study area has a history of high-flooding events (1996, 2008, 2013, and, most recently, 2020). In the year 2020, there was significant flooding due to the 178 mm of total precipitation over a period of five days. The study watershed has an area of 62.2 km² at the USGS gauging station, which is around Downers Grove, Illinois. The study area has an elevation ranging from 204 m to 250 m above mean sea level. Geographically, northern latitudes from 41°50' to 41°57' and western longitudes from 87°59' to 88°6' bound the study catchment area, as shown in Figure 2. The study area is highly residential, with an average imperviousness percentage of about 40%. The range of imperviousness percentages in the watershed is shown in Figure 3. The average soil permeability over the watershed is 62 mm/h [76]. The catchment consists of USGS gauge station 05540160 at the watershed outlet. The river reach

for the hydraulic station lies between the gauging stations 05540160 and 05540228. The study reach is around 5221 m between two gauging stations. The proposed study area does not have any existing precipitation gauging station. The history of flooding events and the unavailability of observed precipitation data in this watershed are the two main reasons for proposing this watershed as a study area.

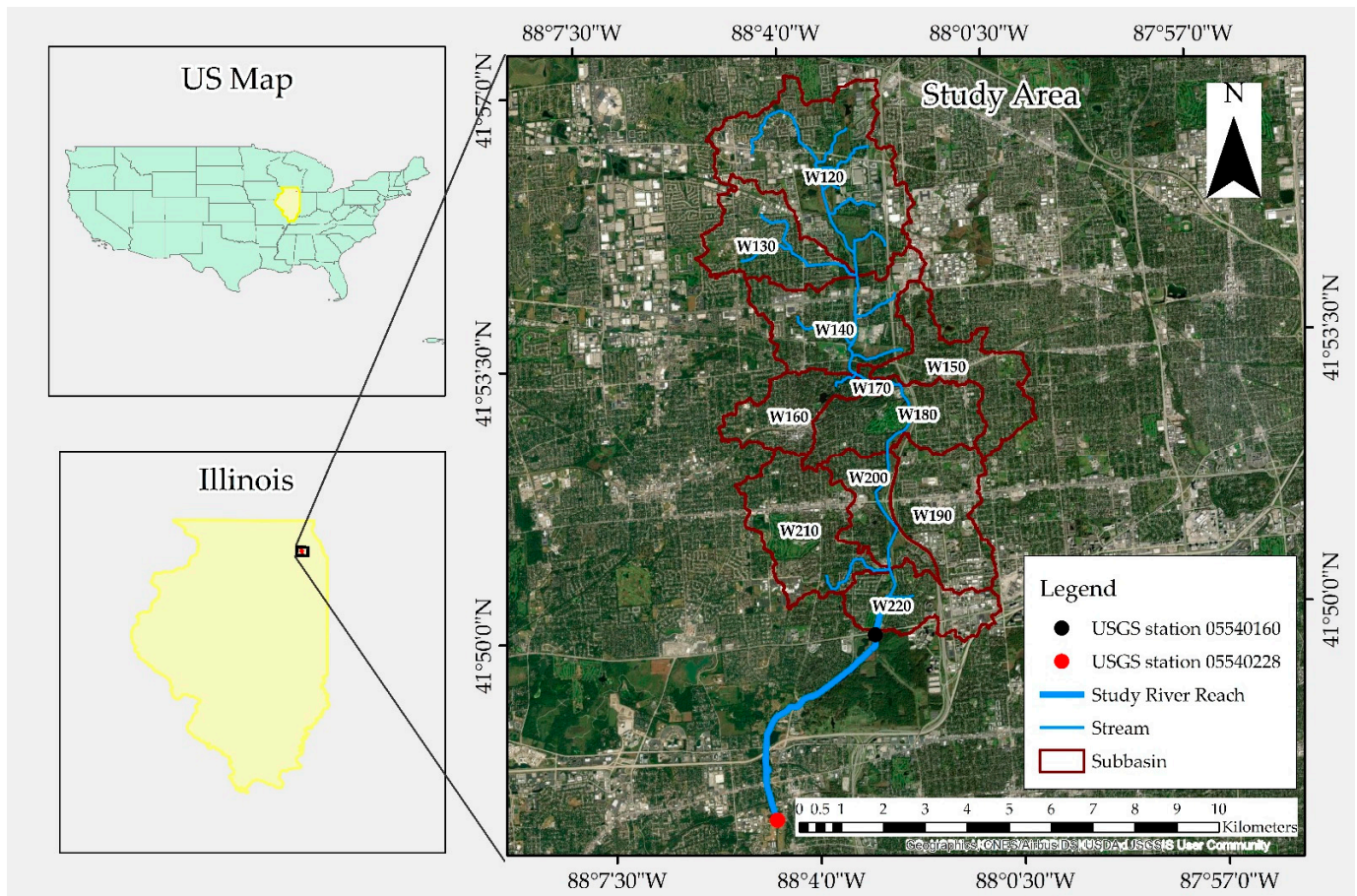


Figure 2. The East Branch DuPage Catchment around Downers Grove, Illinois, with the river system.

2.2. Data

Watershed characteristics datasets, such as land use and land cover, soil group, and DEM datasets, and meteorological model data, such as rainfall and discharge data, are all important data required for hydrology and hydraulic simulation. These datasets were used to estimate hydrology parameters and sub-basin characteristics and to prepare geometric data for hydrology and hydraulic analysis. The data types used in this research and their sources are detailed in Table 1.

2.3. Preprocessing Data

This section describes the extraction of basin characteristics and the meteorological data that were used for the hydrology analysis.

2.3.1. Digital Elevation Model (DEM)

DEM data are spatial data that provide the characteristics of the watershed. A 10 m DEM was retrieved from a United States Department of Agriculture (USDA) website and was clipped for the study catchment using Arc-Map in Arc-GIS.

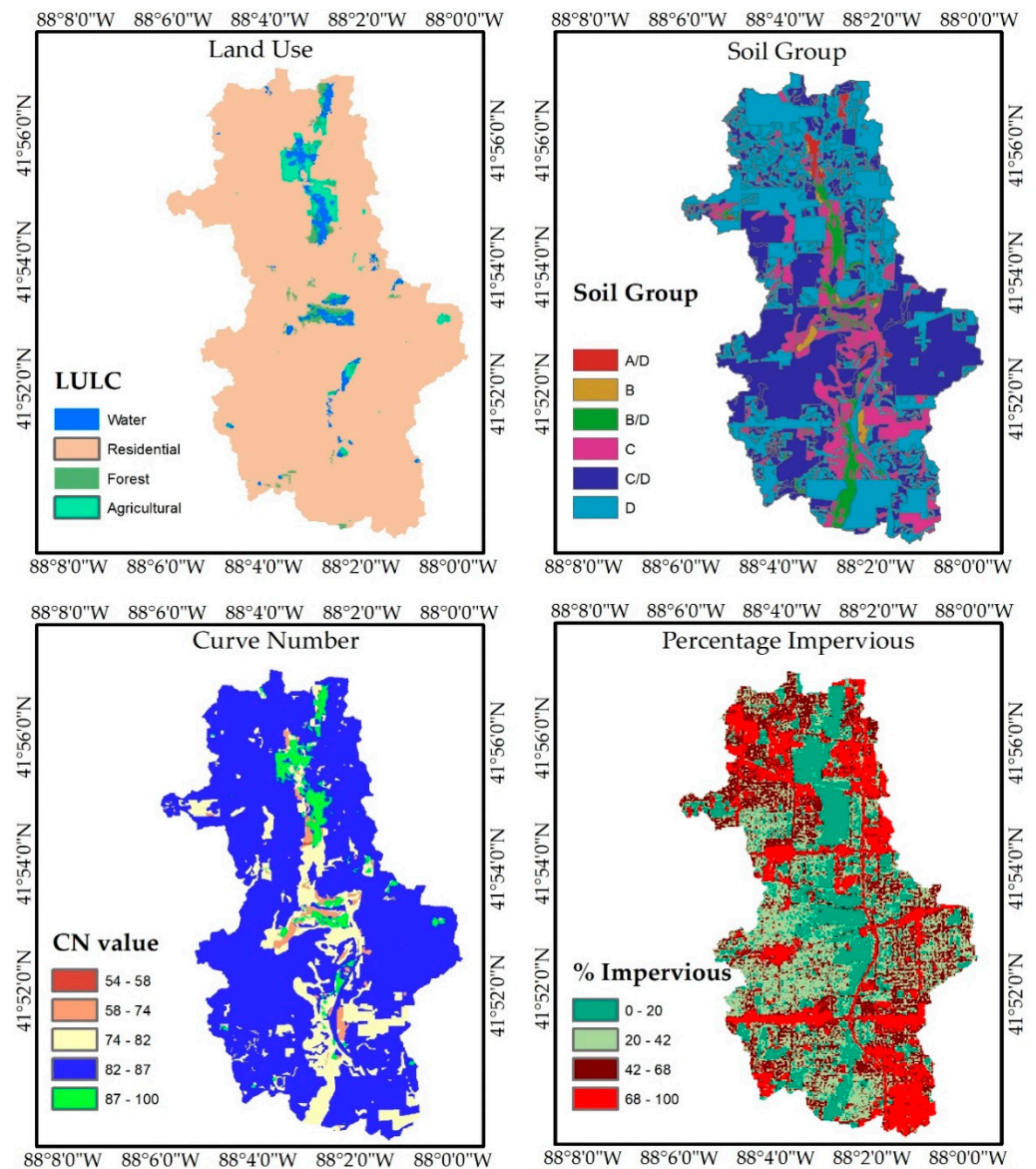


Figure 3. Map depicting characteristics of the study area.

Table 1. Data used for this research with their sources.

Data	Source
Precipitation	Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Cloud Classification System (PERSIANN-CCS).
Soil	United States Department of Agriculture (USDA)
Land Use Land Cover	United States Geological Survey (USGS)
Runoff Data	United States Geological Survey (USGS) water data

2.3.2. Basin Characteristics

LULC data and soil map data were extracted from a USGS and USDA website, respectively. Both datasets were imported into ArcMap to clip for a study boundary and converted to the Shapefile from the raster. Composite curve number values were generated considering pervious and impervious areas. The average curve number of the watershed was 83.4, and the curve number values ranged from 54 to 100, corresponding to high infiltration to water bodies, respectively. The basin characteristics of the study area are shown in Figure 3.

2.3.3. Precipitation Data

Rainfall data are essential meteorological data for hydrology simulations. The study area does not consist of any observed precipitation station; therefore, in this study, precipitation data were obtained from a grid from the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Cloud Classification System (PERSIANN-CCS). The Center for Hydrometeorology and Remote Sensing (CHRS) develops it at the University of California, Irvine, and it is a real-time global high-resolution ($0.04^\circ \times 0.04^\circ$ pixel) satellite precipitation product [77]. The daily time series precipitation data were extracted from a grid using a python environment from 2006 to 2021.

2.4. Hydrologic Modeling Using Arc-GIS and HEC-HMS

HEC-GeoHMS is an extension of Arc-GIS that helps users to extract the essential data to develop the HEC-HMS project. The user must pick an outlet position on the river to begin the extraction procedure. HEC-GeoHMS utilizes terrain preprocessing tools for flow analysis. HEC-GeoHMS can enhance the sub-basin and stream delineations, collect physical attributes of sub-basins and rivers, predict model attributes, and create input files for HEC-HMS. Terrain preprocessing and model development were carried out as shown in Figure 4.

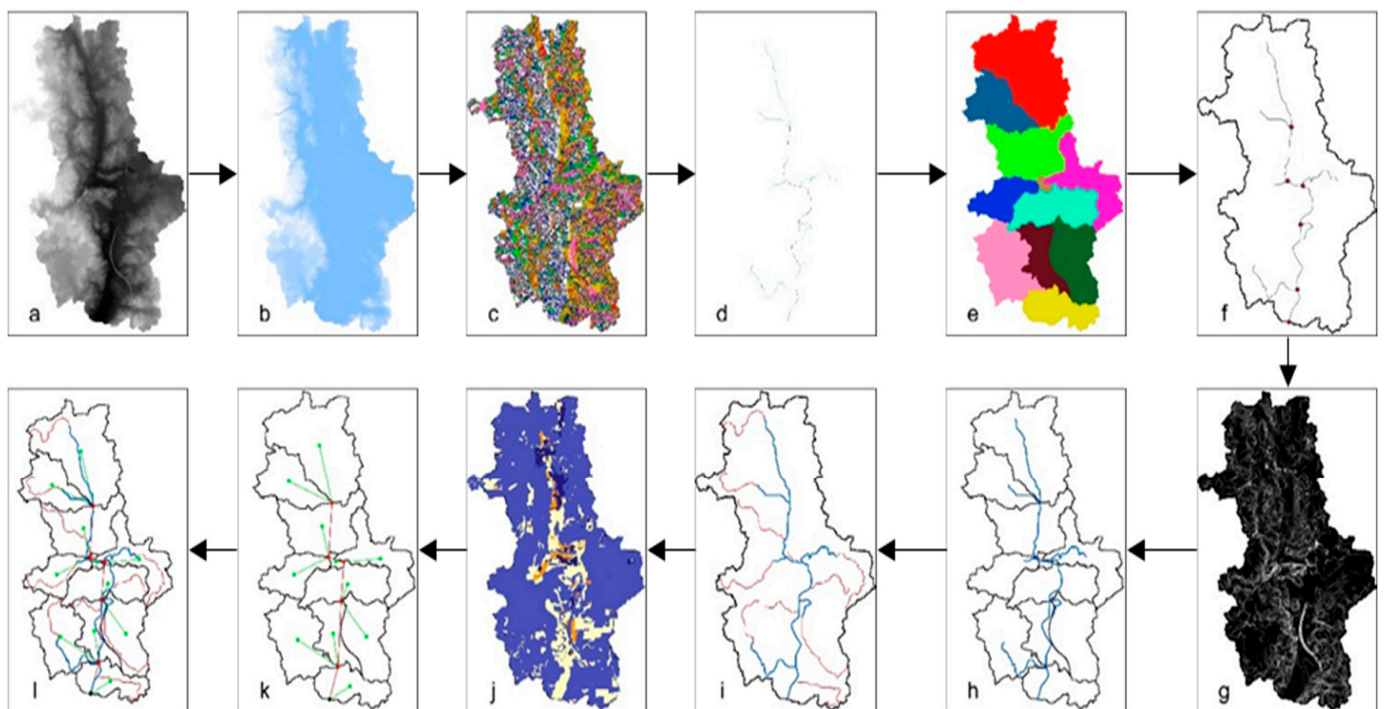


Figure 4. Preprocessing and model development: (a) DEM file; (b) Fill Sinks; (c) Flow Accumulation; (d) Flow Direction; (e) Stream Definition and Catchment Polygon; (f) Drainage Point and Line Processing; (g) Slope; (h) Basin and River Merge; (i) Lonest Flow Path; (j) CN Lag; (k) Sub-basin Nodes and River Links; (l) HEC-HMS input file.

2.4.1. Loss Method: SCS-CN for Rainfall-Runoff

The Soil Conservation Service curve number (SCS-CN) is a loss model that can compute the volume of the river flows [78]. Surface runoff excess depends on the precipitation, soil, and LULC of a particular watershed. Equation (1) is a mathematical expression used to determine the surface runoff.

$$Q = \frac{(P - I_a)^2}{(P - I_a) + S} \quad (1)$$

where

Q = Runoff (inches);

P = Rainfall depth (inches);

I_a = Initial abstraction, and $I_a = 0.2 S$;

S = Potential maximum retention.

The potential maximum retention in inches, S , is calculated using Equation (2):

$$S = \frac{1000}{CN} - 10 \quad (2)$$

2.4.2. Transform Method: SCS Unit Hydrograph

The SCS Unit Hydrograph transforms excess precipitation into a runoff. The SCS proposed the Unit Hydrograph, which is used in the HEC-HMS model. It is a parametric model based on the average Unit Hydrograph, which is created from gauged precipitation and discharge data of various agricultural watersheds collected across the United States. It assumes that a Unit Hydrograph depicts the constant properties of a watershed. The lag time is the sole input variable for this method. It is the time distance between the center of excess rainfall and the hydrograph peak, and HEC-HMS computes it for each sub-basin using Equation (3).

$$T_{lag} = \frac{(S + 1)^{0.7} L^{0.8}}{1900 * Y^{0.5}} \quad (3)$$

where

T_{lag} = lag time (h);

L = hydraulic length of the watershed (ft);

Y = slope of the watershed (%);

S = maximum retention in the watershed (inches).

2.4.3. Routing Method: Muskingum Routing

Discharges from sub-basins were routed through the reaches to the outlet of the watershed using the Muskingum routing method. X and K are the two main parameters used in this method. Theoretically, the K parameter is the wave's travel time through the reach. These parameters can be approximated using observed inflow and outflow hydrographs. The X parameter is a weight coefficient of discharge, whose value fluctuates between 0 and 0.5. The interval between the inflow and outflow hydrographs of an identical station can be used to determine the parameter K . In this model, routing methods parameters were used to calibrate the model.

2.5. Hydrologic Modeling Using Random Forest

This study investigated the capacity of a Random Forest algorithm for predicting the daily discharge using the meteorological and hydrology features. Nonlinear interactions between a dependent variable and several independent variables can be represented using regression tree ensembles such as the Random Forest technique. Despite the popularity of the Random Forest algorithm in a myriad of environmental science fields, its application in the water sector needs to be further explored [79]. Random Forest is the type of supervised machine learning algorithm that can be used for classification and prediction. Random Forest uses the different tree predictors, and the random vector determines their values [61]. Random Forest is a collection of decision trees, where each tree is slightly different from the others. Ensemble learning combines all the decision trees and the average values predicted by each decision tree, solving the regression problem. This algorithm addresses the problem of training data overfitting in decision trees [80]. Random Forest has good performance in large datasets, and its features do not need to be scaled [81]. It is advantageous for features with different scales. Random Forests are appealing for both classification and regression tasks, are computationally fast, are efficient for unstable prediction, and perform well with high-dimensional features [82,83]. This algorithm's key idea is that each tree might make

a fair prediction on its part; however, overfitting seems to occur on some of the data. If numerous trees are built, they will work and overfit in various ways. The average of these results will assist in the reduction of overfitting while holding onto the predictive power of decision trees.

Model Development

Many decision trees with bootstrap aggregation are used to minimize the overfitting issue [84]. A Random Forest Regressor consisting of 100 decision trees, as n -estimators, were applied to this dataset. The max depth parameter defines the maximum depth of the tree. The max depth of the model was fixed to be 100. The max depth by default was 'None', which signifies that the nodes were enlarged until all the leaves had fewer than `min_samples_split` samples. `Min_samples_split` means the total number of samples needed to break the internal node. Since we were trying to maintain the number of decision trees at only 100, max features was set to 'auto', which means that max features was equal to n features (the number of features seen during the model fitting). The parameter `max-leaf nodes = None` refers to an unlimited number of leaf nodes, leaving the decision trees to grow to best fit the model. All of the daily hydrology and meteorological feature samples from 2006 to 2021 were used for training and testing the algorithm. A total of 80% of the dataset was used for the training, and 20% of the dataset was used for the testing of the Random Forest model.

A box plot of daily discharge was created to visualize the patterns of daily discharge as shown in Figure 5c. Daily runoff was checked by plotting the autocorrelation and partial autocorrelation factors. Figure 5a,b show the autocorrelation plot and the partial autocorrelation plot of historical daily runoff observations, respectively. These plots helped us identify a suitable lag period for flow prediction in a watershed [84]. Five sets of discharge values at a lag time of 1 to 5 days were selected to predict the discharge. Similarly, six sets of precipitation at 1 to 5 days of lag time were selected. Table 2 represents the combination of input features used to train the Random Forest Regression. In addition, the cumulative precipitation for 5 days and the day on which the rainfall was greater than 12.7 mm were considered as additional features for predicting the runoff at the outlet of the watershed. NumPy, Pandas, Matplotlib, stats model, Sklearn, and seaborn are the python libraries that were used during data processing, training, and visualization.

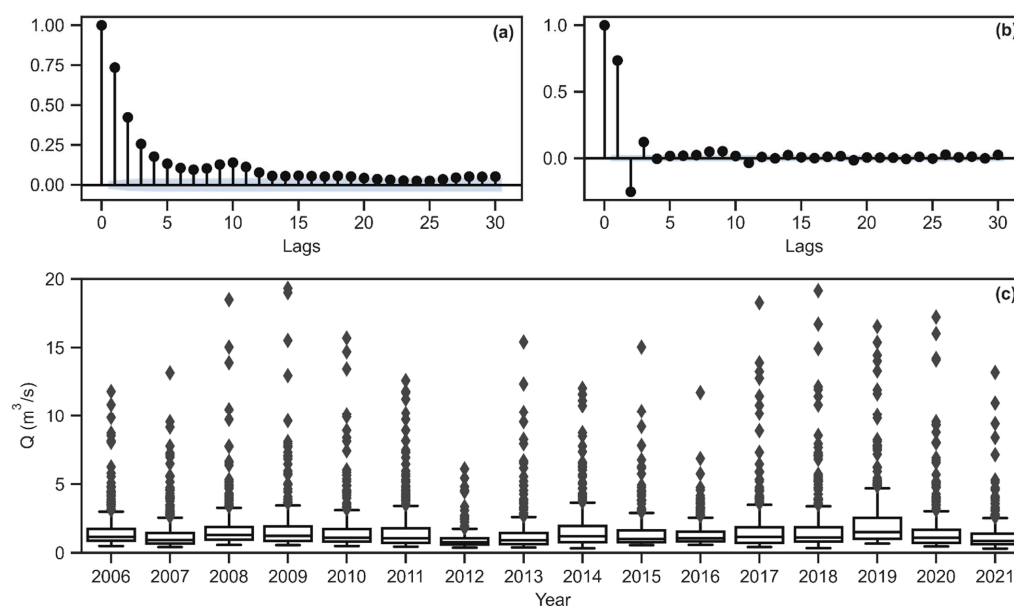


Figure 5. (a) Autocorrelation plot of the historical runoff observations of the DuPage River; (b) Partial autocorrelation plot of the historical runoff observations of the DuPage River; (c) box plot showing the flood events of the DuPage River.

Table 2. The combination of inputs for runoff prediction using Random Forest Regression.

Lag (Days)	The Structure of the Input	Output
5	Discharge of 1 day to the 5-day lag period, Precipitation of 1 day to the 5-day lag period, Sum of 5 days of precipitation (P5 days), Days since last precipitation greater than 0.5 mm. ($p > 0.5$)	One day ahead discharge

The autocorrelation function and the 95% confidence interval are shown in Figure 5a. A strong correlation was found up to 20 lags. The decay of autocorrelation shows the strength of the autoregressive process [29]. Similarly, the partial autocorrelation and 95% confidence interval were calculated. The partial autocorrelation depicted a strong correlation up to a 5-day lag period. Therefore, a lag period of 5 days was selected for the input [29].

2.6. Hydraulic Modeling Using HEC-RAS

Hydraulic modeling using HEC-RAS uses adequate geometry and flow data inputs for an excellent hydraulic model. The 1D HEC-RAS model is commonly employed to analyze flow in mainstream channels and predict the flood extent. Although the 1D model has limited applications, it is cost-effective, durable, and favored when determining flow pathways [85]. When speed is required and flood plain geometry data are scarce, 1D modeling is chosen [86]. HEC-RAS calculates the energy expression using Equation (4), which is based on Saint Venant's equation.

$$Z_2 + Y_2 + \frac{\alpha_2 V_2^2}{2g} = Z_1 + Y_1 + \frac{\alpha_1 V_1^2}{2g} + h_e \quad (4)$$

where

Y_1 and Y_2 = water heights at cross-sections,
 Z_1 and Z_2 = elevations of the stream reach,
 α_1 and α_2 = velocity weighting coefficients,
 V_1 and V_2 = average velocities,
 g = acceleration due to gravity, and
 h_e = energy head loss.

River Geometry Generation

Hydraulic analysis with HEC-RAS starts with extracting the river section geometry data using the RASMAP, which is available in the HEC-RAS model. The process involved in the hydraulic analysis using HEC-RAS is illustrated in the flowchart in Figure 1. The Lidar 1 m DEM for the hydraulic model was obtained from the USGS website. The DEM data were imported into the RAS Mapper tool in the HEC-RAS model and converted into a Digital Terrain Model. In addition, the georeferenced projection file was assigned in RASMAP for the consistent coordinate system. In the RASMAP, the river centerline, bank lines, flow path lines, and cross-section lines were digitized. The Manning's n value was assigned to each cross-section in the entire reach. After the creation of the river geometry and applying the Manning's n value, the steady discharge was used as input data for the steady flow simulation. The water depth achieved from the simulation was then compared to the water depth at gauging stations downstream of the study reach. The Manning's n values at the main channel and over banks were adjusted for the calibration of a model.

2.7. Statistical Performance Indicators

The performance of each model should be examined to determine the best models among different model alternatives. The five evaluation metrics (RMSE, RSR, NSE, PBIAS, and R^2) recommended by [87] and the NRMSE were used in this research to assess the

performance of the hydrology model. The criteria used to evaluate the proposed model's performance are listed in Table 3.

Table 3. List of statistical indexes used to determine the performance of models.

Indices	Mathematical Expression	Satisfactory Range
Root Mean Square Error (<i>RMSE</i>)	$RMSE = \sqrt{\frac{\sum_{i=1}^N (Q_{s,i} - Q_{o,i})^2}{N}}$	
Nash–Sutcliffe efficiency coefficient (<i>NSE</i>)	$NSE = 1 - \left[\frac{\sum_{i=1}^N (Q_{o,i} - Q_{s,i})^2}{\sum_{i=1}^N (Q_{o,i} - \bar{Q}_o)^2} \right]$	$0.5 < NSE \leq 1$
Coefficient of Determination (R^2)	$R^2 = \frac{(\sum_{i=1}^N (Q_{o,i} - \bar{Q}_o) * (Q_{s,i} - \bar{Q}_o))^2}{\sum_{i=1}^N (Q_{o,i} - \bar{Q}_o)^2 * \sum_{i=1}^N (Q_{s,i} - \bar{Q}_o)^2}$	>0.5
Standard Deviation Ratio (<i>RSR</i>)	$RSR = \frac{RMSE}{\text{standard Deviation}}$	$0 < RSR < 0.7$
Percentage bias (<i>PBIAS</i>)	$PBIAS = \frac{\sum_{i=1}^N (Q_{o,i} - Q_{s,i}) * 100}{\sum_{i=1}^N Q_{o,i}}$	$-25\% < PBIAS < +25\%$
Normalized Root Mean Squared Error (<i>NRMSE</i>)	$NRMSE = \frac{\frac{1}{N} \sum_{i=1}^N (Q_{s,i} - Q_{o,i})^2}{\text{Mean}}$	$\leq 30\%$

where $Q_{o,i}$ represents the observed data, $Q_{s,i}$ represents the simulated data from the model, \bar{Q}_o represents the mean value of the total number of observed data samples, and n represents the total number of data samples.

3. Results and Discussion

This section describes the results of the study, and it covers four main topics. In this section, the results of the precipitation product, hydrology, and hydraulic analyses are presented.

3.1. Precipitation

The rainfall data applied in this research were extracted from satellite-based rainfall products for a time period of 16 years (2006–2021). The daily rainfall data obtained for the studied time period are shown in Figure 6a. The daily precipitation data pattern was consistent with the daily observed discharge data. The result shows that the time of peak rainfall data matched the time of peak discharge data. For example, in this watershed outlet, the highest peak discharge of 33.7 m³/s was observed on 14 September 2008 and, similarly, the extracted precipitation product produced the highest precipitation of 61 mm on the same day. In addition, the validation of the extracted precipitation data was supported by the results of the hydrology analysis, which are presented in the following section.

3.2. HEC-HMS Models

Integration of the Arc-Hydro tool and HEC-GeoHMS successfully generated all the sub-basin parameters needed for the hydrology analysis. HEC-GeoHMS is a sophisticated tool that can be used to delineate natural watersheds and perform automatic basin parameter extraction for the HEC-HMS model construction. Table 4 lists the basin parameters obtained from HEC-GeoHMS, including sub-basin area, slope, curve number, and basin lag.

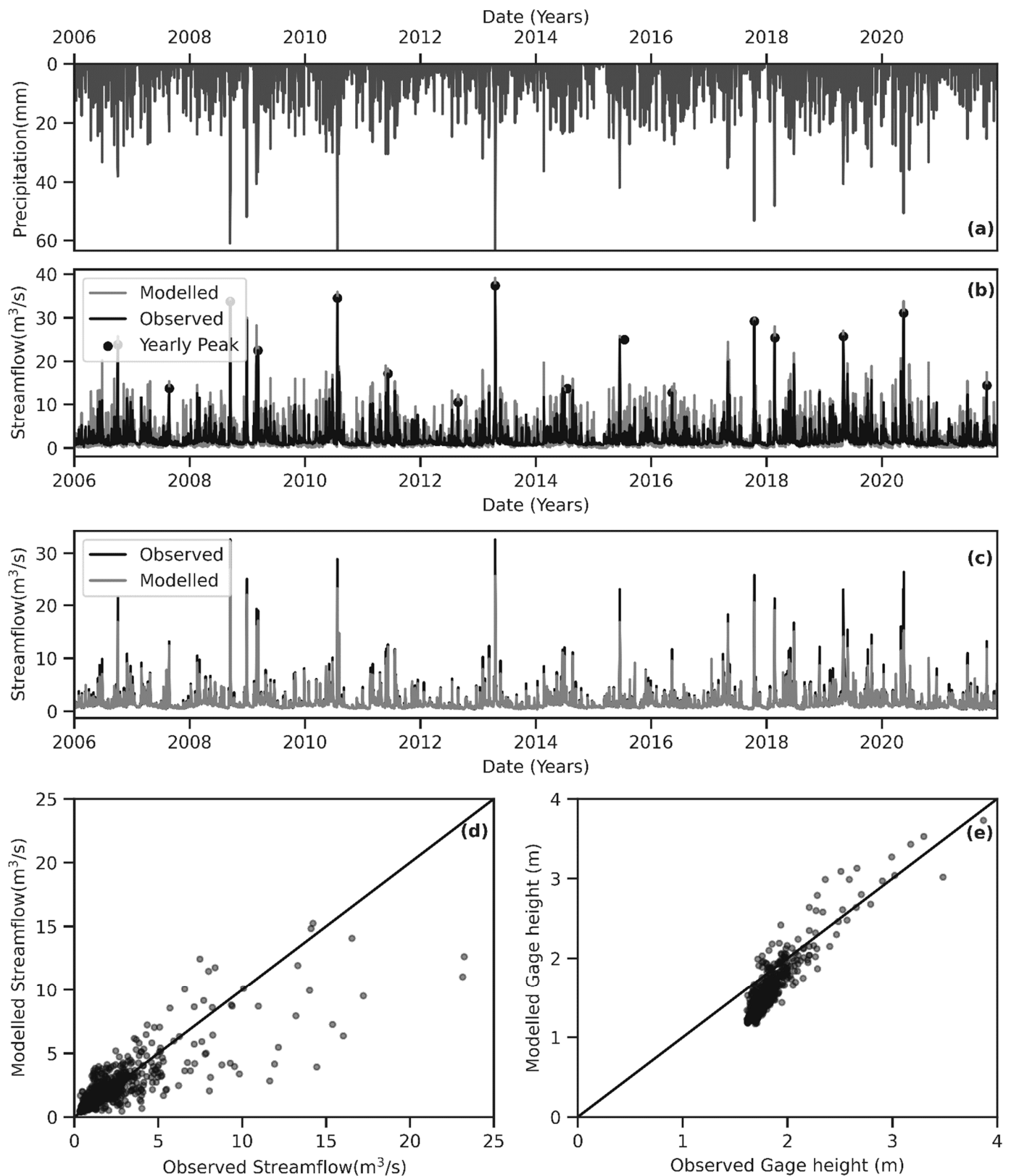


Figure 6. (a) Representation of the generated precipitation product; (b) training and testing of the HEC-HMS model; (c) observed discharge and predicted discharge for Random Forest Regression; (d) observed historical and predicted runoff data; (e) observed gauge height and simulated gauge height from HEC-RAS.

Table 4. Geographic characteristics of the study watershed.

Sub-Basin	Basin Area (km ²)	Basin Slope (%)	Curve Number (CN)	Basin Lag (min)
W220	4.3	2.6	85.8	150
W210	7.0	2.8	84.7	135
W200	3.6	3.1	83.6	133
W190	6.2	1.9	83.9	84
W180	5.9	3.5	83.2	90
W170	0.3	4.5	86.7	84
W160	3.7	2.6	82.3	81
W150	5.5	3.5	83.7	98
W140	7.4	4.5	83.0	86
W130	5.3	2.2	84.2	20
W120	13.0	3.4	84.0	76

The calibration and validation of the HEC-HMS model in this research were performed by adjusting the Muskingum parameters. The measured discharge from the gauging station was compared to the yearly peak discharge produced from an HEC-HMS simulation. Event 1 January 2006 to 31 December 2018 was considered for the model calibration, and Event 1 January 2019 to 31 December 2019 was used for the model validation. The accuracy of the hydrology model using HEC-HMS was determined using a statistical index. The discharge generated using HEC-HMS for the study period is presented in Figure 6b. The root mean square error is one of the most-used methods for evaluating the validity of predictions. The RMSE value during calibration and validation was 1.45 m³/s and 2.45 m³/s, respectively, which is considered a good result. The RSR is calculated by dividing the RMSE by the standard deviation of the measured data, and a value less than 0.7 is considered a good result [88]. The RSR values for the HEC-HMS model were 0.16 and 0.35. The NSE is extensively used in measuring the model performance in hydrology. It ranges from −1 to 1, with 0.5 to 1 being the best values. The NSE method is used to calculate the residual variance in relation to the variance in the measured data. The NSE values were 0.97 and 0.87, respectively, which are close to 1.

The PBIAS shows the average inclination of the calculated data. For a good model, PBIAS values must approach zero or should be between ±25% [89]. Positive numbers suggest that the model is underestimated, whereas negative values indicate that the model is overestimated [90]. The HEC-HMS model overestimated the peak discharge by 5.3% and 9.8% during calibration and validation, respectively. The R² is used to determine the correlation between calculated and measured flow rates. An R² greater than 0.5 indicates satisfactory performance. For the calibration and validation, the R² values were 0.99 and 0.96, respectively. The R² values close to 1 for the HEC-HMS model validated the accuracy of the model.

3.3. Random Forest Regression Model

Random Forest Regression provided good insights into the prediction of daily discharge data. Figure 6c illustrates the observed discharge data and the Random Forest predicted data during the study period. The scatter plot in Figure 6d demonstrates that the Random Forest prediction data were clustered near the regression line under low- and normal-flow conditions. However, Random Forest Regression slightly underestimated the high discharge value, which can also be termed an extreme event. Table 5 shows the evaluation matrix for Random Forest Regression. The RMSE, RSR, NSE, PBIAS, R², and NRMSE values were 0.29 m³/s, 0.23, 0.94, −0.75%, 0.94, and 0.17 for the training period and 0.47 m³/s, 0.56, 0.69, +1.76%, 0.72, and 0.260 for testing period, respectively, as shown in Table 5. The statistical index revealed that the Random Forest model's performance was superior during data training. The values of the statistical index dropped sharply during the testing period. The PBIAS values during training and testing were close to 0%, representing the average inclination of the predicted discharge towards the observed discharge. The values of R² dropped sharply from 0.94 during training to 0.72 during testing. However, the values of the statistical index were within acceptable ranges during the

testing period. Scatter plots were used to analyze the prediction performance of Random Forest Regression with the observed data. In the scatter plot between the observed and predicted values, the more significant deviation was observed for the higher discharge value, which also demonstrates the lower effectiveness of the Random Forest model for peak discharge estimation. The non-peak discharge was more accurately predicted by the machine learning model.

Table 5. Calibration and validation statistics of the HEC-HMS and Random Forest models.

Statistical Index	HEC-HMS Model		Random Forest	
	Calibration	Validation	Training	Testing
RMSE (m ³ /s)	1.45	2.45	0.29	0.47
RSR	0.16	0.35	0.23	0.56
NSE	0.97	0.87	0.94	0.69
PBIAS	−5.30%	−9.80%	−0.75%	+1.76%
R ²	0.99	0.96	0.94	0.72
NRMSE	0.06	0.10	0.17	0.26

Random Forest Regression was used for the prediction of discharge for the given input precipitation. The feature selection based on the lag period of precipitation and discharge was used. The validated results of HEC-HMS and Random Forest were compared to determine their ability to predict the discharge for the study period. After the comparison, we observed that the conventional HEC-HMS model needed more parameter optimization than Random Forest Regression. Similarly, the aim of study was also to prove the suitability of the discharge data predicted by Random Forest for hydraulic analysis. The scatter plot shown in Figure 6e shows the observed gauge height in the gauging station versus the simulated gauge height from the HEC-RAS model. During high-flooding events, the water depth predicted by the hydraulic model using the Random Forest generated discharge was slightly underestimated compared with the observed water depth. As the model showed good performance in generating the water depth under non-flooding conditions, the integration of Random Forest and HEC-RAS could be used to derive useful information while planning the water resource infrastructure and flood control measures in the selected study area. As the performance of a watershed model relies on the precision, robustness, and application of the model under other site conditions, the proposed approach could be tested and analyzed for multiple catchment locations, so that the parameters could be fixed to increase the reliability of the result.

3.4. HEC-RAS Model

The hydraulic analysis was carried out for the East DuPage watershed's downstream reach. For calibration purposes, historical discharge data from flood events in 2020 and 2021 were used, and the results are displayed in Table 6. The study reach consists of only one USGS gauging station at the most downstream location of the study reach with gauge height data beginning from 2020. The hydraulic model was calibrated using water depth data from various flooding events in 2021 and 2022. Figure 6e shows the comparison of simulated and observed data at most downstream stations of the study reach. The Manning's *n* value was adjusted to calibrate the hydraulic model. The water depth produced from the simulation was similar to the observed water depth at the gauging station, as shown in Figure 6e; this result demonstrates the model's consistency and allows it to be used for further investigation. At the upstream cross-section of the reach, daily discharge data from Random Forest were used to calculate the water depth at the downstream reach. The scatter plot in Figure 6e shows that the discharge calculated using Random Forest Regression can be utilized to calculate the flood depth in a river stream. Compared with the observed water depth at the gauging station, the model underestimates the simulated water depth generated from the study.

Table 6. The difference between the observed and simulated water depth.

Event	Discharge (m ³ /s)	Observed Water Depth (m)	Simulated Water Depth (m)	Difference (m)
11 January 2020	8.78	2.79	2.68	0.11
30 March 2020	3.11	2.09	1.98	0.11
29 March 2020	5.07	2.33	2.58	−0.25
30 April 2020	16.03	3.40	3.02	0.38
18 May 2020	26.42	4.41	3.85	0.56
23 October 2020	6.57	2.45	2.91	−0.46
12 December 2020	8.04	2.52	2.61	−0.09
19 March 2021	2.83	1.96	2.03	−0.07
26 June 2021	10.96	2.90	3.27	−0.37
27 August 2021	2.21	1.81	1.89	−0.08
26 October 2021	8.38	2.50	2.48	0.02

4. Discussion

The results of the hydrology simulation provide strong support for the effectiveness of the satellite precipitation product for the hydrology simulation in an ungauged catchment. Both the HEC-HMS and Random Forest models accurately recreated the discharge characteristics, such as the flood peak and timing, during the study period. These findings are consistent with those of previous studies that showed that PERSIANN-CCS precipitation products could effectively simulate the hydrology in ungauged watersheds [91,92]. The statistical index in Table 5 from the model calibration and validation suggests that Random Forest can be effectively applied for estimating the daily discharge at watershed outlets. The good performance of Random Forest for the hydrology analysis proved its appropriateness for rainfall-runoff simulation in data-scarce regions. The results of Random Forest are in agreement with a previous study's finding of good performance as an alternative prediction method in the hydrology domain [93]. The statistical index in Table 5 proved the suitability of both Random Forest and HEC-HMS for rainfall-runoff simulation. The results illustrate that Random Forest slightly underestimated the peak discharge during the high-flooding events; however, during the non-flooding period, the discharge predicted by Random Forest was better than that predicted by the HEC-HMS model. Figure 6e provided good support for the effectiveness of the Random-Forest-generated discharge for hydraulic simulation. The result indicates that the water depth simulated by HEC-RAS at the most downstream cross-section was slightly underestimated compared with the observed water depth at the gauging station. This result may be due to the use of the slightly underestimated peak discharge obtained from the Random Forest model. The overall result of this research work supports the integration of machine learning and a physical-based model for rainfall-runoff and flood depth prediction in data-scarce regions.

5. Conclusions

This study evaluated the feasibility of HEC-HMS and Random Forest for rainfall-runoff simulation and an integrated approach of machine learning and HEC-RAS for hydraulic analysis. HEC-HMS requires a large number of input variables, which may not always be available in a data-scarce region. In this scenario, the Random Forest model can be used for the prediction of discharge in the watershed. In addition, the Random Forest model is simple to build and takes less time. In this study, a PERSIANN-CCS NetCDF file was used to generate time-series precipitation data. The result supports the usage of PERSIANN-CCS daily precipitation data for rainfall-runoff simulation. Based on the models' reasonably strong performance, the obtained precipitation, LULC, DEM, and SSURGO soil input data are sufficiently dependable for discharge simulation. Because the data sources employed in this study yield reasonably reliable results, they are recommended for hydrology investigations. The continuous simulation of rainfall-runoff processes in the basin using physical and machine learning models yielded good results. Peak flows were underestimated in the Random Forest model and slightly overestimated in the HEC-HMS

model. An integrated HEC-RAS and Random Forest Regression model yielded good results in predicting the runoff flood depth downstream of a watershed. Given these findings, it is possible to say that the Random Forest model could aid in rainfall-runoff simulation as a complement to the physical model. This discharge could be used in hydraulic modeling for flood depth and flood extent analysis, which could be helpful to researchers in further research. The model's accuracy in predicting the flow can be increased by removing the outliers; high flood values are considered here in order to compensate for the prediction of the high flood values from Random Forest Regression. In the future, researchers could work in the following areas:

1. In this study, we used the PERSIANN precipitation product, and future work may be more accurate if there is a precipitation gauging station. Furthermore, researchers could also use other precipitation products, such as Next-Generation Weather Data (NEXRAD) and Climate Hazards Group Infrared Precipitation (CHIRPS);
2. In this study, precipitation was only used as an input variable for the Random Forest model; other variables, such as temperature, infiltration, evaporation, and radiation, could be used in future work. In addition, feature selection of input variables could be performed for the most accurate selection;
3. Other machine learning and data-driven models, such as support vector regression (SVR), long short-term memory (LSTM), and artificial neural networks (ANNs), could be used as prediction models. Future research directions could be guided by the selection of the best machine learning model in terms of accuracy, robustness, and reliability;
4. Although the study area is a small watershed in DuPage County, future research could focus on a more dynamic, heterogeneous, and meteorologically unique basin.

Author Contributions: Conceptualization, A.K.; formal analysis, A.B., investigation, A.B., U.P. and S.R.; software, A.B. and U.P.; supervision, A.K.; writing—initial draft preparation, A.B., U.P., S.R. and A.K.; writing—review & editing, A.B., U.P. and A.K. 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 available in a publicly accessible repository. The data download information is available in Table 1.

Acknowledgments: The authors would like to thank the reviewers for their valuable suggestions. The authors acknowledge the support of Southern Illinois University and Carbondale's Vice-Chancellor for Research. The research, simulation, and analysis were done with open-source software and datasets.

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

References

1. Merwade, V.; Olivera, F.; Arabi, M.; Edleman, S. Uncertainty in Flood Inundation Mapping: Current Issues and Future Directions. *J. Hydrol. Eng.* **2008**, *13*, 608–620. [[CrossRef](#)]
2. Merz, B.; Kreibich, H.; Schwarze, R.; Thielen, A. Review Article: Assessment of Economic Flood Damage. *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 1697–1724. [[CrossRef](#)]
3. Gaume, E.; Bain, V.; Bernardara, P.; Newinger, O.; Barbuc, M.; Bateman, A.; Blaškovičová, L.; Blöschl, G.; Borga, M.; Dumitrescu, A.; et al. A Compilation of Data on European Flash Floods. *J. Hydrol.* **2009**, *367*, 70–78. [[CrossRef](#)]
4. Ghazali, D.; Guericolas, M.; Thys, F.; Sarasin, F.; Arcos González, P.; Casalino, E. Climate Change Impacts on Disaster and Emergency Medicine Focusing on Mitigation Disruptive Effects: An International Perspective. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1379. [[CrossRef](#)] [[PubMed](#)]
5. Faccini, F.; Luino, F.; Paliaga, G.; Sacchini, A.; Turconi, L.; de Jong, C. Role of Rainfall Intensity and Urban Sprawl in the 2014 Flash Flood in Genoa City, Bisagno Catchment (Liguria, Italy). *Appl. Geogr.* **2018**, *98*, 224–241. [[CrossRef](#)]

6. Sapountzis, M.; Kastridis, A.; Kazamias, A.-P.; Karagiannidis, A.; Nikopoulos, P.; Lagouvardos, K. Utilization and Uncertainties of Satellite Precipitation Data in Flash Flood Hydrological Analysis in Ungauged Watersheds. *Glob. NEST J.* **2021**, *23*, 388–399. [[CrossRef](#)]
7. Pathak, P.; Kalra, A.; Ahmad, S. Temperature and Precipitation Changes in the Midwestern United States: Implications for Water Management. *Int. J. Water Resour. Dev.* **2017**, *33*, 1003–1019. [[CrossRef](#)]
8. Jenkins, K.; Surminski, S.; Hall, J.; Crick, F. Assessing Surface Water Flood Risk and Management Strategies under Future Climate Change: Insights from an Agent-Based Model. *Sci. Total Environ.* **2017**, *595*, 159–168. [[CrossRef](#)]
9. Kundzewicz, Z.W.; Kanae, S.; Seneviratne, S.I.; Handmer, J.; Nicholls, N.; Peduzzi, P.; Mechler, R.; Bouwer, L.M.; Arnell, N.; Mach, K.; et al. Flood Risk and Climate Change: Global and Regional Perspectives. *Hydrol. Sci. J.* **2014**, *59*, 1–28. [[CrossRef](#)]
10. Guerreiro, S.B.; Dawson, R.J.; Kilsby, C.; Lewis, E.; Ford, A. Future Heat-Waves, Droughts and Floods in 571 European Cities. *Environ. Res. Lett.* **2018**, *13*, 034009. [[CrossRef](#)]
11. Min, S.-K.; Zhang, X.; Zwiers, F.W.; Hegerl, G.C. Human Contribution to More-Intense Precipitation Extremes. *Nature* **2011**, *470*, 378–381. [[CrossRef](#)] [[PubMed](#)]
12. Vörösmarty, C.J.; de Guenni, L.B.; Wollheim, W.M.; Pellerin, B.; Bjerklie, D.; Cardoso, M.; D’Almeida, C.; Green, P.; Colon, L. Extreme Rainfall, Vulnerability and Risk: A Continental-Scale Assessment for South America. *Philos. Trans. R. Soc. A* **2013**, *371*, 20120408. [[CrossRef](#)] [[PubMed](#)]
13. Woznicki, S.A.; Baynes, J.; Panlasigui, S.; Mehaffey, M.; Neale, A. Development of a Spatially Complete Floodplain Map of the Conterminous United States Using Random Forest. *Sci. Total Environ.* **2019**, *647*, 942–953. [[CrossRef](#)] [[PubMed](#)]
14. Archer, D.R.; Fowler, H.J. Characterising Flash Flood Response to Intense Rainfall and Impacts Using Historical Information and Gauged Data in Britain: Flash Flood Response to Intense Rainfall in Britain. *J. Flood Risk Manag.* **2018**, *11*, S121–S133. [[CrossRef](#)]
15. Kastridis, A.; Stathis, D. The Effect of Rainfall Intensity on the Flood Generation of Mountainous Watersheds (Chalkidiki Prefecture, North Greece). In *Perspectives on Atmospheric Sciences*; Karacostas, T., Bais, A., Nastos, P.T., Eds.; Springer Atmospheric Sciences; Springer International Publishing: Cham, Switzerland, 2017; pp. 341–347. ISBN 978-3-319-35094-3.
16. Schoppa, L.; Disse, M.; Bachmair, S. Evaluating the Performance of Random Forest for Large-Scale Flood Discharge Simulation. *J. Hydrol.* **2020**, *590*, 125531. [[CrossRef](#)]
17. Talei, A.; Chua, L.H.C.; Quek, C. A Novel Application of a Neuro-Fuzzy Computational Technique in Event-Based Rainfall–Runoff Modeling. *Expert Syst. Appl.* **2010**, *37*, 7456–7468. [[CrossRef](#)]
18. Singh, V.P.; Frevert, D.K. *Watershed Models*; Taylor and Francis: Abingdon, UK, 2005.
19. Halwatura, D.; Najim, M.M.M. Application of the HEC-HMS Model for Runoff Simulation in a Tropical Catchment. *Environ. Model. Softw.* **2013**, *46*, 155–162. [[CrossRef](#)]
20. US Army Corps of Engineers. *Hydrologic Modeling System (HEC-HMS) Application Guide Version 3.1.0*; Institute for Water Resources: Davis, CA, USA, 2008.
21. Bajwa, H.S.; Tim, U.S. *Toward Immersive Virtual Environments for GIS-Based Floodplain Modeling and Visualization*; ESRI: Redlands, CA, USA, 2002.
22. Senthil Kumar, A.; Sudheer, K.; Jain, S.; Agarwal, P. Rainfall-runoff modelling using artificial neural networks: Comparison of network types. *Hydrol. Process.* **2005**, *19*, 1277–1291. [[CrossRef](#)]
23. Rezaeianzadeh, M.; Stein, A.; Tabari, H.; Abghari, H.; Jalalkamali, N.; Hosseinipour, E.Z.; Singh, V.P. Assessment of a Conceptual Hydrological Model and Artificial Neural Networks for Daily Outflows Forecasting. *Int. J. Environ. Sci. Technol.* **2013**, *10*, 1181–1192. [[CrossRef](#)]
24. 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](#)]
25. Sahoo, S.; Russo, T.A.; Elliott, J.; Foster, I. Machine Learning Algorithms for Modeling Groundwater Level Changes in Agricultural Regions of the U.S. *Water Resour. Res.* **2017**, *53*, 3878–3895. [[CrossRef](#)]
26. Rajaei, T.; Khani, S.; Ravansalar, M. Artificial Intelligence-Based Single and Hybrid Models for Prediction of Water Quality in Rivers: A Review. *Chemom. Intell. Lab. Syst.* **2020**, *200*, 103978. [[CrossRef](#)]
27. Zounemat-Kermani, M.; Batelaan, O.; Fadaee, M.; Hinkelmann, R. Ensemble Machine Learning Paradigms in Hydrology: A Review. *J. Hydrol.* **2021**, *598*, 126266. [[CrossRef](#)]
28. Jordan, M.I.; Mitchell, T.M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349*, 255–260. [[CrossRef](#)] [[PubMed](#)]
29. Ghimire, S.; Yaseen, Z.M.; Farooque, A.A.; Deo, R.C.; Zhang, J.; Tao, X. OPEN Streamflow Prediction Using. *Sci. Rep.* **2021**, *11*, 17497. [[CrossRef](#)]
30. Mewes, B.; Oppel, H.; Marx, V.; Hartmann, A. Information-Based Machine Learning for Tracer Signature Prediction in Karstic Environments. *Water Resour. Res.* **2020**, *56*, e2018WR024558. [[CrossRef](#)]
31. Parisouj, P.; Mohebzadeh, H.; Lee, T. Employing Machine Learning Algorithms for Streamflow Prediction: A Case Study of Four River Basins with Different Climatic Zones in the United States. *Water Resour. Manag.* **2020**, *34*, 4113–4131. [[CrossRef](#)]
32. Adnan, R.M.; Petroselli, A.; Heddam, S.; Santos, C.A.G.; Kisi, O. Short Term Rainfall-Runoff Modelling Using Several Machine Learning Methods and a Conceptual Event-Based Model. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 597–616. [[CrossRef](#)]

33. Shamshirband, S.; Hashemi, S.; Salimi, H.; Samadianfard, S.; Asadi, E.; Shadkani, S.; Kargar, K.; Mosavi, A.; Nabipour, N.; Chau, K.-W. Predicting Standardized Streamflow Index for Hydrological Drought Using Machine Learning Models. *Eng. Appl. Comput. Fluid Mech.* **2020**, *14*, 339–350. [[CrossRef](#)]
34. Nguyen, D.T.; Chen, S.-T. Real-Time Probabilistic Flood Forecasting Using Multiple Machine Learning Methods. *Water* **2020**, *12*, 787. [[CrossRef](#)]
35. Zhou, Y.; Cui, Z.; Lin, K.; Sheng, S.; Chen, H.; Guo, S.; Xu, C.-Y. Short-Term Flood Probability Density Forecasting Using a Conceptual Hydrological Model with Machine Learning Techniques. *J. Hydrol.* **2022**, *604*, 127255. [[CrossRef](#)]
36. Kalra, A.; Ahmad, S.; Nayak, A. Increasing Streamflow Forecast Lead Time for Snowmelt-Driven Catchment Based on Large-Scale Climate Patterns. *Adv. Water Resour.* **2013**, *53*, 150–162. [[CrossRef](#)]
37. Rezaei, K.; Pradhan, B.; Vadiati, M.; Nadiri, A.A. Suspended Sediment Load Prediction Using Artificial Intelligence Techniques: Comparison between Four State-of-the-Art Artificial Neural Network Techniques. *Arab. J. Geosci.* **2021**, *14*, 215. [[CrossRef](#)]
38. Choubin, B.; Darabi, H.; Rahmati, O.; Sajedi-Hosseini, F.; Kløve, B. River Suspended Sediment Modelling Using the CART Model: A Comparative Study of Machine Learning Techniques. *Sci. Total Environ.* **2018**, *615*, 272–281. [[CrossRef](#)] [[PubMed](#)]
39. Rezaei, K.; Vadiati, M. A Comparative Study of Artificial Intelligence Models for Predicting Monthly River Suspended Sediment Load. *J. Water Land Dev.* **2020**, *45*, 107–118. [[CrossRef](#)]
40. Wang, S.; Peng, H.; Liang, S. Prediction of Estuarine Water Quality Using Interpretable Machine Learning Approach. *J. Hydrol.* **2022**, *605*, 127320. [[CrossRef](#)]
41. Deng, T.; Chau, K.-W.; Duan, H.-F. Machine Learning Based Marine Water Quality Prediction for Coastal Hydro-Environment Management. *J. Environ. Manag.* **2021**, *284*, 112051. [[CrossRef](#)]
42. Melesse, A.M.; Khosravi, K.; Tiefenbacher, J.P.; Heddam, S.; Kim, S.; Mosavi, A.; Pham, B.T. River Water Salinity Prediction Using Hybrid Machine Learning Models. *Water* **2020**, *12*, 2951. [[CrossRef](#)]
43. Asadollah, S.B.H.S.; Sharafati, A.; Motta, D.; Yaseen, Z.M. River Water Quality Index Prediction and Uncertainty Analysis: A Comparative Study of Machine Learning Models. *J. Environ. Chem. Eng.* **2021**, *9*, 104599. [[CrossRef](#)]
44. Hussein, E.A.; Thron, C.; Ghaziasgar, M.; Bagula, A.; Vaccari, M. Groundwater Prediction Using Machine-Learning Tools. *Algorithms* **2020**, *13*, 300. [[CrossRef](#)]
45. Khedri, A.; Kalantari, N.; Vadiati, M. Comparison Study of Artificial Intelligence Method for Short Term Groundwater Level Prediction in the Northeast Gachsaran Unconfined Aquifer. *Water Supply* **2020**, *20*, 909–921. [[CrossRef](#)]
46. Zhu, S.; Piotrowski, A.P. River/Stream Water Temperature Forecasting Using Artificial Intelligence Models: A Systematic Review. *Acta Geophys.* **2020**, *68*, 1433–1442. [[CrossRef](#)]
47. Chang, H.; Psaris, M. Local Landscape Predictors of Maximum Stream Temperature and Thermal Sensitivity in the Columbia River Basin, USA. *Sci. Total Environ.* **2013**, *461–462*, 587–600. [[CrossRef](#)] [[PubMed](#)]
48. Weierbach, H.; Lima, A.R.; Willard, J.D.; Hendrix, V.C.; Christianson, D.S.; Lubich, M.; Varadharajan, C. Stream Temperature Predictions for River Basin Management in the Pacific Northwest and Mid-Atlantic Regions Using Machine Learning. *Water* **2022**, *14*, 1032. [[CrossRef](#)]
49. Feigl, M.; Lebedzinski, K.; Herrnegger, M.; Schulz, K. Machine-learning methods for stream water temperature prediction. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 2951–2977. [[CrossRef](#)]
50. Zhang, J.; Xu, J.; Dai, X.; Ruan, H.; Liu, X.; Jing, W. Multi-Source Precipitation Data Merging for Heavy Rainfall Events Based on Cokriging and Machine Learning Methods. *Remote Sens.* **2022**, *14*, 1750. [[CrossRef](#)]
51. Radhakrishnan, C.; Chandrasekar, V.; Reising, S.C.; Berg, W. Rainfall Estimation from TEMPEST-D CubeSat Observations: A Machine-Learning Approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3626–3636. [[CrossRef](#)]
52. Guo, W.-D.; Chen, W.-B.; Yeh, S.-H.; Chang, C.-H.; Chen, H. Prediction of River Stage Using Multistep-Ahead Machine Learning Techniques for a Tidal River of Taiwan. *Water* **2021**, *13*, 920. [[CrossRef](#)]
53. Chiang, S.; Chang, C.-H.; Chen, W.-B. Comparison of Rainfall-Runoff Simulation between Support Vector Regression and HEC-HMS for a Rural Watershed in Taiwan. *Water* **2022**, *14*, 191. [[CrossRef](#)]
54. Ni, L.; Wang, D.; Singh, V.P.; Wu, J.; Wang, Y.; Tao, Y.; Zhang, J. Streamflow and Rainfall Forecasting by Two Long Short-Term Memory-Based Models. *J. Hydrol.* **2020**, *583*, 124296. [[CrossRef](#)]
55. Yin, H.; Wang, F.; Zhang, X.; Zhang, Y.; Chen, J.; Xia, R.; Jin, J. Rainfall-Runoff Modeling Using Long Short-Term Memory Based Step-Sequence Framework. *J. Hydrol.* **2022**, *610*, 127901. [[CrossRef](#)]
56. Tikhmarine, Y.; Souag-Gamane, D.; Ahmed, A.N.; Sammen, S.S.; Kisi, O.; Huang, Y.F.; El-Shafie, A. Rainfall-Runoff Modelling Using Improved Machine Learning Methods: Harris Hawks Optimizer vs. Particle Swarm Optimization. *J. Hydrol.* **2020**, *589*, 125133. [[CrossRef](#)]
57. Tamiru, H.; Dinka, M.O. Application of ANN and HEC-RAS Model for Flood Inundation Mapping in Lower Baro Akobo River Basin, Ethiopia. *J. Hydrol. Reg. Stud.* **2021**, *36*, 100855. [[CrossRef](#)]
58. Samantaray, S.; Das, S.S.; Sahoo, A.; Satapathy, D.P. Monthly Runoff Prediction at Baitarani River Basin by Support Vector Machine Based on Salp Swarm Algorithm. *Ain Shams Eng. J.* **2022**, *13*, 101732. [[CrossRef](#)]
59. Adnan, R.M.; Liang, Z.; Heddam, S.; Zounemat-Kermani, M.; Kisi, O.; Li, B. Least Square Support Vector Machine and Multivariate Adaptive Regression Splines for Streamflow Prediction in Mountainous Basin Using Hydro-Meteorological Data as Inputs. *J. Hydrol.* **2020**, *586*, 124371. [[CrossRef](#)]

60. Worland, S.C.; Farmer, W.H.; Kiang, J.E. Improving Predictions of Hydrological Low-Flow Indices in Ungaged Basins Using Machine Learning. *Environ. Model. Softw.* **2018**, *101*, 169–182. [[CrossRef](#)]
61. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
62. Zhou, P.; Li, Z.; Snowling, S.; Baetz, B.W.; Na, D.; Boyd, G. A Random Forest Model for Inflow Prediction at Wastewater Treatment Plants. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 1781–1792. [[CrossRef](#)]
63. Meng, Y.; Yang, M.; Liu, S.; Mou, Y.; Peng, C.; Zhou, X. Quantitative Assessment of the Importance of Bio-Physical Drivers of Land Cover Change Based on a Random Forest Method. *Ecol. Inform.* **2021**, *61*, 101204. [[CrossRef](#)]
64. Li, B.; Yang, G.; Wan, R.; Dai, X.; Zhang, Y. Comparison of Random Forests and Other Statistical Methods for the Prediction of Lake Water Level: A Case Study of the Poyang Lake in China. *Hydrol. Res.* **2016**, *47*, 69–83. [[CrossRef](#)]
65. Bachmair, S.; Svensson, C.; Prosdocimi, I.; Hannaford, J.; Stahl, K. Developing Drought Impact Functions for Drought Risk Management. *Nat. Hazards Earth Syst. Sci.* **2017**, *17*, 1947–1960. [[CrossRef](#)]
66. Erdal, H.I.; Karakurt, O. Advancing Monthly Streamflow Prediction Accuracy of CART Models Using Ensemble Learning Paradigms. *J. Hydrol.* **2013**, *477*, 119–128. [[CrossRef](#)]
67. Muñoz, P.; Orellana-Alvear, J.; Willems, P.; Céleri, R. Flash-Flood Forecasting in an Andean Mountain Catchment—Development of a Step-Wise Methodology Based on the Random Forest Algorithm. *Water* **2018**, *10*, 1519. [[CrossRef](#)]
68. Tyralis, 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](#)]
69. 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](#)]
70. Feng, Q.; Liu, J.; Gong, J. Urban Flood Mapping Based on Unmanned Aerial Vehicle Remote Sensing and Random Forest Classifier—A Case of Yuyao, China. *Water* **2015**, *7*, 1437–1455. [[CrossRef](#)]
71. Quiroga, V.M.; Kurea, S.; Udo, K.; Manoa, A. Application of 2D Numerical Simulation for the Analysis of the February 2014 Bolivian Amazonia Flood: Application of the New HEC-RAS Version 5. *Ribagua* **2016**, *3*, 25–33. [[CrossRef](#)]
72. Brunner, G. *HEC-RAS, River Analysis System Hydraulic Reference Manual*; U.S. Army Corps of Engineers: Davis, CA, USA, 2016.
73. İcaga, Y.; Tas, E.; Kilit, M. Flood Inundation Mapping by GIS and a Hydraulic Model (Hec Ras): A Case Study of Akarcay Bolvadin Subbasin, in Turkey. *Acta Geobalcánica* **2016**, *2*, 111–118. [[CrossRef](#)]
74. Abaya, S.W.; Mandere, N.; Ewald, G. Floods and Health in Gambella Region, Ethiopia: A Qualitative Assessment of the Strengths and Weaknesses of Coping Mechanisms. *Glob. Health Action* **2009**, *2*, 2019. [[CrossRef](#)]
75. US Army Corps of Engineers. *Dupage River, Illinois Feasibility Report and Integrated Environmental Assessment*; US Army Corps of Engineers: Chicago, IL, USA, 2019.
76. StreamStats. Available online: <https://Streamstats.Usgs.Gov/Ss/> (accessed on 15 June 2022).
77. Nguyen, P.; Shearer, E.J.; Tran, H.; Ombadi, M.; Hayatbini, N.; Palacios, T.; Huynh, P.; Braithwaite, D.; Updegraff, G.; Hsu, K.; et al. The CHRS Data Portal, an Easily Accessible Public Repository for PERSIANN Global Satellite Precipitation Data. *Sci. Data* **2019**, *6*, 180296. [[CrossRef](#)]
78. Mockus, V. *National Engineering Handbook Section 4 Hydrology*; US Soil Conservation Service: Washington, DC, USA, 1972; p. 127.
79. Saadi, M.; Oudin, L.; Ribstein, P. Random Forest Ability in Regionalizing Hourly Hydrological Model Parameters. *Water* **2019**, *11*, 40. [[CrossRef](#)]
80. Müller, A.; Guido, S. *Introduction to Machine Learning with Python: A Guide for Data Scientists*, 1st ed.; O’Reilly: Farnham, UK, 2016.
81. Park, H.; Kim, K.; Lee, D.K. Prediction of Severe Drought Area Based on Random Forest: Using Satellite Image and Topography Data. *Water* **2019**, *11*, 705. [[CrossRef](#)]
82. Biau, G.; Scornet, E. A Random Forest Guided Tour. *Test* **2016**, *25*, 197–227. [[CrossRef](#)]
83. Gregorutti, B.; Michel, B.; Saint-Pierre, P. Correlation and Variable Importance in Random Forests. *Stat. Comput.* **2017**, *27*, 659–678. [[CrossRef](#)]
84. Hussain, D.; Khan, A.A. Machine Learning Techniques for Monthly River Flow Forecasting of Hunza River, Pakistan. *Earth Sci. Inf.* **2020**, *13*, 939–949. [[CrossRef](#)]
85. Gharbi, M.; Soualmia, A.; Dartus, D.; Masbernat, L. Comparison of 1D and 2D Hydraulic Models for Floods Simulation on the Medjerda River in Tunisia. *J. Mater. Environ. Sci.* **2016**, *7*, 3017–3026.
86. Pathan, A.I.; Agnihotri, P.G. Application of New HEC-RAS Version 5 for 1D Hydrodynamic Flood Modeling with Special Reference through Geospatial Techniques: A Case of River Purna at Navsari, Gujarat, India. *Model. Earth Syst. Environ.* **2021**, *7*, 1133–1144. [[CrossRef](#)]
87. Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. *Trans. ASABE* **2015**, *58*, 1763–1785. [[CrossRef](#)]
88. Kumar, N.; Singh, S.K.; Srivastava, P.K.; Narsimlu, B. SWAT Model Calibration and Uncertainty Analysis for Streamflow Prediction of the Tons River Basin, India, Using Sequential Uncertainty Fitting (SUFI-2) Algorithm. *Model. Earth Syst. Environ.* **2017**, *3*, 30. [[CrossRef](#)]
89. Abbaspour, K.C.; Rouholahnejad, E.; Vaghefi, S.; Srinivasan, R.; Yang, H.; Kløve, B. A Continental-Scale Hydrology and Water Quality Model for Europe: Calibration and Uncertainty of a High-Resolution Large-Scale SWAT Model. *J. Hydrol.* **2015**, *524*, 733–752. [[CrossRef](#)]

90. Gupta, H.V.; Sorooshian, S.; Yapo, P.O. Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. *J. Hydrol. Eng.* **1999**, *4*, 135–143. [[CrossRef](#)]
91. Hong, Y.; Gochis, D.; Cheng, J.; Hsu, K.; Sorooshian, S. Evaluation of PERSIANN-CCS Rainfall Measurement Using the NAME Event Rain Gauge Network. *J. Hydrometeorol.* **2007**, *8*, 469–482. [[CrossRef](#)]
92. Joshi, N.; Bista, A.; Pokhrel, I.; Kalra, A.; Ahmad, S. Rainfall-Runoff Simulation in Cache River Basin, Illinois, Using HEC-HMS. In *World Environmental and Water Resources Congress 2019*; American Society of Civil Engineers: Pittsburgh, PA, USA, 2019; pp. 348–360.
93. Desai, S.; Ouarda, T.B.M.J. Regional Hydrological Frequency Analysis at Ungauged Sites with Random Forest Regression. *J. Hydrol.* **2021**, *594*, 125861. [[CrossRef](#)]