5-1-2023

EVALUATING THE PERFORMANCE OF PROCESS-BASED AND MACHINE LEARNING MODELS FOR RAINFALL-RUNOFF SIMULATION WITH APPLICATION OF SATELLITE AND RADAR PRECIPITATION PRODUCTS

Amrit Bhusal
Southern Illinois University Carbondale, amrit.bhusal96@gmail.com

EVALUATING THE PERFORMANCE OF PROCESS-BASED AND MACHINE LEARNING MODELS FOR RAINFALL-RUNOFF SIMULATION WITH APPLICATION OF SATELLITE AND RADAR PRECIPITATION PRODUCTS

by

Amrit Bhusal

B.E., Kathmandu University, 2019

A Thesis
Submitted in Partial Fulfillment of the Requirements for the
Master of Science Degree

School of Civil, Environmental, and Infrastructure Engineering
in the Graduate School
Southern Illinois University Carbondale
May 2023

THESIS APPROVAL

EVALUATING THE PERFORMANCE OF PROCESS-BASED AND MACHINE LEARNING MODELS FOR RAINFALL-RUNOFF SIMULATION WITH APPLICATION OF SATELLITE AND RADAR PRECIPITATION PRODUCTS

by

Amrit Bhusal

A Thesis Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Master of Science

in the field of Civil Engineering

Approved by:

Dr. Ajay Kalra, Chair

Dr. Prabir Kolay

Dr. Habibollah Fakhraei

Graduate School
Southern Illinois University Carbondale
December 13, 2022

AN ABSTRACT OF THE THESIS OF

Amrit Bhusal, for the Master of Science degree in Civil Engineering, presented on December 13, 2022, at Southern Illinois University Carbondale.

TITLE: EVALUATING THE PERFORMANCE OF PROCESS-BASED AND MACHINE LEARNING MODELS FOR RAINFALL-RUNOFF SIMULATION WITH APPLICATION OF SATELLITE AND RADAR PRECIPITATION PRODUCTS

MAJOR PROFESSOR: Dr. Ajay Kalra

Hydrology Modeling using HEC-HMS (Hydrological Engineering Centre-Hydrologic Modeling System) 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. In addition, Point gauge observations have historically been the primary source of the necessary rainfall data for hydrologic models. However, point gauge observation does not provide accurate information on rainfall's spatial and temporal variability, which is vital for hydrological models. Therefore, this study also evaluates the performance of satellite and radar precipitation products for hydrological analysis. The results revealed that integrated Machine Learning and physical-based model could provide more confidence in rainfall-runoff and flood depth prediction. Similarly, the study revealed that radar data performance was superior to the gauging station's rainfall data for the hydrologic analysis in large watersheds. The discussions in this research will encourage researchers and system managers to improve current rainfall-runoff simulation models by application of Machine learning and radar rainfall data.

ACKNOWLEDGEMENTS

I would like to convey my profound gratitude to Dr. Ajay Kalra, my thesis committee chair and advisor, for his valuable time, insightful discussion, and supervision throughout my research, which helped me to achieve this project. I also want to thank Dr. Prabir Kolay and Dr. Habibollah Fakhraei, who served on my thesis committee, for their knowledge, unwavering support, and advice in advancing my research understanding.

I would like to thank Dr. Sangmin Shin for his supervision and providing me with funding during my Master Studies. I would like to thank Computational Hydraulics International for providing free academic access to PCSWMM through a university grant. I would like to thank the University of Illinois System (Award #107688) for providing research support for the current study.

I would also like to thank the Department of Civil, Environmental, and Infrastructure Engineering for providing me with a platform to conduct this research. I appreciate Jennifer, the office manager, for making the required logistical arrangements. I want to thank my seniors, Dr. Balbhadra Thakur, and Dr. Neekita Joshi, for their guidance during this project. I also appreciate the assistance of my coworkers Mr. Utsav Parajuli, Mr. Amrit Babu Ghimire, Mr. Anjan Parajuli, Mr. Abhiru Aryal, Ms. Albira Aacharya, and Mr. Mandip Banjara. I also want to express my gratitude to my friends Amit Neupane, Nation Adhikari, Jeewasmi Gouli, and Babin Dangal for supporting me throughout my graduate studies.

My tremendous gratefulness to my father, mother, and family for their unconditional love and support throughout my educational and personal journey. I also want to thank all of my colleagues, professors, and staff at SIUC's School of Civil, Environmental, and Infrastructure Engineering for their support and collaboration throughout my graduate degree.

TABLE OF CONTENTS

CHAPTER PAGE

ABSTRACT…………….. i

ACKNOWLEDGEMENTS ii

LIST OF TABLES …... vi

LIST OF FIGURES………………………………………………………………………. vii

CHAPTERS

CHAPTER 1- Introduction 1

1.1 - Research background 1

1.2 - Problem statetment and objectives 3

1.3 Research outline 4

CHAPTER 2 - Application of machine learning and process-based models for rainfall-runoff simulation in Dupage river basin, Illinois 5

2.1 - Introduction 5

2.2 - Data and methods 10

2.2.1 - Study area 11

2.2.2 - Data.. 13

2.2.3 - Preprocessing data 13

2.2.3.1 - Digital elevation model 13

2.2.3.2 - Basin characteristics 14

2.2.3.3 - Precipitation data 14

2.2.4 - Hydrologic modelling using ARC-GIS and HEC-HMS 15

2.2.4.1 - Loss method 16

2.2.4.2 - Transform method 16

2.2.4.3 - Routing method 17

2.2.5 - Hydrologic modelling using random forest 17

2.2.5.1 - Model development 18

2.2.6 - Hydraulic modelling 20

2.2.6.1 - River geometry 21

2.2.7 - Statistical performance indicator 21

2.3 - Results 22

2.3.1 - Precipitation 22

2.3.2 - HEC-HMS model 23

2.3.3 - Random forest regression 25

2.3.4 - HEC-RAS model 28

2.4 - Discussion 29

2.5 - Conclusion 30

CHAPTER 3 - Evaluating the performance of PCSWMM using NEXRAD data for urbanized watersheds 33

3.1 - Introduction 33

3.2 - Methodology 37

3.2.1 - Site description 37

3.2.2 - Data 38

3.2.3 - NEXRAD radar data 39

3.2.4 - Hydrological modelling using PCSWMM 39

3.2.5 - Radar processing using PCSWMM in RAP project 41

3.2.6 - Rainfall event selection 42

3.2.7 - Model performance evaluation 42

3.3 - Results and discussion 43

3.3.1 - Ellerbe creek watershed 43

3.3.1.1 - NEXRAD rainfall estimates 43

3.3.1.2 - Streamflow analysis 45

3.3.2 - River des peres watershed 51

3.4 - Conclusion 52

CHAPTER 4 - Conclusion and recommendation 55

REFERENCES……….. 60

APPENDIX - PUBLISHER PERMISSION……………………………………………..……..83

VITA………………….. 84

LIST OF TABLES

TABLE PAGE

Table 1- Data Used for this research with their sources. 13

Table 2 - The combination of input for runoff prediction using random forest regression 20

Table 3 - List of statistical indexes for determining the performance of models. 22

Table 4 - Geographic characteristics of the study watershed. 24

Table 5 - Calibration and Validation summary statistics of HEC-HMS and Random Forest 28

Table 6 - Observed and simulated water depth. 29

Table 7 - Selection of Rainfall events 42

Table 8 - Indices applied for evaluation criteria 43

Table 9 - Comparison of observed and NEXRAD generated rainfall data 44

Table 10 - Statistical evaluation of gauge and radar generated discharge hydrograph at node 4. 46

Table 11 - Statistical evaluation of gauge and radar generated discharge hydrograph at Ellerbe creek watershed outlet 49

Table 12 - Table 6 Statistical evaluation of radar generated discharge hydrograph at node River Des Peres watershed outlet 51

LIST OF FIGURES

FIGURE PAGE

Figure 1 - Figure portraying the flowchart of hydrology simulation using Random Forest and HEC-HMS and hydraulic analysis using HEC-RAS.. 11

Figure 2 - The East Branch DuPage Catchment around Downers Grove, Illinois, with the river system. 12

Figure 3 - Map depicting characteristics of study area. 14

Figure 4 - Pre-processing and model development... 15

Figure 5 - a) Autocorrelation plot of the historical runoff observation of the DuPage River; b) Partial autocorrelation plot of the historical runoff observation of the DuPage River; c) Box Plot showing the flood events of the DuPage River. 19

Figure 6 - a) Representation of generated precipitation product; b) Training and testing for HEC-HMS Model; c) observed discharge and predicted discharge for random forest regression; d) Observed historical and Predicted runoff data, e) Observed gage height and Predicted gage height.. 27

Figure 7 - The figure depicting the characteristics of study watersheds and their location.......... 38

Figure 8 - Watersheds discretization for modelling.. 40

Figure 9 - Regression analysis between gauge recorded and NEXRAD estimated rainfall data, (a) Event 1, (b) Event 2, (c) Event 3, (d) Event 4.. 45

Figure 10 - Graphical representation of observed discharge, radar and gauge generated discharge at node 4, (a) Event 1, (b) Event 2, (c) Event 3, (d) Event 4 47

Figure 11 -Graphical representation of observed discharge, radar and gauge generated

discharge at Ellerbe creek watershed outlet, (a) Event 1, (b) Event 2, (c) Event 3, (d) Event 4 49

Figure 12 - Graphical representation of observed, and radar generated discharge at River Des Peres watershed outlet, (a) Event 1, (b) Event 2, (c) Event 3, (d) Event 4 52

CHAPTER 1

INTRODUCTION

1.1 RESEARCH BACKGROUND

Flooding events are among the most disastrous and costly calamities affecting a large number of populations worldwide. Floods are considered to have a significant global impact compared to other natural disasters such as drought, heatwave, and wildfires, as they can cause extensive damage and destruction of lives and property of the worldwide population. Nearly 30% of the total economic damages from all-natural catastrophes worldwide are attributed to flood destruction (Abbott et al. 1986). There has been a widespread acknowledgment of the effects of climate change and urbanization(Kalra et al. 2008, 2013b; Kalra and Ahmad 2011; Pathak et al. 2017; Sagarika et al. 2014), which resulted in a significant increase in the frequency and severity of urban floods in many parts of the world (Thakali et al. 2016; Thakur et al. 2020b; Acharya et al. 2020).

Climate change is anticipated to substantially impact the timing, duration, and intensity of rainfall events, which would significantly intensify flooding events in many urban cities worldwide and increase future flood risk and related losses in the absence of effective mitigation(Thakur et al. 2020c; b). Similarly, urbanization has transformed the pervious natural topography to the more impervious surface due to infrastructural development and anthropogenic activities. The increase in impervious surfaces decreases the abstraction capacity of soil by which all rainfalls are converted into runoff, creating significant flooding events.

Governments, researchers, and engineers are focusing on analyzing the influence of climate change and urbanization on flood risk to prepare cities, apply viable adaptation strategies, and make informed choices about mitigation techniques(Pathak et al. 2018). Planning and implementing flood management techniques in vulnerable locations is only possible with accurate rainfall-runoff modeling(Shrestha et al. 2020a). In this context, various process-based models have been developed till data to accurately analyze rainfall-runoff simulations during high-intensity rainfall events. Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS) and Personal Computer Storm Water Management (PCSWMM) are widely used to analyze the rainfall-runoff of rural and urban watersheds. However, calibrating and verifying a physical-based model like HEC-HMS and PCSWMM requires a substantial amount of topographic and meteorological data, which is difficult to obtain in data scarce regions. Additionally, there are disadvantages to utilizing a physical-based hydrological model because it is challenging to comprehend the intricate, nonlinear, and interconnected hydrology.

In this context, different Machine learning models has been developed till date which has shown good performance in hydrology domain. Random Forest, developed by (Breiman 2001) is a Machine learning model which has shown good performance in hydrology domains such as flood mapping and risk analysis. However, Random Forest is rarely used in the analysis of Rainfall-Runoff simulation. Therefore, this study evaluates the performance of process based model and Random Forest Model to generate the runoff hydrograph during different events of rainfall.

Rainfall data, a meteorological data, is the most important datasets for the hydrologic analysis of the watershed. In hydrology, accurate geographical and temporal variation of rainfall data are essential for generating accurate flood discharge. Rainfall is typically measured and estimated using rain gauging stations. However, in most watersheds, the gauging station is not always available. In addition, it is challenging to determine the spatial variability of rainfall at the watershed scale through a rainfall gauging station. Therefore, this study applies satellite and radar precipitation products to accurately predict the flooding discharge during high intensity flooding events.

1.2 PROBLEM STATETMENT AND OBJECTIVES

Flooding events are among the most disastrous and costly calamities affecting many populations worldwide. Climate change has synergistically interacted with urbanization to produce extreme flooding and drought events, which has significantly impacted many cities of the world(Joshi et al. 2020). Hydrological analysis is the foremost stage in studying the impact of such extreme flooding events in urban areas and applying possible flood mitigation strategies(Bhandari et al. 2018). Different process-based models, such as HEC-HMS and PCSWMM, have been developed to replicate urban hydrological processes accurately. However, the process-based models require a large number of data sets which may not always be available in the data-scarce region. In this context, machine learning tools have served as an alternative tool to analyze rainfall-runoff simulation in a data-scarce region. Furthermore, rainfall datasets, crucial data for hydrological analysis, are not always available in the watershed. In such conditions, radar and satellite-based precipitation datasets can be applied for hydrological analysis.

In this context the primary objective of this study is to determine the effectiveness of process-based models and random forest models for the hydrological analysis in a data scarce region by the satellite precipitation product. In the first part of this research, the effectiveness of Random Forest is compared with the HEC-HMS model using Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks-Cloud Classification System (PERSIANN-CSS) precipitation, a satellite precipitation product. In addition, this study also determined the appropriateness of applying Random Forest generated discharge for hydraulic modeling using Hydrologic Analysis Center's River Analysis Model (HEC-RAS).

The second part of this study integrates the PCSWMM with NOAA Next Generation Radar (NEXRAD-III) precipitation product for the hydrological analysis of the urban watershed. In addition, this study determines the appropriateness of NEXRAD-III precipitation product for hydrological analysis by comparing it with the discharge hydrograph generated from gauging station precipitation data.

Research Question #1: Is the Random Forest Model application feasible for the rainfall-runoff simulations in the data-scarce urban watershed?

Hypothesis #1: The random forest model is anticipated to perform as an alternative model for hydrological analysis.

Research Question #2: What is the effectiveness of satellite and radar rainfall products for the hydrological investigation in an urban watershed?

Hypothesis #2: Satellite and radar precipitation products can offer spatial and temporal variations of precipitation data at a watershed scale.

1.3 RESEARCH OUTLINE

The project follows a manuscript structure beginning with an introduction. Two manuscripts are combined in this study for the completion of this research. The second chapter, titled "*Application of Machine Learning and Process-based models for rainfall-runoff simulation in Du Page River basin, Illinois,*" addresses research question #1. Furthermore, the third chapter, titled " *Evaluating the performance of PCSWMM using NEXRAD data for urbanized watersheds* " and the second chapter addresses research question #2. Chapter four outlines this project's outcomes and suggests future research recommendations.

CHAPTER 2

APPLICATION OF MACHINE LEARNING AND PROCESS-BASED MODELS FOR RAINFALL-RUNOFF SIMULATION IN DUPAGE RIVER BASIN, ILLINOIS

2.1 INTRODUCTION

Floods are one of the most common and costly natural catastrophes all around the world(Gaume et al. 2009; Merwade et al. 2008; Merz et al. 2010). The magnitude and frequency of extreme flooding events have increased considerably worldwide over the previous few decades (Ghazali et al. 2018). Climate change, urbanization, and other anthropogenic activities are causing a flood risk globally (Faccini et al. 2018; Joshi et al. 2021). A water-related natural hazard such as floods, drought, and a landslide has become the new normal due to the uncertainty in rainfall patterns and magnitude caused by climate change and urbanization (Parajuli et al. 2017; Pathak et al. 2017; Shrestha et al. 2020b). Flooding is projected to become more common in the coming years as the frequency of extreme precipitation events increases (Guerreiro et al. 2018; Jenkins et al. 2017).

Flood severity increases, resulting in high flood fatalities, massive economic loss, and social consequences (Min et al. 2011). Given the negative consequences of flooding, developing floodplain management plans to avoid and mitigate flood damage is critical (Vörösmarty et al. 2013). The flood risk assessment depends on a precise estimation of peak runoff, calculated by rainfall-runoff simulation (Woznicki et al. 2019a). Accurate rainfall-runoff simulation is a prominent topic in hydrology research (Archer and Fowler 2018). 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 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 (Kastridis and Stathis 2017).

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 like HEC-HMS can compute an actual hydrology system (Schoppa et al. 2020). The hydrology modeling using the HEC-HMS model can investigate urban floods, flood frequency, flood warning system, and effectiveness of spillways and detention ponds over a watershed (Talei et al. 2010). 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 is entered, edited, managed, and stored via a system. Fourth, the simulation results are reported and illustrated using a functional system (Singh and Frevert 2010). Finally, the calibration procedure, which compares simulated outcomes 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 (Halwatura and Najim 2013). A physical-based model such as HEC-HMS necessitates a large amount of data such as land use and land cover, soil group data, infiltration data, and a significant amount of time to calibrate to ensure the correctness of the model (Scharffenberg 2016). Furthermore, there are drawbacks to using a physical-based hydrology model, owing to the difficulties in completely understanding the complicated, nonlinear, and interrelated hydrology (Senthil Kumar et al. 2005). The hydrology model using HEC-HMS can be unsuitable for a larger watershed with scarce data. Therefore, as a complement to the physical model, recently, the application of Machine learning and the data-driven model has been used across hydrology domains (Kim et al. 2015; Rezaeianzadeh et al. 2013).

Machine learning is a kind of artificial intelligence that can make an accurate prediction by training and testing datasets. Machine learning provides a solution to a real-world problem by studying previously observed data and has been effective in generating accurate results (Sahoo et al. 2017). ML provides adequate computation power (Rajaee et al. 2020; Zounemat-Kermani et al. 2021) and is used in a wide variety of research and applications in hydrology(Ahmad et al. 2010; Thakur et al. 2020a). Some examples of ML applications in the hydrology domain are rainfall-runoff prediction (Jordan and Mitchell 2015; Mewes et al. 2020), flood forecasting (Adnan et al. 2021; Bhandari et al. 2019; Kalra et al. 2013b; Kalra and Ahmad 2007; Parisouj et al. 2020), sedimentation study (Nguyen and Chen 2020; Shamshirband et al. 2020; Zhou et al. 2022b), water quality prediction (Carrier et al. 2013; Choubin et al. 2018; Kalra et al. 2013a, 2018; Rezaei et al. 2021; Rezaei and Vadiati 2020) , groundwater prediction (Deng et al. 2021; Rahaman et al. 2019; Wang et al. 2022), river temperature prediction (Asadollah et al. 2021; Hussein et al. 2020; Khedri et al. 2020; Melesse et al. 2020), and rainfall estimation (Chang and Psaris 2013; Zhu and Piotrowski 2020). In recent years, ML algorithms have significantly improved and are also widely used for rainfall-runoff simulation (Feigl et al. 2021; Weierbach et al. 2022), thanks to the rapid advancement of computer technology. Recently, many researchers performed rainfall-runoff predictions using different Machine Learning and Data-driven models. Some examples of these models are long short-term memory (Radhakrishnan et al. 2022; Zhang et al. 2022), artificial neural networks (Chiang et al. 2022; Guo et al. 2021), support vector machines (Ni et al. 2020; Yin et al. 2022), and the random forest model (Tikhamarine et al.

2020; Woznicki et al. 2019a). Random Forest is a popular machine learning tool, and Breiman developed it first in 2001 (Tamiru and Dinka 2021). The 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 (Samantaray et al. 2022). It is a classification and regression tree-based ensemble learning algorithm (Tamiru and Dinka 2021). 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 (Adnan et al. 2020).

Previous studies showed that the 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 (Adnan et al. 2020; Breiman 2001; Meng et al. 2021; Worland et al. 2018; Zhou et al. 2019). However, random forest is the least used for hydrology analysis among the data-driven and machine learning models (Li et al. 2016). Among the few applications of Random Forest, most of these studies focused on flood risk hazards (Bachmair et al. 2017; Woznicki et al. 2019b) and mapping (Erdal and Karakurt 2013). Therefore, in this study, an assessment of Random Forest for rainfall-runoff prediction in a watershed is performed. The main aim of this research is to determine the effectiveness of the Random Forest model for rainfall-runoff operation in the scarce data region. Therefore, this research also used satellite precipitation product for rainfall-runoff simulation and determined its appropriateness in hydrology research. Furthermore, this study also assessed the appropriateness of using Random Forest generated discharge for hydraulic modeling using (Hydrologic Analysis Center's River Analysis Model) HEC-RAS.

HEC-RAS is the most widely accepted model (Muñoz et al. 2018), for analyzing channel flow and floodplain characterization (Tyralis et al. 2019). Users can compute one-dimensional steady and unsteady flow, two-dimensional unsteady flow, sediment transport computations, and water quality models by using HEC-RAS (Tyralis et al. 2019). Regularizing geometric data and identifying hydraulic structures (weirs, culverts, reservoirs, pump stations, bridges, levees, gates, blockage and ineffective regions, land use, Manning roughness coefficient, streambed slope, and ice cover) and their situations are achievable with HEC-RAS (Wang et al. 2015). 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 is employed as an input parameter in this study. While 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 (Feng et al. 2015). 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 for the 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 by comparing the effectiveness of the random model with a physical-based model, section 4 provides the discussion of the results, and section 5 provides the major conclusions from the current analysis.

2.2 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 basin characteristics data such as Digital Elevation Model (DEM), Land Use and Land Cover (LULC), Soil group, and meteorological data such as daily precipitation and discharge data. The integrated use of Arc-Hydro, HEC-GeoHMS, and HEC-HMS was performed 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, the 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 approach of Random Forest and HEC-RAS model for hydraulic analysis.

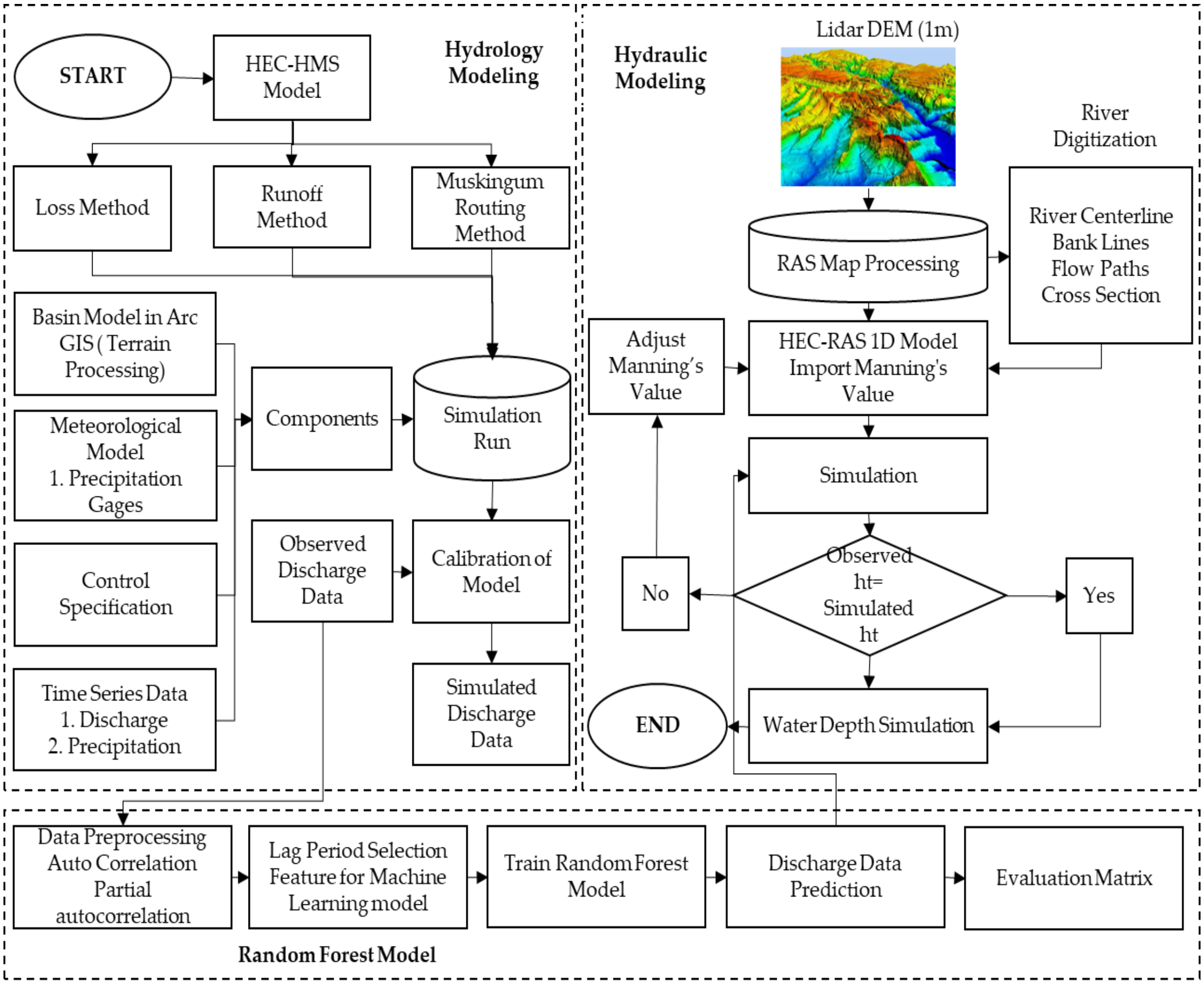


Figure 1 - Figure portraying the flowchart of hydrology simulation using Random Forest and HEC-HMS and hydraulic analysis using HEC-RAS.

2.2.1 STUDY AREA

This research used the East Branch DU PAGE watershed as a study area. Over the last twenty years, the study area has observed significant urbanization. The study area had a history of high flooding events in 1996, 2008, 2013, and most lately in 2020. In the year 2020, there was significant flooding due to 178mm of total precipitation over a period of five days. The study watershed has an area of 62.2 km^2 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′ bounds the study catchment area, as shown in Figure 2. The study area is highly residential, with an average impervious percentage of about 40%, the range of percentage imperviousness in a watershed is shown in Figure 2. The average soil permeability over the watershed is 62mm/hour. 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 to gauging station 05540228. The study reach is around 5221 m between two gauging stations. The proposed research area does not consist of any 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.

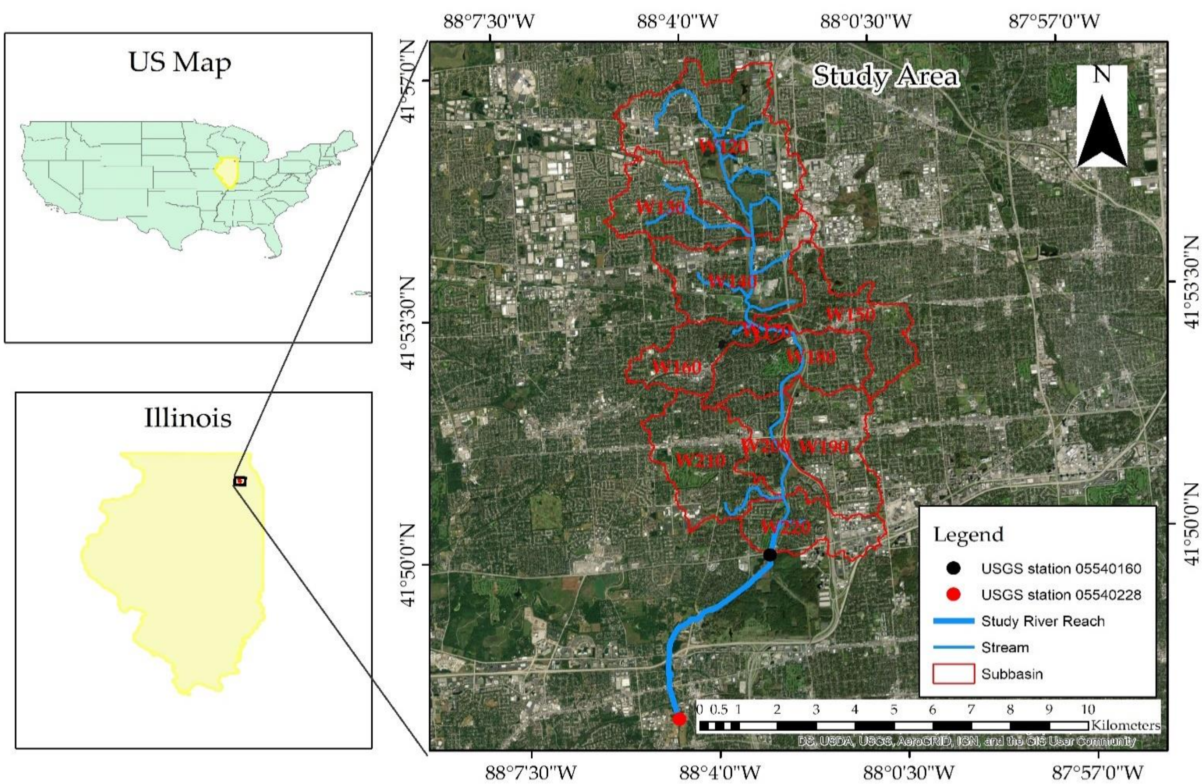


Figure 2 - The East Branch DuPage Catchment around Downers Grove, Illinois, with the river system.

2.2.2 DATA

Watershed characteristics datasets such as land-use and land cover, soil group, DEM, and meteorological model data such as rainfall and 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 type used in this research and their sources are detailed in Table 1.

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

2.2.3 PREPROCESSING DATA

This section describes the extraction of basin characteristics of the study catchment.

2.2.3.1 DIGITAL ELEVATION MODEL

DEM is spatial data that provides the characteristics of the watershed. 10 m * 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.

2.2.3.2 BASIN CHARACTERISTICS

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

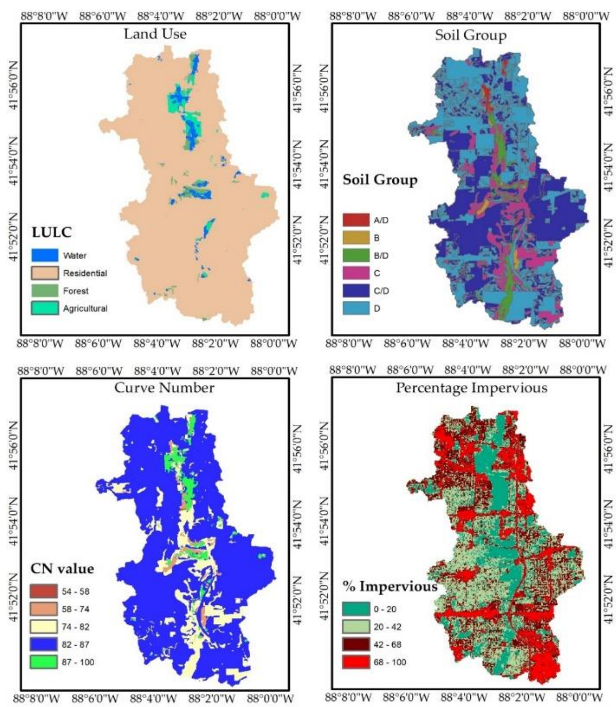


Figure 3 - Map depicting characteristics of study area.

2.2.3.3 PRECIPITATION DATA

Rainfall is essential meteorological data for the hydrology simulation. The study area does not consist of any observed precipitation station; therefore, in this study, precipitation data were obtained from a grid from PERSIANN-CCS (Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks-Cloud Classification System). 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° x 0.04° pixel) satellite precipitation product (Nguyen et al. 2019). The daily time series precipitation data was extracted from a grid using a python environment from 2006 to 2021.

2.2.4 HYDROLOGIC MODELLING USING ARC-GIS AND HEC-HMS

HEC-GeoHMS is an extension of an 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 is carried out as shown in Figure 4.

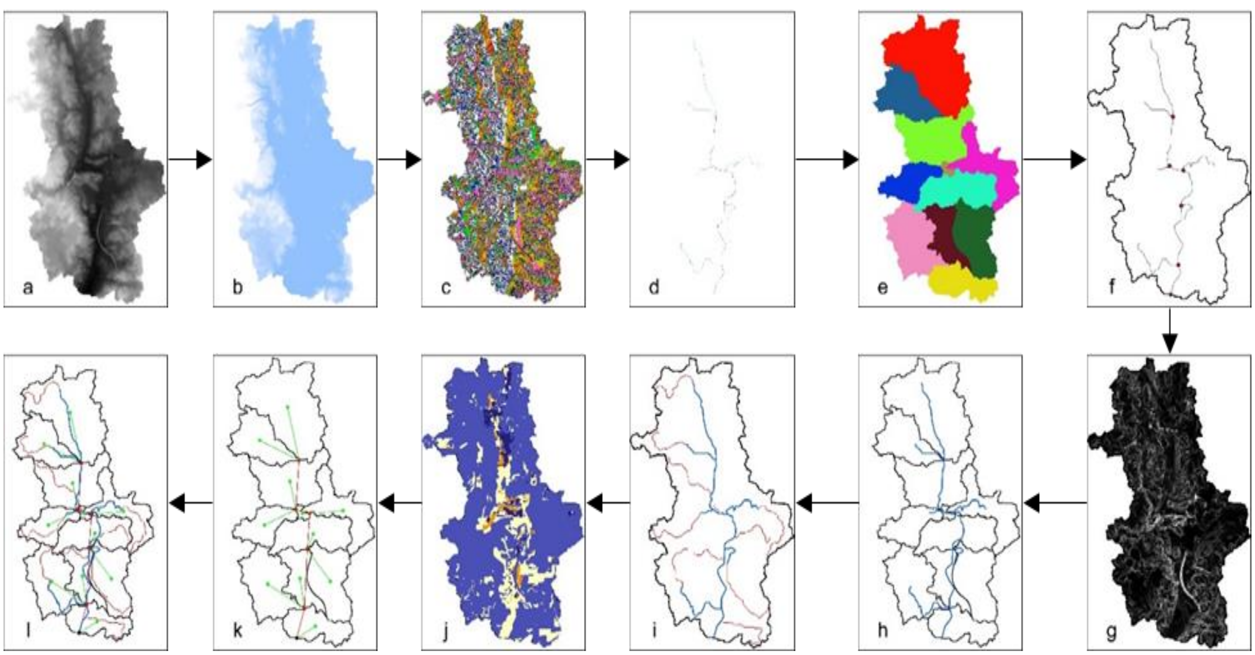


Figure 4 - Pre-processing and model development: (a) DEM file; (b)Fill Sinks; (c)Flow Accumulation; (d)Flow Direction; (e)Stream definition and Catchment Polygon; (f)Drainage Point and line processing; (g) Slope; (h)Basin and River merge; (i)Lonest Flow path; (j)CN lag; (k)Sub-basin nodes and river links; (l)HEC-HMS input file.

2.2.4.1 LOSS METHOD

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

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

Where, Q= Runoff (inch); P= Rainfall depth (inch); I_a= Initial abstraction, and I_a=0.2S; S = Potential maximum retention. The potential maximum retention in inches, S, is calculated using Equation 2:

$$S = \frac{1000}{CN} - 10 \tag{2}$$

2.2.4.2 TRANSFORM METHOD

The SCS Unit Hydrograph transforms excess precipitation into runoff. 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**,

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

Where, T_{lag} = lag time (hrs.); L = hydraulic length of the watershed (ft); Y = slope of the watershed (%); S = maximum retention in the watershed (inches).

2.2.4.3. ROUTING METHOD

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 time passing in a reach length. These parameters can be approximated using observed inflow and outflow Hydrographs, and the X parameter is a weight coefficient of discharge, which value fluctuates between 0 and 0.5. The interval between the inflow and outflow Hydrographs of an identical station can be used to determine parameter K. In this model, routing methods parameters were used to calibrate the model.

2.2.5. 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 like the Random Forest technique. Despite the popularity of the Random Forest algorithm in myriad environmental sciences fields, its application in the water sector has still remained to be explored more (Saadi et al. 2019). 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 (Breiman 2001). Random Forest is a collection of decision trees, where each tree slightly varies from one other. 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. The Random forest has a good performance in the large dataset, and its features need not to be scaled (Park et al. 2019). It is advantageous for the features with different scales. Random forests are appealing for both classification and regression tasks, computationally fast, efficient for unstable prediction, and perform well with high-dimensional features (Biau and Scornet 2016; Gregorutti et al. 2017). This algorithm's key idea is that each tree might make a fair prediction on its part; however, overfitting seems to occur on a certain part of 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.

2.2.5.1 MODEL DEVELOPMENT

Many decision trees with bootstrap aggregation are used to minimize the overfitting issue (Hussain and Khan 2020). 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. Max depth of the model is fixed to be 100. Max depth by default is 'None' which signifies that the nodes are enlarged until all the leaves have fewer than min_samples_splits samples. Min_samples_split means the total samples needed to break internal node. Since we are trying to maintain the number of decision trees at only 100, the max features are stated auto, which means max features is equal to n features (The number of features seen during the model fitting). The parameter max-leaf nodes = None, refers to the unlimited number of leaf nodes, leaving the decision trees to grow best to fit the model. Total number of daily hydrology and meteorological feature samples from 2006 to 2021 were used for training and testing the algorithm. 80% of total datasets were used for the training, and 20% of entire datasets were used for the testing of the Random Forest model.

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 and Figure 5b shows the Autocorrelation and the partial Autocorrelation plot of historical daily runoff observations respectively. This plot helped identifying a suitable lag period for flow prediction in a watershed (Hussain and Khan 2020). 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 lag time were selected.

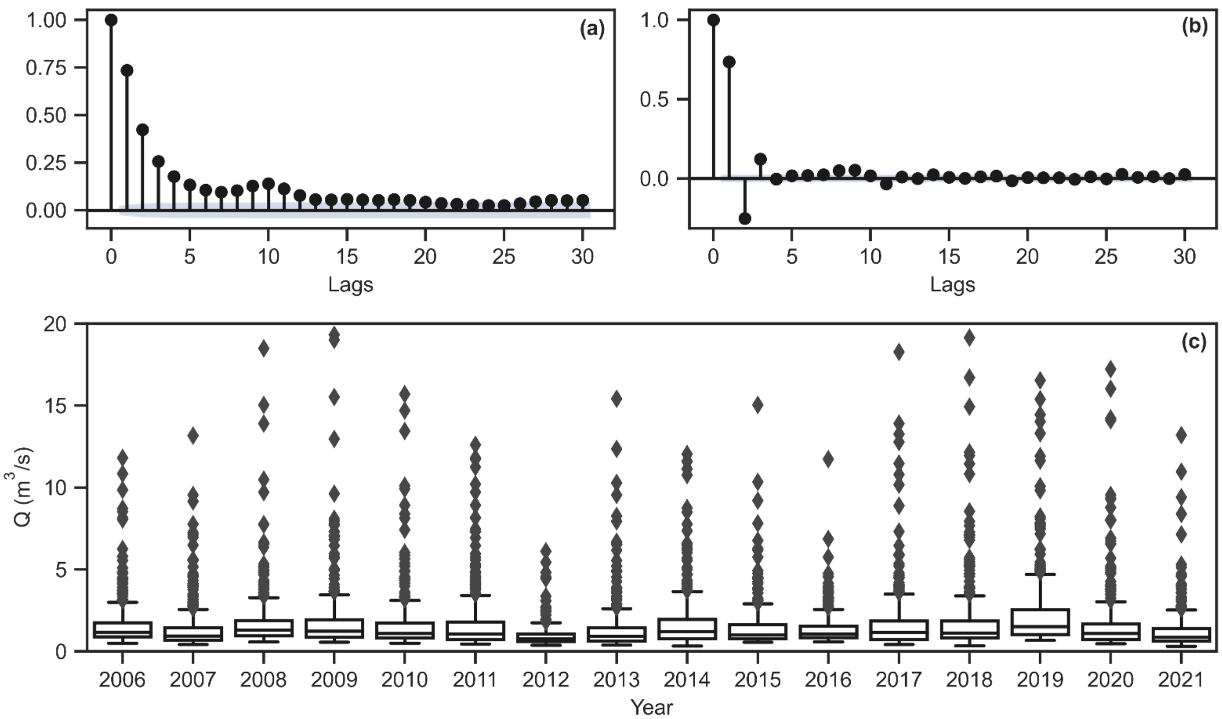


Figure 5 - a) Autocorrelation plot of the historical runoff observation of the DuPage River; b) Partial autocorrelation plot of the historical runoff observation of the DuPage River; c) Box Plot showing the flood events of the DuPage River.

Figure 5a represents the combination of input features for training the random forest regression. In addition, the cumulative precipitation for 5 days and the day in which rainfall was greater than 0.5 inches was considered as an additional feature for predicting the runoff at the outlet of the watershed. NumPy, Pandas, Matplotlib, stats model, Sklearn, and seaborn are the python libraries that area used during data processing, training, and visualizing.

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. 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.

Table 2 - The combination of input for runoff prediction using random forest regression

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

2.2.6 HYDRAULIC MODELLING

Hydraulic modeling using HEC-RAS uses adequate geometry and flow data input for an excellent hydraulic model. The 1D HEC-RAS model is commonly employed to analyze flow in mainstream channels and predict flood extent. Although the 1D model has limited applications, it is cost-effective, durable, and favored when determining flow pathways (Gharbi et al. 2016). When speed is of the essence and flood plain geometry data is scarce, 1D modeling is chosen (Pathan and Agnihotri 2021). HEC-RAS calculates the energy expression using Equation 4, which is based on Saint Venant's equation.

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

Where, Y_1 and Y_2 = water heights at cross-sections; Z_1 and Z_2 =elevations of 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.

2.2.6.1 RIVER GEOMETRY

Hydraulic Analysis with the 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 a flowchart in Figure 1. The Lidar 1m DEM for the hydraulic model was obtained from the USGS website. The DEM data was imported into the RAS Mapper tool in the HEC-RAS and was converted into 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 at entire reach. After the creation of river geometry and applying manning's n value, the steady discharge was used as an 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. Manning's n values at the main channel and over banks were adjusted for calibration of a model.

2.2.7 STATISTICAL PERFORMANCE INDICATOR

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 (Moriasi et al. 2015) and NRMSE were used in this research to assess the performance of the hydrology model. The criteria used to evaluate the proposed models' performance is listed in Table 3**.**

Table 3 - List of statistical indexes for determining the performance of models.

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

Where, $Q_{o,i}$ represents the observed data, $Q_{s,i}$ represents the simulated data from the model, $\overline{Q}_{o,i}$, represents the mean value of total number of data's, and n represent the total number of data points.

2.3 RESULTS

This section describes the results of the study and has four main topics. In this section, results of precipitation product, hydrology analysis and hydraulic analysis is presented.

2.3.1 PRECIPITATION

The rainfall data applied in this research was extracted from satellite-based rainfall products for a time period of 16 years (2006-2021). The daily rainfall data obtained for a study time period is shown in Figure 6a. The daily precipitation data pattern is consistent with the daily observed discharge data. The result shows that the time of peak rainfall data matched the peak discharge data. For example, in this watershed outlet, the highest peak discharge of 33.7 m^3/s was observed on Sept 14, 2008, and similarly, extracted precipitation product produced the highest precipitation of 61 mm on the same day. In addition, the validation of extracted precipitation data was supported by the results of the hydrology analysis, which is presented in the following section.

2.3.2 HEC-HMS MODEL

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

The calibration and validation of HEC-HMS model in this research was done by adjusting the Muskingum parameters. The measured discharge from the gauging station was compared to the yearly peak discharge produced from a simulation. Event 2006 Jan 1 to 2018 Dec 31 was considered for the model calibration, and Event 2019 Jan 1 to 2019 Dec 31 was used for the model validation. The accuracy of hydrology model using HEC-HMS was proved using statistical index. The discharge generated using HEC-HMS for the study period is presented in Figure 6b. A root means square error is one of the most used methods for evaluating the validity of prediction. The results of RMSE during calibration and validation were 1.45 m^3/s and 2.45 m^3/s, respectively, which is considered as a good result. RSR is calculated by dividing the RMSE by the standard deviation of the measured data, and with values less than 0.7, is considered as a good result (Kumar et al. 2017). The RSR values for the HEC-HMS model were 0.16 and 0.35. The NSE is extensively used in performance measures in hydrology. It ranges from – 1 to 1, with 0.5 to 1 being the best value. The NSE method is used to calculate the residual variance in relation to the variance of measured data, and the NSE values were 0.97 and 0.87, respectively, which are near one.

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

PBIAS shows the average inclination of the calculated data. For a good model, PBIAS values must approach zero or be less than 25% (Abbaspour et al. 2015). Positive numbers suggest that the model is underestimated, whereas negative values indicate overestimating the model (Gupta et al. 1999). Our model overestimated the peak value by +5.3 percent for calibration and +9.8 percent for validation. R^2 is used to determine a correlation between calculated and measured flow rates. R^2 greater than 0.5 indicates satisfactory performance. For calibration and validation, R^2 values were 0.99 and 0.96, respectively. R^2 value close to 1 for the HEC-HMS model validates the accuracy of the model.

2.3.3 RANDOM FOREST REGRESSION

Random Forest regression provided good insights into the prediction of daily discharge data. Figure 6c provides a good representation of the observed value and the predicted value while using Random Forest. The scatter plot in Figure 6d showed that Random Forest prediction was clustered near the regression line on the low and normal flow conditions. However, the Random Forest regression slightly overestimated high discharge value, which can also be termed an extreme event. **Table 5** shows the evaluation matrix for the random forest regression. The value of RMSE, RSR, NSE, PBIAS, R^2, NRMSE were 0.29 m^3/s, 0.23, 0.94, -0.75%, 0.94, 0.17 for the training period and 0.47 m3/s, 0.56, 0.69, +1.76%, 0.72, 0.260 for testing period, respectively as shown in **Table 5**. Statistical Index revealed that the Random Forest model performance was superior during data training. The values of the statistical index dropped sharply during the testing period. PBIAS values for training and testing values are near to 0%, representing the average inclination of predicted discharge towards the observed discharge. The values of R^2 dropped sharply from 0.94 during training to 0.72 during the data testing. However, the values of a statistical index are within acceptable ranges during a testing period. Scatter plots analyze the prediction performance of random forest regression with the observed data. Larger deviation on the line was seen on higher flow values. It means the model has prediction error on the prediction of the peak flows. The non-peak discharge was more accurately predicted by the machine learning model.

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

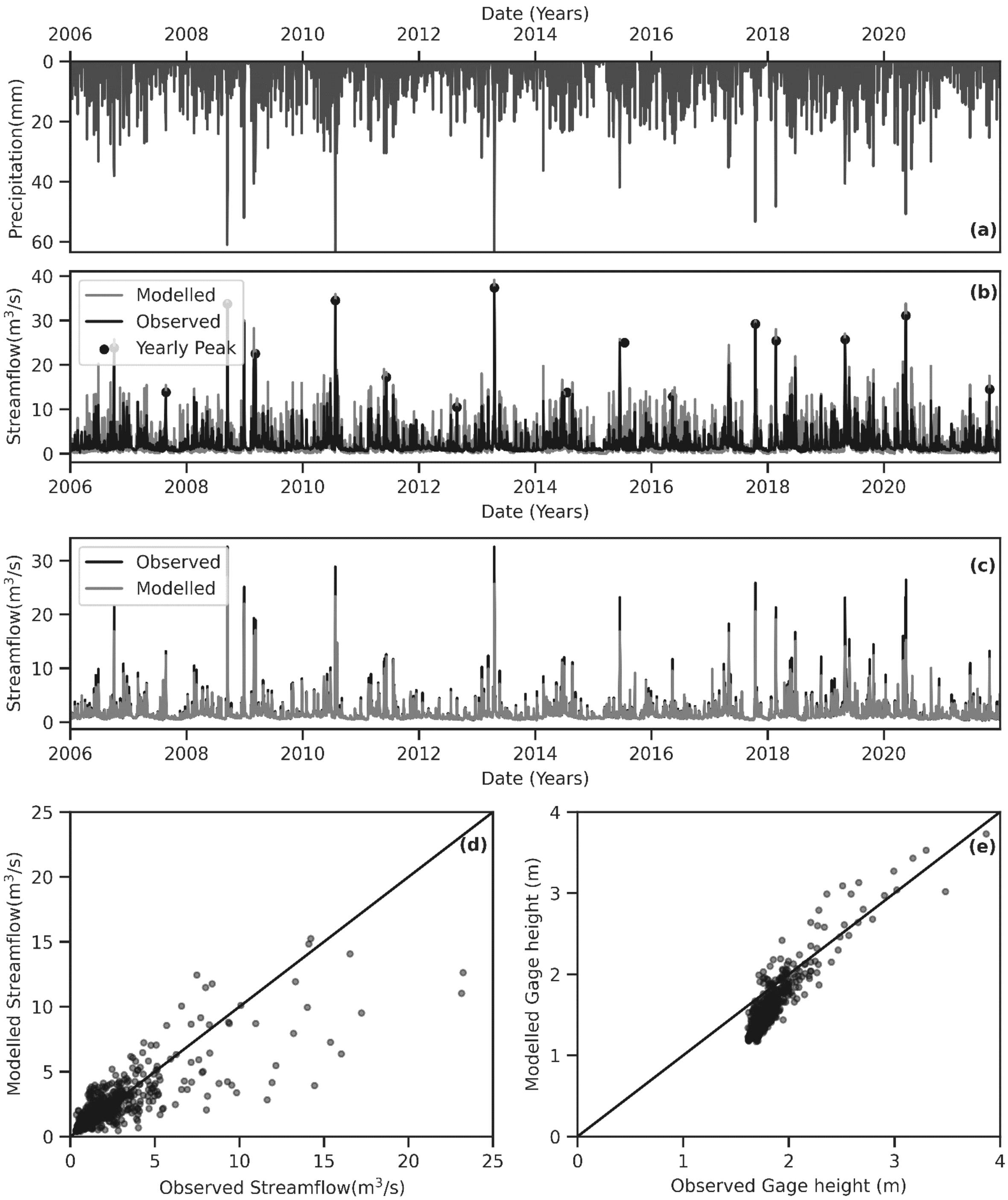


Figure 6 - a) Representation of generated precipitation product; b) Training and testing for HEC-HMS Model; c) observed discharge and predicted discharge for random forest regression; d) Observed historical and Predicted runoff data, e) Observed gage height and Predicted gage height

Table 5 - Calibration and Validation summary statistics of HEC-HMS and Random Forest model.

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

2.3.4 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 the Figure 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 showed the comparison of simulated and observed data at most downstream stations of a study reach. The manning's n value was adjusted to calibrate the hydraulic model. The water depth produced from a simulation was similar to the observed water depth at the gauging station, as shown in Figure 6e; this demonstrates the model's consistency and allows it to be used for further investigation. At the upstream cross-section of reach, daily discharge data from Random Forest was 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 to the observed water depth at the gauging station, the model underestimates the simulated water depth generated from the study.

Table 6 - Observed and simulated water depth.

Event	Discharge (m^3/s)	Observed Water depth (m)	Simulated Water depth (m)	Difference (m)
2020 Jan 11	8.78	2.79	2.68	0.11
2020 Mar 30	3.11	2.09	1.98	0.11
2020 Mar 29	5.07	2.33	2.58	-0.25
2020 Apr 30	16.03	3.40	3.02	0.38
2020 May 18	26.42	4.41	3.85	0.56
2020 Oct 23	6.57	2.45	2.91	-0.46
2020 Dec 12	8.04	2.52	2.61	-0.09
2021 Mar 19	2.83	1.96	2.03	-0.07
2021 Jun 26	10.96	2.90	3.27	-0.37
2021 Aug 27	2.21	1.81	1.89	-0.08
2021 Oct 26	8.38	2.50	2.48	0.02

2.4 DISCUSSION

The outcome of the hydrology simulation demonstrated strong support for the effectiveness of the satellite precipitation product for the hydrology simulation in an ungauged catchment. Both HEC-HMS and Random Forest models accurately recreated the discharge characteristics such as flood peak and timing during the study period. These findings are consistent with previous studies, which showed that PERSIANN-CCS precipitation products could effectively simulate hydrology in ungauged watersheds (Hong et al. 2007; Joshi et al.

2019a). The statistical index, listed in Table 5 from model calibration and validation, suggested that Random Forest can be effectively applied for estimating 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 result of Random Forest agrees with the previous study's findings of the good performance as an alternative predicting method in the hydrology domain (Desai and Ouarda 2021). The statistical index in Table 5 proved the suitability of both Random Forest and HEC-HMS for rainfall-runoff simulation. Results illustrated 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 the HEC-HMS model. **Figure 6e** provided good support for the effectiveness of Random Forest generated discharge for hydraulic simulation. The result depicted that the simulated water depth by HEC-RAS at the most downstream cross-section was slightly underestimated compared to the observed water depth at the gauging station. This result may be because of using the slightly underestimated 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 the scarce data region.

2.5 CONCLUSION

This study evaluates the feasibility of HEC-HMS and Random Forest for rainfall-runoff simulation and an integrated approach of the Machine Learning and HEC-RAS for hydraulic analysis. HEC-HMS requires large number of input variables, which may not be always available in a data scarce region, in this scenario, Random Forest model can be used for the prediction of discharge in the watershed. In addition, Random Forest model is simple to build and takes less time. In this study, PERSIANN-CCS NetCDF file was used to generate time series precipitation data. The result shows good support on the usage of PERSIANN-CCS daily precipitation data for the rainfall-runoff simulation. Based on the models' reasonably strong performance, precipitation data, LULC, DEM, and SSURGO soil input data obtained are sufficiently dependable to simulate discharge. Because the data sources employed in this study yield reasonably reliable results, they are recommended for hydrology investigations. The basin's continuous simulation of rainfall-runoff processes using physical and machine learning has yielded good results. Peak flows were under-predicted in the Random Forest and slightly over predicted in the HEC-HMS model. An Integrated Model of HEC-RAS and Random Forest regression showed a good result in predicting runoff flood depth at 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 for further research. The model's accuracy for predicting the flow can be increased by removing the outliers; high flood values are considered here to compensate for the prediction of the high flood values from the random forest regression. Future Researchers can work in the following areas.

In this study, we used the PERSIANN precipitation product, and the future work may be more accurate if there is any gauging precipitation station. Furthermore, Researchers can also use other precipitation products such as Next Generation Weather Data (NEXRAD) and Climate Hazards Group Infrared Precipitation (CHIRPS).

In this study, precipitation data is only used as an input variable for the Random Forest model; other variables such as temperature, infiltration, evaporation, and radiation can be used in future work. In addition, feature selection of input variables can be made for the most accurate selection.

The other machine learning and data-driven models, such as support vector regression (SVR), long short-term memory (LSTM), and artificial neural network (ANN), can be used as the other prediction model. The future research direction can be guided for the best selection from the multiple machine learning models based on accuracy, robustness, and reliability. Although the study area is a small watershed in DuPage County, future research could focus on a more dynamic, heterogeneous, and meteorologically unique basin.

Chapter 2 is published in Hydrology, MDPI. Bhusal. A., U. Parajuli, S. Regmi, and A. Kalra. 2022. "Application of Machine Learning and Process-Based Models for Rainfall-Runoff Simulation in DuPage River Basin, Illinois." Hydrology, 9 (7): 117. https://doi.org/10.3390/hydrology9070117.

CHAPTER 3

EVALUATING THE PERFORMANCE OF PCSWMM USING NEXRAD DATA FOR URBANIZED WATERSHEDS

3.1 INTRODUCTION

Floods are among the most destructive events worldwide, and its frequency and magnitude have been increasing significantly in the past few decades (Aryal et al. 2022; Chen et al. 2018; Merz et al. 2010; Tamaddun et al. 2016; Thakur et al. 2017). The worldwide flood casualties rose from an average of $7 billion USD every year in the 1980s to roughly $24 billion USD each year in the 2000s (Kundzewicz et al. 2014), and this statistic is expected to rise to $52 billion USD by 2050 (Hallegatte et al. 2013). Urban floods can damage urban infrastructure, interrupt city services, and have significant adverse socioeconomic effects (Zhou et al. 2022a). Additionally, it is predicted that by 2030, there will be 5 billion urban inhabitants worldwide, and a single flood might significantly impact millions of people's lives (Gaines 2016). In this context, accurate peak discharge and discharge hydrograph predictions during catastrophic flooding events are essential for flood control decision-making(Kalra et al. 2021b; a). The threat of high-intensity rainfall events can be reduced by deploying flood control systems in flood-prone areas with correct evaluation of the rainfall-runoff simulation through hydrological modelling(Bhandari et al. 2017; Ghimire et al. 2016; Nyaupane et al. 2018; Pokhrel et al. 2020; Thakali et al. 2017).

Rainfall data is the most critical input parameter for development of hydrologic models which significantly impacts the model's accuracy (Price et al. 2014). The accuracy of time-variable rainfall data is paramount for successfully verifying the hydrologic models (Pechlivanidis et al. 2016). In addition, accurate information on rainfall's spatial and temporal variability is essential in forecasting extreme climatology conditions, hydrological simulation, flood and drought monitoring, ground water monitoring and water resource management (Cristiano et al. 2017; Joshi and Kalra 2021). Insufficient and inappropriate rainfall data causes verification errors in the hydrologic models, which lowers the model's accuracy and dependability for simulating the actual watershed (Vallabhaneni et al. 2004) .

Point gauge observations have historically been the primary source of the necessary rainfall data for hydrologic models. Point rain gauges detect rainfall at a specific location of the rain gauge. However, hydrological models require information on the area-averaged rainfall for their accuracy. Therefore, rainfall estimations from gauges become insufficient due to the poor depiction of areal precipitation by rainfall gauging stations, especially in scenarios with sparse gauge network designs (Porcù et al. 2014). Researchers have developed a number of approximation strategies that estimate the areal average rainfall data based on mathematical and geometrical calculations. Thiessen polygon, kriging, and inverse distance weighting methods are some common approaches that develop the areal average rainfall data by interpolating rainfall data from gauging stations. However, the results obtained from such interpolation may not always be correct due to rainfall variability within a small area, as well as sparse or nonexistent gauging stations in the watershed(Haberlandt 2007). This scenario makes it difficult for system managers and researchers to accurately derive the hydrological response of watershed during extreme rainfall period and application of flood protection system, causing a significant socioeconomic impact.

In this context, radar-based rainfall data has been developed in recent years as an alternative to the gauging station’s rainfall data that provides information on average areal rainfall. Radar rainfall data resolves the natural variability of rainfall over temporal and spatial scales that are pertinent to hydrologic applications which is why it is represented as a major advance over gauging rainfall data despite its indirect method of detecting rainfall (Hamedi and Fuentes 2015). Many previous researchers have used radar rainfall data in their watershed to determine its effectiveness for the development of outlet hydrograph. Some studies concentrate on accurately simulating total precipitation volumes, while others focus on the impact of rainfall's spatial and temporal variability on the accuracy of the hydrographs(Chaubey et al. 1999; Krajewski et al. 1991). The results are inconsistent, mixed, and sometimes contradictory based on the numerous studies conducted using radar data. Some studies found that radar-based rainfall data can produce significant errors in developing flood hydrographs compared to the rainfall data obtained from gauging stations (Johnson et al. 1999; Cole and Moore 2008). However, several researchers contrasted this by stating that radar data can predict floods as correctly as rain gauge data (Bedient et al. 2000; Lopez et al. 2005; Pessoa et al. 1993). The past literature suggests that although hydrological models have been used in numerous research studies to analyze the geographical and point differences between rain gauge and radar rainfall data, definitive findings have yet to be reached. Several variables influence the outcomes, including model complexity, various catchment areas, runoff-generating mechanisms, and variations in radar and rain gauge rainfall datasets (Neary et al. 2004). As a result, it could not draw any broad judgments concerning the validity of radar rainfall data for hydrological modeling.

Therefore, this study determines the performance of radar-based rainfall data for the rainfall runoff simulation in the study watershed. This study used NEXRAD III rainfall data as it is more widely available to the general public and significantly less expensive than using a dense network of gauge stations to achieve the same accuracy (Kalin and Hantush 2006). Similarly, in addition to precise rainfall observation, reliable hydrologic models are also required for the most accurate estimation of runoff hydrograph at watershed outlets during rainfall events. Several agencies have updated their hydrologic models so that radar data can be used to benefit from the increasingly accessible radar data. The Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS), Soil and Water Assessment Tool (SWAT) and Personal Computer Storm Water Management (PCSWMM) are some of the commonly used software for the rainfall-runoff simulation. Previous research has proved the effectiveness of integrating radar and satellite rainfall with HEC-HMS model for the rainfall runoff simulation (Ahmed et al. 2022; Bhusal et al. 2022; Hamedi and Fuentes 2015). Similarly, Sexton et al. (2010); Tuppad et al. (2010); Price et al. (2014) used the integration of SWAT and radar rainfall data for the rainfall-runoff analysis. However, the integration of PCSWMM and radar data is rarely applied for the rainfall runoff simulation. The latest version of PCSWMM supports using radar rainfall data for the real-time application of historical rainfall events and future climate change analysis. The PCSWMM provides a separate project tool known as Radar Acquisition and Processing (RAP) project, from where the user can directly generate the time series rainfall data from radar Network Common Data Form (NETCDF) file for different sub-catchments. Therefore this research took the first step to integrate the PCSWMM model with NEXRAD radar by involving the RAP project tool in the PCSWMM for the rainfall-runoff simulation in different scales of watersheds

This article examines the potential benefits of employing radar-based rainfall data to enhance rainfall-runoff simulation in two watersheds. The study compares two rainfall inputs—NEXRAD III radar products and gauge observations—with hydrologic simulations employing PCSWMM. The structure of this article is as follows. The next section presents the hydrological, meteorological, and geographical data used in this study with an approach for creating a hydrological model and extracting radar rainfall data. In the third section, the results of the hydrologic simulations are reported, and statistical methods are applied to compare the model performance with the radar data to the gauge rainfall data. The research concludes with conclusions and suggestions for enhancing rainfall-runoff simulation.

3.2 METHODOLOGY

This section introduces the proposed study watersheds and provides information on the data used in developing the hydrological models. In addition, this section describes the methodological approach of extracting radar rainfall data and developing PCSWMM hydrological models to study watersheds.

3.2.1 SITE DESCRIPTION

This study proposed two watersheds of different spatial scales to evaluate the performance of radar rainfall data. This study selected Ellerbe Creek and River Des Peres watersheds as study areas which are shown in Figure 7. Ellerbe creek is located in Durham, North Carolina, United States (US), with a drainage area of 57 km^2 and a highly urbanized area with a percentage imperviousness of about 40%. The land use of Ellerbe creek is comprised of 74.6 % residential area, 1 % open water, 15.4 % forest, and 9 % agricultural area. The elevation of Ellerbe creek ranges from 73 m to 173 m above mean sea level. The latitude of Ellerbe creek ranges from 35°59'0" to 36°4'0” in the north, and the longitude ranges from 78° 50' 0" to 78° 58' 0" in the west. Ellerbe creek has one rainfall gauging station at latitude 36°01'43" and longitude 78°54'09".

Similarly, River Des Peres is located in St. Louis, Missouri, US, with a drainage area of 150 km^2 and drains into the Missouri River at St. Louis. The watershed is more urbanized than Ellerbe creek, with a percentage imperviousness of about 55%. The land use of River Des Peres watershed is comprised of 95 % residential area, 1 % open water, 2 % forest, and 2% agricultural area. The elevation of the watershed ranges from 120 m to 225 m above mean sea level. The latitude of River Des Peres ranges from 38°32'0" to 38°40'0" in the north, and the longitude ranges from 90° 16' 0" to 90° 28' 0" in the west. River Des Peres has no rainfall gauging station within the watershed.

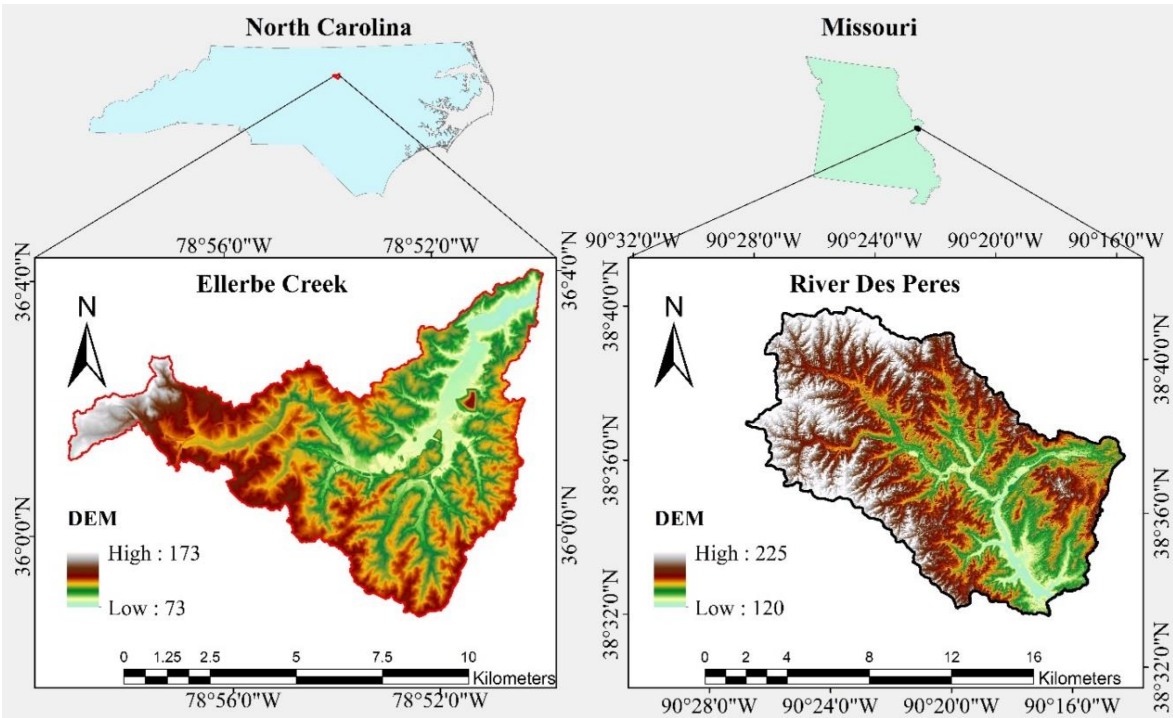


Figure 7 - The figure depicting the characteristics of study watersheds and their location

3.2.2 DATA

Topographical datasets, meteorological data, and hydrological data are all important data required for hydrological analysis of the watershed. Digital Elevation Model (DEM), Land Use and Land Cover data (LULC), and Soil Group Data are important topographical datasets for developing accurate hydrological analysis. All the topographical datasets are extracted from publicly available online platforms and preprocessed in ArcGIS 10.8 before it is imported to PCSWMM models. DEM data was extracted from the United States Department of Agriculture-

National Resources Conservation Service (USDA-NRCS) websites (https://datagateway.nrcs.usda.gov/). One meter DEM data was extracted from USDA-NRSC websites based on the boundary of study watersheds. Similarly, Soil Group dataset was also obtained from USDA-NRCS website. LULC dataset was extracted from Multi-Resolution Land Characteristics (MRLC) Consortium website (https://www.mrlc.gov/).

3.2.3 NEXRAD RADAR DATA

This study applied the NEXRAD III radar data to generate the rainfall data for the study watershed. NEXRAD data is developed by the joint contribution of the Federal Aviation Administration, the U.S. Air Force, and the NOAA National Weather Service. NEXRAD has 160 radars, also known as Weather Surveillance Radar-1988 Doppler-National Weather Service, throughout the United States, and it was first started in 1991(Fulton et al. 1998). NEXRAD III radar data are refined using NEXRAD II, and rainfall data are approximated from reflectivity, also known as (Z-R) relation. These radars have improved hydrologic forecast operations and services and the NWS forecast and warning program by enhancing the detection of severe air velocities, storms, and tornadoes (Fulton et al. 1998). The NEXRAD network surrounds nearly all the continental United States' radar space. Using doppler radars to detect atmospheric precipitation and winds, scientists can supervise and forecast weather phenomena like rain, ice pellets, snow, hail, tornadoes, and some non-weather objects like birds and insects (Fulton et al. 1998). The raw reflectivity data from the WSR-88D needs to go through three stages of hydrometeorological processing before being converted to rainfall depth.

3.2.4 HYDROLOGICAL MODELLING USING PCSWMM

The PCSWMM, a commercial version of SWMM, is a dynamic rainfall-runoff routing model used for a single and long-term event simulation of water quantity and quality for urban and rural watersheds (Rossman 2017). This study integrates the ArcGIS, Hydrologic Engineering Center - River Analysis System (HEC-RAS), and PCSWMM to delineate the watersheds of the proposed study areas. In the first step, the topographical datasets are processed in ArcGIS and imported to the PCSWMM. Similarly, RASMAP in HEC-RAS was used to extract the geometric features of the rivers. The geometric features of the river developed in the HEC-RAS were imported as junctions, conduits, and transects in the PCSWMM. After the topographic data and geometric features were imported into PCSWMM, the watershed was delineated in which the proposed watersheds were divided into different sub-catchments. The Ellerbe Creek and River Des Peres watersheds were divided into 10 and 16 sub-catchments, respectively, which is shown in Figure 8.

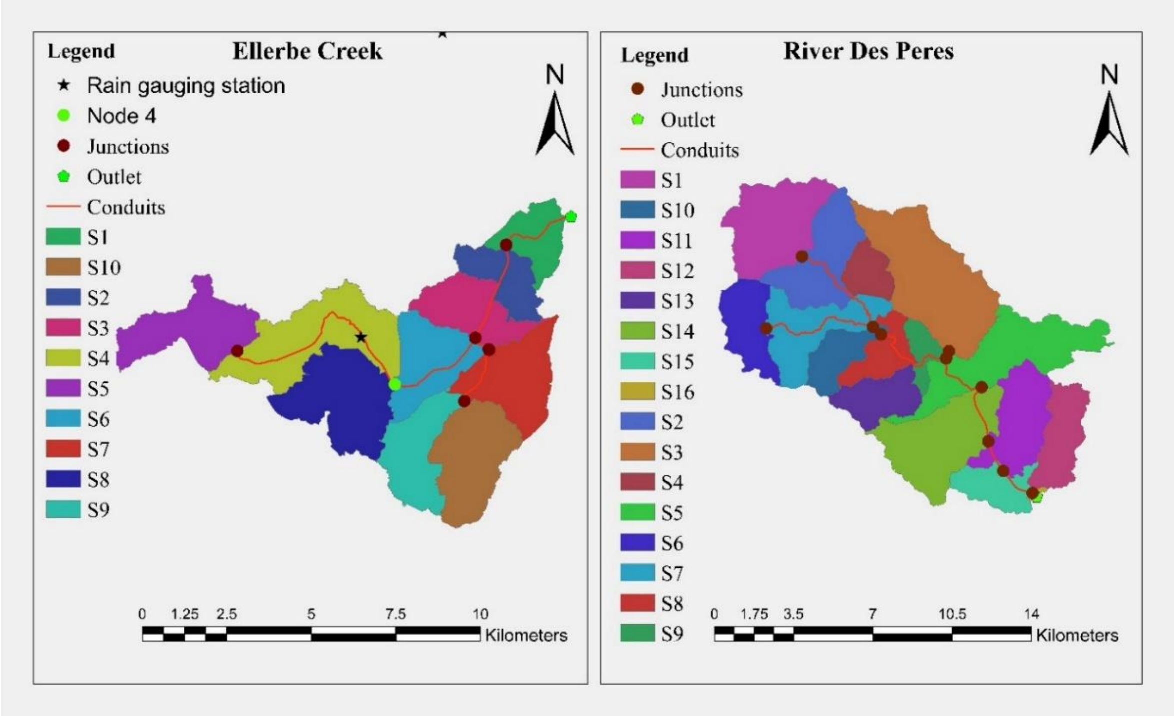


Figure 8 - Watersheds discretization for modelling

The SCS-CN was applied for infiltration from pervious regions, and the dynamic wave was employed for runoff routing. Area, width, slope, and curve number are essential attributes of sub-catchments in the PCSWMM model. Area, width, and slope were automatically generated from DEM data while delineating watersheds in a PCSWMM model. The curve number grid was prepared in ArcGIS by integrating LULC and soil group data and populated in each sub-catchment. Manning's roughness value of each transect was tabulated based on the LULC map of the study watersheds. The default value was used for all the remaining parameters for the sub-catchments that were excluded from the study.

3.2.5 RADAR DATA PROCESSING USING RAP PROJECT IN PCSWMM

The latest version of PCSWMM supports using radar rainfall data for the real-time application of historical rainfall events and future climate change analysis. This study uses the RAP processing modules of the PCSWMM for extracting the rainfall data from the radar file. In the first step, the weather radar (Level III) data was downloaded from the NOAA website. NEXRAD III data consists of 1-hr accumulated rainfall data over a grid 4×4 km^2, also known as Hydrologic Rainfall Analysis Product. Since Ellerbe creek and River Des Peres are in different locations in the United States, so the radar was downloaded from two different NOAA stations. NEXRAD: KRAX and NEXRAD: KLSX were used for downloading the radar data for Ellerbe creek and River Des Peres, respectively. The location of KRAX radar station is at Latitude 35°9'56" and Longitude 78°29'23". Similarly, the location of KLSX radar station is at Latitude 38°41'56" and Longitude 90°40'58". The downloaded Radar data was imported in the RAP modules of the PCSWMM. The RAP project's radar scans are divided into layers, each of which contains reflectivity and rainfall values related to each scan cell. Once the radar files were imported and processed through the RAP module, the radar cell's value was used to create a time series rainfall value for each sub catchment.

3.2.6 RAINFALL EVENT SELECTION

This study selected four rainfall events for Ellerbe Creek and River Des Peres watersheds to compare the simulated hydrograph with the observed hydrograph. The rainfall events that contribute to the flooding in the watershed were selected as the study events for both. In addition, the availability of the record of the discharge hydrograph at the USGS gauging station was considered for the selection of the rainfall events. The snowfall events were discarded to bypass the difficulty of calculating the watershed's melting snow and areal snow distribution. Table 7 provides the list of the selected rainfall events for both watersheds.

Table 7 - Selection of Rainfall events

Rainfall Event	Ellerbe Watershed	River Des Peres Watershed
Event 1	October 11, 2018	September 27, 2020
Event 2	August 23, 2019	September 21, 2021
Event 3	October 25, 2021	July 4, 2022
Event 4	November 24, 2022	November 6, 2022

3.2.7. MODEL PERFORMANCE EVALUATION

This study evaluates the performance of radar data by comparing the radar-generated discharge hydrograph with the observed hydrograph at the watershed outlet. The assessment of the performance of radar rainfall data requires the use of graphical and statistical methods. (Cole and Moore 2008; Johnson et al. 1999). Root mean square error (RMSE), Nash-Sutcliffe efficiency coefficient (NSE), and coefficient of determination (R^2) were applied in this study to evaluate the performance of any hydrological analysis, which are listed in Table 8.

Table 8 - Indices applied for evaluation criteria

Mathematical Expression	Acceptable Range
$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Q_{y,i} - Q_{x,i})^2}{N}}$	
$NSE = 1 - [\frac{\sum_{i=1}^{N}(Q_{x,i} - Q_{y,i})^2}{\sum_{i=1}^{N}(Q_{x,i} - \overline{Q}_x)^2}]$	0.5<NSE≤1
$R^2 = \frac{(\sum_{i=1}^{N}(Q_{x,i} - \overline{Q_{x,i}}) * (Q_{y,i} - \overline{Q_{x,i}}))^2}{\sum_{i=1}^{N}(Q_{x,i} - \overline{Q_{x,i}})^2 * \sum_{i=1}^{N}(Q_{y,i} - \overline{Q_{x,i}})^2}$	>0.5

where $Q_{x,i}$ represent the observed data, $Q_{y,i}$ represents the simulated data, $\overline{Q_{x,i}}$ represents the average value of the total number of observed data, N denotes the total amount of data, and SD represents the standard deviation.

3.3 RESULTS AND DISCUSSION

This section presents the results obtained in this research. In the first part, performance of integrating PCSWMM and NEXRAD data for the Ellerbe creek watershed is presented. In the successive section, the performance of NEXRAD data is presented for River Des Peres watershed.

3.3.1 ELLERBE CREEK WATERSHED

3.3.1.1 NEXRAD RAINFALL ESTIMATES

A comparison of the two rainfall types was performed to understand better how the input rainfall data can affect model outcomes during hydrologic analysis. The rainfall data obtained from the USGS gauging stations (observed data) were compared with the rainfall data generated from the NEXRAD radar using RAP tool in the PCSWMM model. Table 9 listed the comparison of NEXRAD-generated rainfall data with observed rainfall data from the USGS gauging station.

The results revealed that the rainfall data extracted from NEXRAD was underestimated for all rainfall events compared to the observed rainfall data. The total rainfall amount extracted from NEXRAD was slightly underestimated for Events 1 and 2. However, the total rainfall amount generated from NEXRAD is significantly lower than the observed data from gauging stations for Event 3 and Event 4. These results reveal that the NEXRAD data is more reliable for high-intensity rainfall events than the low-intensity rainfall events.

Table 9 - Comparison of observed and NEXRAD generated rainfall data

Event	Total Rainfall Amount		Percentage Differences (mm)
	Observed Rainfall (mm)	NEXRAD Rainfall (mm)	
Event 1	64.5	61.6	4.5 %
Event 2	77.4	70.6	8.7 %
Event 3	44.1	28.9	34.0 %
Event 4	29.2	17.2	41.0 %

Similarly, Figure 9 shows the regression analysis between observed rainfall data and NEXRAD estimated rainfall data. R^2 between the observed and NEXRAD-generated rainfall data for Events 1,2,3, and 4 was 0.91, 0.76, 0.95, and 0.86, respectively. The R^2 value for all the rainfall events ranges from 0.76 to 0.95 and falls in a satisfactory range. These results revealed that NEXRAD data could accurately predict the temporal variability of rainfall data for all the events. In addition, the scatter plot for Events 1, 2, and 3 also demonstrated that NEXRAD could overestimate a small gauge rainfall value and underestimate for large gauge rainfall value. This outcome is consistent with earlier study that assessed the reliability of NEXRAD developed rainfall data (Skinner et al. 2009).

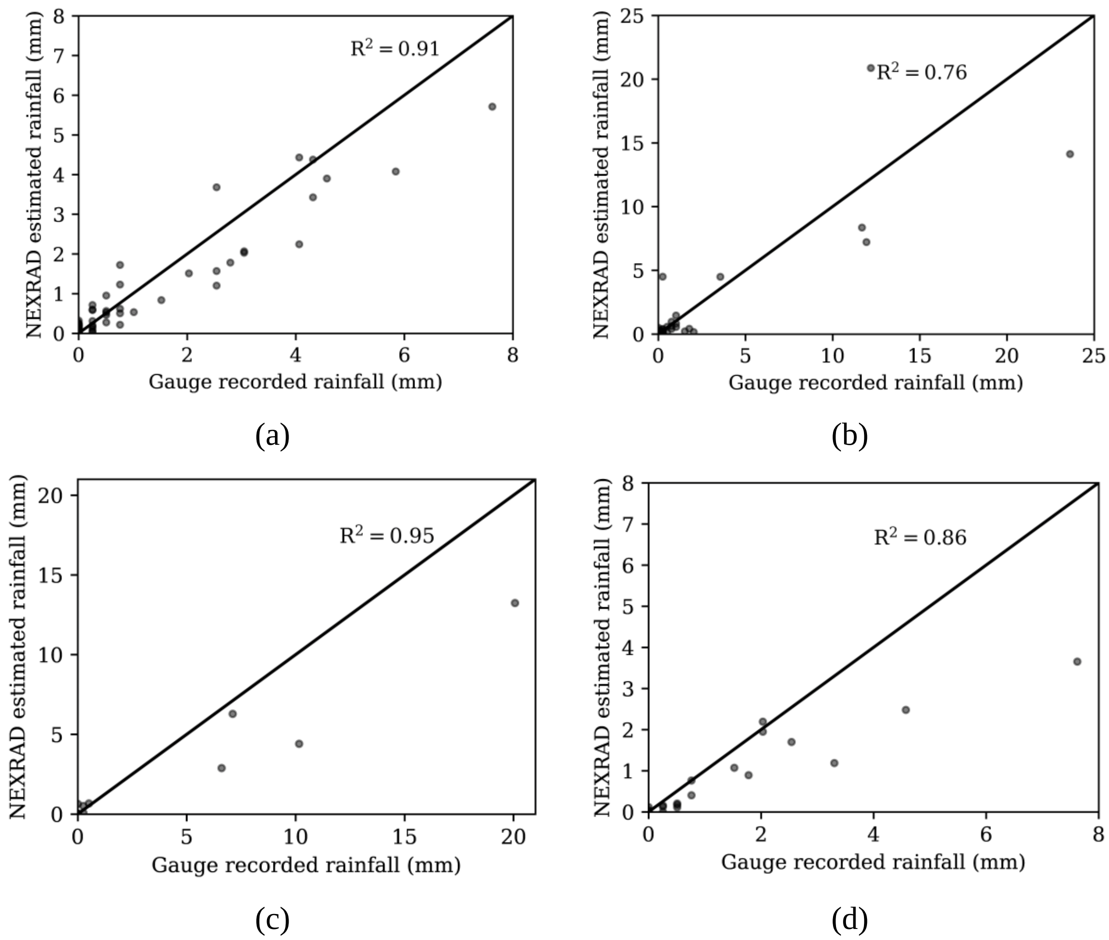


Figure 9 - Regression analysis between gauge recorded and NEXRAD estimated rainfall data, (a) Event 1, (b) Event 2, (c) Event 3, (d) Event 4

3.3.1.2 STREAMFLOW ANALYSIS

Stream flow analysis of the Ellerbe Creek watershed was performed in two locations of the watershed. The stream flow analysis was performed at the junction near the USGS rainfall gauging station and the watershed outlet. Figure 10 shows the discharge hydrograph generated from the gauging station and NEXRAD rainfall data at node 4. The results showed that the performance of both rainfall data obtained from the gauging station and NEXRAD was almost similar in generating the discharge hydrograph, except in Event 4. For the Event 4 analysis, the discharge hydrograph developed by the rainfall gauging station was excellent, while comparing it with the observed hydrograph, with the NSE and R^2 of 0.92 and 0.93 (Table 10). However, for the same rainfall events, NEXRAD estimated rainfall data developed a significantly poor discharge hydrograph with the NSE and R^2 of 0.42 and 0.52, respectively. The average statistical index revealed that discharge was more accurately estimated by applying the rainfall data obtained from the gauging station. The average RMSE, NSE, and R^2 were 2.39 m^3/s, 0.66, and 0.81 for the NEXRAD-generated discharge hydrograph and 2.29 m^3/s, 0.73, and 0.89 for the gauging station-generated discharge hydrograph.

Table 10 - Statistical evaluation of gauge and radar generated discharge hydrograph at node 4.

Event	Radar			Gauge		
	RMSE (m^3/s)	NSE	R^2	RMSE	NSE	R^2
Event 1	1.61	0.89	0.93	3.20	0.62	0.92
Event 2	2.99	0.54	0.93	2.59	0.62	0.92
Event 3	2.31	0.78	0.89	2.38	0.77	0.81
Event 4	2.65	0.42	0.52	1.01	0.92	0.93
Average	2.39	0.66	0.81	2.29	0.73	0.89

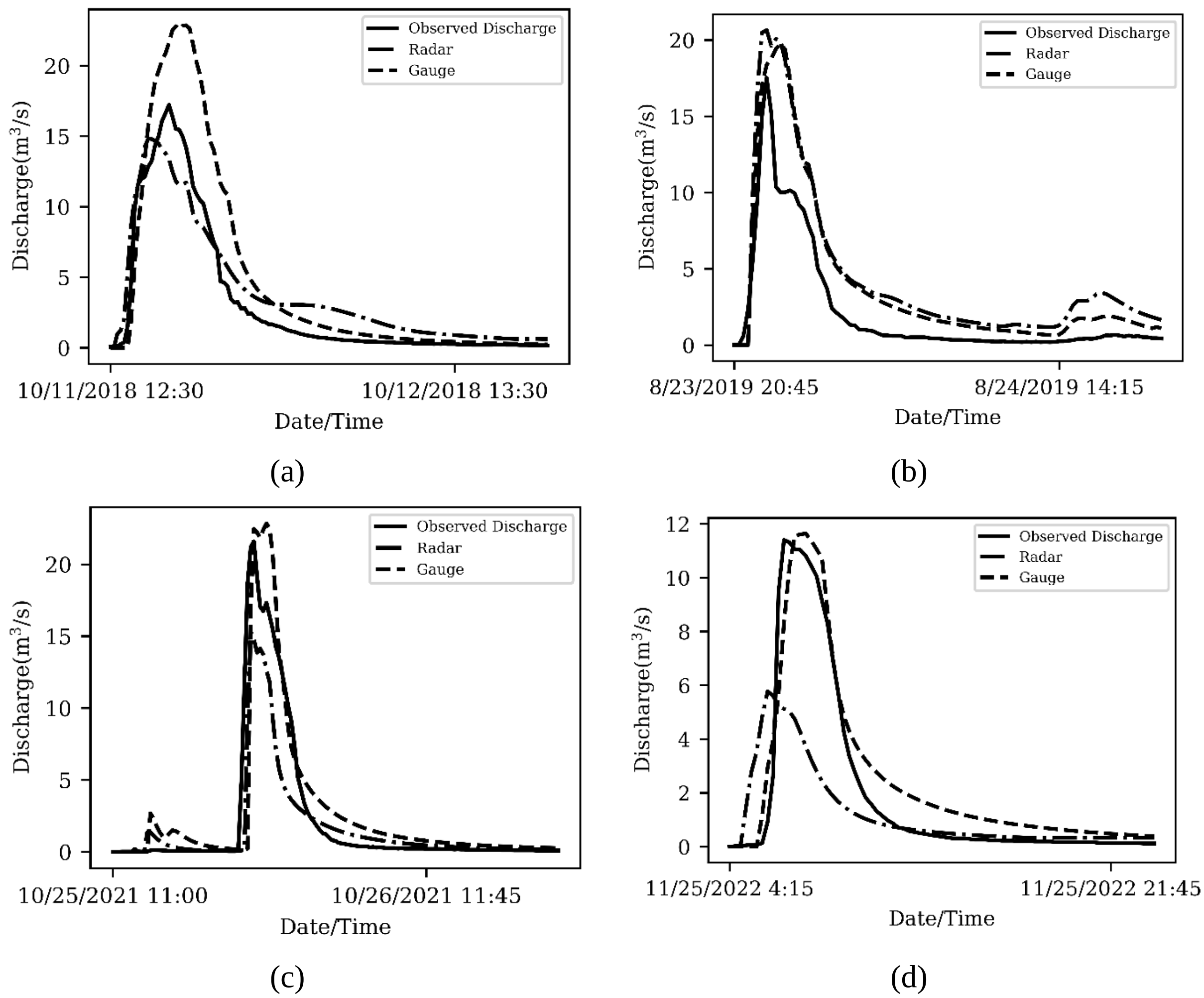


Figure 10 - Graphical representation of observed discharge, radar and gauge generated discharge at node 4, (a) Event 1, (b) Event 2, (c) Event 3, (d) Event 4

Similarly, Figure 11 shows the PCSWMM performance for stream flow analysis at the Ellerbe Creek watershed outlet. The results demonstrated that the rainfall data from the gauging station overestimated the discharge hydrograph at the watershed outlet. For Event 1, Event 3, and Event 4, the gauging station rainfall data resulted in a significantly higher peak discharge while comparing it with the observed discharge at the gauging station. The performance of the rainfall gauging station is significantly poor during Event 4, where the peak discharge simulated using gauge rainfall data is more than double the peak discharge of observed discharge data. In addition, the statistical index value was significantly poor during the Event 4 analysis (Table 11). The NSE and R^2 were -0.94 and 0.48, respectively, which fell in the unacceptable ranges for the hydrologic analysis. Similarly, the average of RMSE, NSE, and R^2 was 7.59 m^3/s, 0.20, and 0.80. The average R^2 value of 0.80 represents that, although gauging stations' rainfall data overestimated the peak discharge, it accurately predicted the nature or pattern of the observed discharge hydrograph at the watershed's outlet.

Similarly, the NEXRAD generated discharge hydrograph was found consistent with observed hydrograph in terms of pattern and peak discharge at watershed outlet. The NEXRAD rainfall data overestimated the peak discharge for Event 2, however, the statistical indices values still fall under the acceptable ranges. Although for Events 1, Event 3 and Event 4, NEXRAD slightly underestimated the peak discharge, the statistical indices value obtained are excellent with highest NSE and R^2 values of both 0.92. The average value of RMSE, NSE and R^2 was 5.63 m^3/s, 0.71 and 0.82, all within the acceptable ranges.

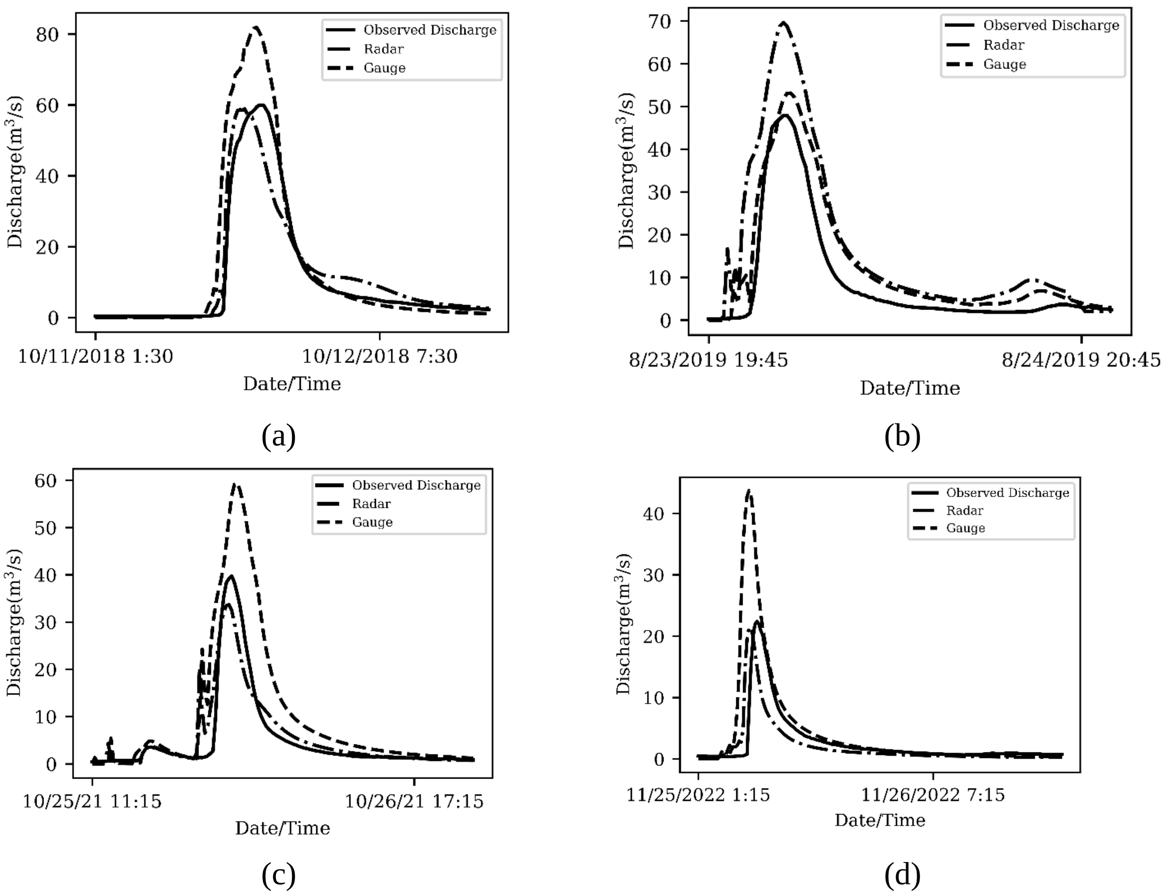


Figure 11 - Graphical representation of observed discharge, radar and gauge generated discharge at Ellerbe creek watershed outlet, (a) Event 1, (b) Event 2, (c) Event 3, (d) Event 4

Table 11 - Statistical evaluation of gauge and radar generated discharge hydrograph at Ellerbe creek watershed outlet

Event	Radar			Gauge		
	RMSE (m^3/s)	NSE	R^2	RMSE	NSE	R^2
Event 1	5.12	0.92	0.92	9.09	0.74	0.93
Event 2	11.09	0.51	0.89	5.53	0.84	0.94
Event 3	3.34	0.86	0.87	9.11	0.22	0.86
Event 4	2.97	0.57	0.59	6.64	-0.94	0.48
Average	5.63	0.71	0.82	7.59	0.21	0.80

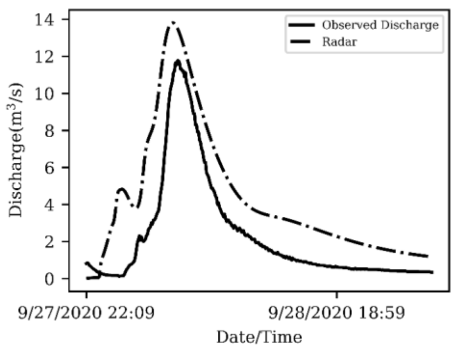
The overall results demonstrate that in almost all the investigation cases for the Ellerbe Creek watershed, integration of the radar data with the PCSWMM model showed a good performance statistically and graphically in generating the discharge hydrograph similar to the USGS observed hydrograph. This is because of the ability of NEXRAD radar to capture the spatial variability of rainfall amounts throughout the watershed. However, the results of integrating the gauging station's rainfall and PCSWMM were mixed. The application of gauging station rainfall data was effective at generating the discharge hydrograph at node 4 and ineffective at generating the hydrograph at the watershed's outlet. There are two possible reasons for the mixed performance of rainfall data from the gauging stations. First, when the hydrological analysis was performed at Node 4, the contributing upstream drainage area to the node 4 was only 22 km^2 and the distance between the rain gauge station and the farthest boundary of the watershed was only 2.5 km. The rainfall gauging station effectively captured the spatial variability of rainfall amount within the 2.5 km vicinity of the gauging station. However, when the analysis was performed at the Ellerbe Creek watershed outlet with a contributing drainage area of about 57 km^2, the distance between the rain gauge station and the farthest location of the watershed was almost 6 km. The rainfall gauging station was found ineffective in capturing the spatial variability of rainfall amount at a distance 6 km farther from its location. Second, the effectiveness of rainfall gauging station also depends upon the wind direction and velocity of storm events (Tsanis et al. 2002; Sexton et al. 2010) . For the state of North Carolina, the storms generally travel from east to west direction. Therefore, rainfall captured by gauging station better approximates rainfall patterns in the west direction of the gauging station.

3.3.2 RIVER DES PERES WATERSHED

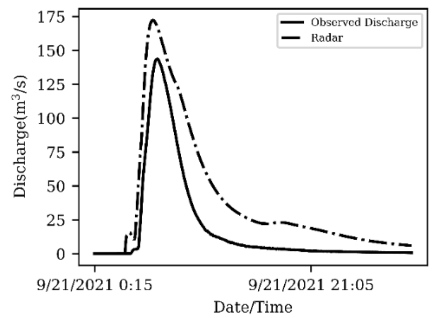
The performance evaluation of integrating the PCSWMM model and NEXRAD radar data was also performed in a River Des Peres watershed, where the contributing drainage area is about 150 km^2. The River Des Peres watershed does not have any rainfall gauging stations within the watershed; therefore, the performance evaluation was done with the observed discharge at the watershed outlet. Figure 12 shows the graphical representation of observed and simulated discharge at the watershed outlet. NEXRAD radar predicted discharge hydrographs were similar in pattern to the observed discharge hydrograph. The NEXRAD radar data slightly underestimated the peak discharge for Event 3 and Event 4 and slightly overestimated the peak discharge for Event 1 and Event 2. The average observed and radar-predicted peak discharge for the four events was 102.92 m^3/s and 75.07 m^3/s, which is listed in Table 12. Similarly, the average RMSE, NSE, and R^2 were 11.35 m^3/s, 0.66, and 0.87, respectively, within the acceptable ranges. The results demonstrated that NEXRAD radar could capture the spatial and temporal variability of total rainfall amount throughout the watershed.

Table 12 - Table 6 Statistical evaluation of radar generated discharge hydrograph at node River Des Peres watershed outlet

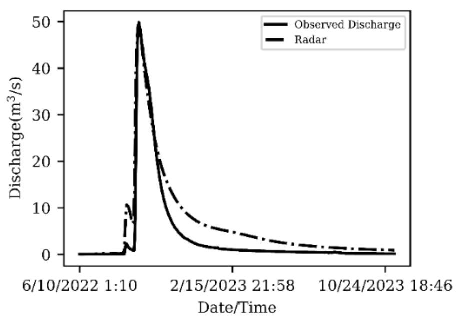
Event	RMSE (m^3/s)	NSE	R^2	Radar Generated Peak Discharge	Observed Peak Discharge	% Peak Difference
Event 1	2.40	0.55	0.87	13.81	11.76	17.49 %
Event 2	25.45	0.62	0.92	172.16	143.48	19.99 %
Event 3	3.40	0.88	0.93	49.37	49.84	0.93 %
Event 4	14.14	0.60	0.67	87.24	95.2	8.35 %
Average	11.35	0.66	0.85	102.92	75.07	11.69%



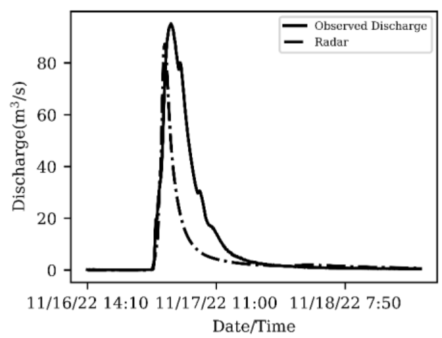
(a)



(b)



(c)



(d)

Figure 12 - Graphical representation of observed, and radar generated discharge at River Des Peres watershed outlet, (a) Event 1, (b) Event 2, (c) Event 3, (d) Event 4

3.4. CONCLUSION

This research was the first to integrate the PCSWMM model with NEXRAD radar by involving the RAP project tool in the PCSWMM for the rainfall-runoff simulation in different scales of watersheds. The premise of the proposed investigation was that incorporating radar data in the PCSWMM model may potentially enhance hydrologic modeling. This study chooses two watersheds (i.e., Ellerbe creek and River Des Peres watersheds) of different locations, drainage areas, and characteristics to test the proposed investigation.

Ellerbe Creek watershed consists of one rainfall gauging station in the most upstream region of the watershed. Therefore, in the first phase of this research, the rainfall data obtained from the USGS gauging stations were compared with the rainfall data generated from the 15-

minute NEXRAD radar using RAP tools in the PCSWMM model. Although the NEXRAD slightly underestimated for large gauge rainfall value and overestimated the small gauge rainfall value, the overall performance of NEXRAD radar on estimating the rainfall amount fell under the acceptable ranges with the average R^2 value of 0.87.

In the second phase of this study, the NEXRAD generated and USGS gauging rainfall data were applied in the PCSWMM model for the Ellerbe Creek watershed hydrological analysis. Two locations (i.e., adjacent to the rainfall gauging station and watershed outlet) were considered for investing the performance of NEXRAD and the gauging station's rainfall data for the hydrologic analysis. The results showed a better efficiency of gauging station rainfall data when the analysis was performed at node 4 (near the rainfall gauging stations). However, the performance of rainfall data from the gauging station was poor in depicting a discharge hydrograph at the watershed outlet; more specifically, the performance was unacceptable for some rainfall events. It may be because the single or sparse rain gauging stations in the more significant watershed cannot detect the spatial variability of rainfall events throughout the watershed. Similarly, the integration of NEXRAD data and the PCSWMM model showed an acceptable performance at both investigating locations of the Ellerbe Creek watershed for rainfall-runoff simulations of all the proposed events. The average value of RMSE, NSE, and R^2 was 5.63 m^3/s, 0.70, and 0.82 at the Ellerbe Creek watershed outlet while applying the NEXRAD radar data; in contrast, they were 7.59 m^3/s, 0.21 and 0.80 for applying rainfall data from gauging stations.

In the final steps of this investigation, the performance of NEXRAD data was evaluated in a River Des Peres watershed, which has a significantly larger drainage area (150 Km^2) compared to the Ellerbe Creek watershed. The River Des Peres watershed does not have any rainfall gauging stations; therefore, the performance of NEXRAD radar data was only evaluated by comparing the discharge hydrograph obtained by integrating NEXRAD radar data and PCSWMM model with observed discharge data at the watershed outlet. The RMSE, NSE, and R2 were 11.35 m^3/s, 0.66, and 0.85, all within the acceptable ranges.

In conclusion, the results supported the investigation premises. This study demonstrated that using rainfall data from NEXRAD can be a feasible replacement for using rainfall data from surface rain gauges in the larger watersheds where rainfall gauging stations are typically scarce or nonexistent. Significant rainfall that caused observed flood peaks in both watersheds was detected using radar-derived rainfall data. In addition, the hydrograph and streamflow peak predictions made by radar-driven PCSWMM models for the larger watersheds were correct for most of the events.

This study also recommends studying the integration of NEXRAD radar data and the PCSWMM model for event-based and continuous simulation in the future. Future researchers can also integrate other radar and satellite-based grid rainfall data, which is supported by the PCSWMM model. In addition, future researchers can compare the performance of the PCSWMM model with other models such as HEC-HMS, SWAT, and Machine learning models for rainfall-runoff analysis by integrating different satellite and radar-based rainfall products.

CHAPTER 4

CONCLUSION AND RECOMMENDATION

This research focused on the performance evaluation of different hydrological models and rainfall products for hydrological analysis. In the first part of this investigation, performance of Random Forest model is compared with the process-based model HEC-HMS for rainfall runoff-simulation with application of satellite rainfall product. Similarly, the second part of this project focuses on performance evaluation of integration of the PCSWMM model with NEXRAD radar by involving the RAP project tool in the PCSWMM for the rainfall-runoff simulation in different scales of watersheds. The conclusions obtained in this project are explained in successive paragraphs.

The first section of 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

Similarly, the research conducted in second part of this project was the first to integrate the PCSWMM model with NEXRAD radar by involving the RAP project tool in the PCSWMM for the rainfall-runoff simulation in different scales of watersheds. The premise of the proposed investigation was that incorporating radar data in the PCSWMM model may potentially enhance hydrologic modeling. This study chooses two watersheds (i.e., Ellerbe creek and River Des Peres watersheds) of different locations, drainage areas, and characteristics to test the proposed investigation. Ellerbe Creek watershed consists of one rainfall gauging station in the most upstream region of the watershed. Therefore, in the first phase of this research, the rainfall data obtained from the USGS gauging stations were compared with the rainfall data generated from the 15-minute NEXRAD radar using RAP tools in the PCSWMM model. Although the NEXRAD slightly underestimated for large gauge rainfall value and overestimated the small gauge rainfall value, the overall performance of NEXRAD radar on estimating the rainfall amount fell under the acceptable ranges with the average R^2 value of 0.87. In the second phase of this study, the NEXRAD generated and USGS gauging rainfall data were applied in the PCSWMM model for the Ellerbe Creek watershed hydrological analysis. Two locations (i.e., adjacent to the rainfall gauging station and watershed outlet) were considered for investing the performance of NEXRAD and the gauging station's rainfall data for the hydrologic analysis. The results showed a better efficiency of gauging station rainfall data when the analysis was performed at node 4 (near the rainfall gauging stations). However, the performance of rainfall data from the gauging station was poor in depicting a discharge hydrograph at the watershed outlet; more specifically, the performance was unacceptable for some rainfall events. It may be because the single or sparse rain gauging stations in the more significant watershed cannot detect the spatial variability of rainfall events throughout the watershed. Similarly, the integration of NEXRAD data and the PCSWMM model showed an acceptable performance at both investigating locations of the Ellerbe Creek watershed for rainfall-runoff simulations of all the proposed events. The average value of RMSE, NSE, and R^2 was 5.63 m^3/s, 0.70, and 0.82 at the Ellerbe Creek watershed outlet while applying the NEXRAD radar data; in contrast, they were 7.59 m^3/s, 0.21 and 0.80 for applying rainfall data from gauging stations. In the final steps of this investigation, the performance of NEXRAD data was evaluated in a River Des Peres watershed, which has a significantly larger drainage area (150 Km^2) compared to the Ellerbe Creek watershed. The River Des Peres watershed does not have any rainfall gauging stations; therefore, the performance of NEXRAD radar data was only evaluated by comparing the discharge hydrograph obtained by integrating NEXRAD radar data and PCSWMM model with observed discharge data at the watershed outlet. The RMSE, NSE, and R2 were 11.35 m^3/s, 0.66, and 0.85, all within the acceptable ranges. In conclusion, the results supported the investigation premises. This study demonstrated that using rainfall data from NEXRAD can be a feasible replacement for using rainfall data from surface rain gauges in the larger watersheds where rainfall gauging stations are typically scarce or nonexistent. Significant rainfall that caused observed flood peaks in both watersheds was detected using radar-derived rainfall data. In addition, the hydrograph and streamflow peak predictions made by radar-driven PCSWMM models for the larger watersheds were correct for most of the events.

This study also recommends studying the integration of NEXRAD radar data and the PCSWMM model for event-based and continuous simulation in the future. Future researchers can also integrate other radar and satellite-based grid rainfall data, which is supported by the PCSWMM model. In addition, future researchers can compare the performance of the PCSWMM model with other models such as HEC-HMS, SWAT, and Machine learning models for rainfall-runoff analysis by integrating different satellite and radar-based rainfall products.

In the future, researchers could work in the following areas:

1. 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.

2. 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.

3. This study also recommends studying the integration of NEXRAD radar data and the PCSWMM model for event-based and continuous simulation in the future. Future researchers can also integrate other radar and satellite-based grid rainfall data, which is supported by the PCSWMM model. In addition, future researchers can compare the performance of the

PCSWMM model with other models such as HEC-HMS, SWAT, and Machine learning models for rainfall-runoff analysis by integrating different satellite and radar-based rainfall products.

REFERENCES

Abbaspour, K. C., E. Rouholahnejad, S. Vaghefi, R. Srinivasan, H. Yang, and B. Kløve. 2015. "A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model." *Journal of Hydrology,* 524: 733–752. https://doi.org/10.1016/j.jhydrol.2015.03.027.

Abbott, M. B., J. C. Bathurst, J. A. Cunge, P. E. O'Connell, and J. Rasmussen. 1986. "An introduction to the European Hydrological System — Systeme Hydrologique Europeen, 'SHE', 1: History and philosophy of a physically-based, distributed modelling system." *Journal of Hydrology*, 87 (1): 45–59. https://doi.org/10.1016/0022-1694(86)90114-9.

Acharya, B., and Joshi, B. (2020). "Flood frequency analysis for an ungauged Himalayan River basin using different methods: A case study of Modi Khola, Parbat, Nepal." *Meteorology Hydrology and Water Management*, 8(2), 46–51.

Adnan, R. M., Z. Liang, S. Heddam, M. Zounemat-Kermani, O. Kisi, and B. Li. 2020. "Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs." *Journal of Hydrology*, 586: 124371. https://doi.org/10.1016/j.jhydrol.2019.124371.

Adnan, R. M., A. Petroselli, S. Heddam, C. A. G. Santos, and O. Kisi. 2021. "Short term rainfall-runoff modelling using several machine learning methods and a conceptual event-based model." *Stoch Environ Res Risk Assess,* 35 (3): 597–616. https://doi.org/10.1007/s00477-020-01910-0.

Ahmad, S., A. Kalra, and H. Stephen. 2010. "Estimating soil moisture using remote sensing data: A machine learning approach." *Advances in water resources,* 33 (1): 69–80. Elsevier.

Ahmed, S. I., R. Rudra, P. Goel, A. Khan, B. Gharabaghi, and R. Sharma. 2022. “A Comparative Evaluation of Using Rain Gauge and NEXRAD Radar-Estimated Rainfall Data for Simulating Streamflow.” *Hydrology*, 9 (8): 133. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/hydrology9080133.

Archer, D. r., and H. j. Fowler. 2018. “Characterising flash flood response to intense rainfall and impacts using historical information and gauged data in Britain.” *Journal of Flood Risk Management*, 11 (S1): S121–S133. https://doi.org/10.1111/jfr3.12187.

Aryal, A., A. Acharya, and A. Kalra. 2022. “Assessing the Implication of Climate Change to Forecast Future Flood Using CMIP6 Climate Projections and HEC-RAS Modeling.” *Forecasting*, 4 (3): 582–603. https://doi.org/10.3390/forecast4030032.

Asadollah, S. B. H. S., A. Sharafati, D. Motta, and Z. M. Yaseen. 2021. “River water quality index prediction and uncertainty analysis: A comparative study of machine learning models.” *Journal of Environmental Chemical Engineering*, 9 (1): 104599. https://doi.org/10.1016/j.jece.2020.104599.

Bachmair, S., C. Svensson, I. Prosdocimi, J. Hannaford, and K. Stahl. 2017. “Developing drought impact functions for drought risk management.” *Natural Hazards and Earth System Sciences*, 17: 1947–1960. https://doi.org/10.5194/nhess-17-1947-2017.

Bedient, P. B., B. C. Hoblit, D. C. Gladwell, and B. E. Vieux. 2000. “NEXRAD Radar for Flood Prediction in Houston.” *Journal of Hydrologic Engineering*, 5 (3): 269–277. American Society of Civil Engineers. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:3(269).

Bhandari, M., N. Nyaupane, S. R. Mote, A. Kalra, and S. Ahmad. 2017. “2D Unsteady Routing and Flood Inundation Mapping for Lower Region of Brazos River Watershed.”

Bhandari, S., A. Kalra, K. Tamaddun, and S. Ahmad. 2018. “Relationship between Ocean-Atmospheric Climate Variables and Regional Streamflow of the Conterminous United States.” *Hydrology*, 5 (2): 30. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/hydrology5020030.

Bhandari, S., B. Thakur, A. Kalra, W. P. Miller, V. Lakshmi, and P. Pathak. 2019. “Streamflow Forecasting Using Singular Value Decomposition and Support Vector Machine for the Upper Rio Grande River Basin.” *JAWRA Journal of the American Water Resources Association*, 55 (3): 680–699. https://doi.org/10.1111/1752-1688.12733.

Bhusal, A., U. Parajuli, S. Regmi, and A. Kalra. 2022. “Application of Machine Learning and Process-Based Models for Rainfall-Runoff Simulation in DuPage River Basin, Illinois.” *Hydrology*, 9 (7): 117. https://doi.org/10.3390/hydrology9070117.

Biau, G., and E. Scornet. 2016. “A random forest guided tour.” *TEST*, 25 (2): 197–227. https://doi.org/10.1007/s11749-016-0481-7.

Breiman, L. 2001. “Random Forests.” *Machine Learning*, 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Carrier, C., A. Kalra, and S. Ahmad. 2013. “Using Paleo Reconstructions to Improve Streamflow Forecast Lead Time in the Western United States.” *JAWRA Journal of the American Water Resources Association*, 49 (6): 1351–1366. https://doi.org/10.1111/jawr.12088.

Chang, H., and M. Psaris. 2013. “Local landscape predictors of maximum stream temperature and thermal sensitivity in the Columbia River Basin, USA.” *Science of The Total Environment*, 461–462: 587–600. https://doi.org/10.1016/j.scitotenv.2013.05.033.

Chaubey, I., C. T. Haan, J. M. Salisbury, and S. Grunwald. 1999. “Quantifying Model Output Uncertainty Due to Spatial Variability of Rainfall1.” *JAWRA Journal of the American*

Water Resources Association, 35 (5): 1113–1123. https://doi.org/10.1111/j.1752-1688.1999.tb04198.x.

Chen, C., A. Kalra, and S. Ahmad. 2018. "Hydrologic responses to climate change using downscaled GCM data on a watershed scale." *Journal of Water and Climate Change,* 10 (1): 63–77. https://doi.org/10.2166/wcc.2018.147.

Chiang, S., C.-H. Chang, and W.-B. Chen. 2022. "Comparison of Rainfall-Runoff Simulation between Support Vector Regression and HEC-HMS for a Rural Watershed in Taiwan." *Water*, 14 (2): 191. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/w14020191.

Choubin, B., H. Darabi, O. Rahmati, F. Sajedi-Hosseini, and B. Kløve. 2018. "River suspended sediment modelling using the CART model: A comparative study of machine learning techniques." *Science of The Total Environment*, 615: 272–281. https://doi.org/10.1016/j.scitotenv.2017.09.293.

Cole, S. J., and R. J. Moore. 2008. "Hydrological modelling using raingauge- and radar-based estimators of areal rainfall." *Journal of Hydrology*, 358 (3): 159–181. https://doi.org/10.1016/j.jhydrol.2008.05.025.

Cristiano, E., M.-C. ten Veldhuis, and N. van de Giesen. 2017. "Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review." *Hydrology and Earth System Sciences*, 21 (7): 3859–3878. Copernicus GmbH. https://doi.org/10.5194/hess-21-3859-2017.

Deng, T., K.-W. Chau, and H.-F. Duan. 2021. "Machine learning based marine water quality prediction for coastal hydro-environment management." *Journal of Environmental Management*, 284: 112051. https://doi.org/10.1016/j.jenvman.2021.112051.

Desai, S., and T. B. M. J. Ouarda. 2021. "Regional hydrological frequency analysis at ungauged sites with random forest regression." *Journal of Hydrology*, 594: 125861. https://doi.org/10.1016/j.jhydrol.2020.125861.

Erdal, H. I., and O. Karakurt. 2013. "Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms." *Journal of Hydrology*, 477: 119–128. https://doi.org/10.1016/j.jhydrol.2012.11.015.

Faccini, F., F. Luino, G. Paliaga, A. Sacchini, L. Turconi, and C. de Jong. 2018. "Role of rainfall intensity and urban sprawl in the 2014 flash flood in Genoa City, Bisagno catchment (Liguria, Italy)." *Applied Geography*, 98: 224–241. https://doi.org/10.1016/j.apgeog.2018.07.022.

Feigl, M., K. Lebiedzinski, M. Herrnegger, and K. Schulz. 2021. "Machine-learning methods for stream water temperature prediction." *Hydrology and Earth System Sciences*, 25 (5): 2951–2977. Copernicus GmbH. https://doi.org/10.5194/hess-25-2951-2021.

Feng, Q., J. Liu, and J. Gong. 2015. "Urban Flood Mapping Based on Unmanned Aerial Vehicle Remote Sensing and Random Forest Classifier—A Case of Yuyao, China." *Water*, 7 (4): 1437–1455. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/w7041437.

Gaines, J. M. 2016. "Flooding: Water potential." *Nature*, 531 (7594): S54–S55. Nature Publishing Group. https://doi.org/10.1038/531S54a.

Gaume, E., V. Bain, P. Bernardara, O. Newinger, M. Barbuc, A. Bateman, L. Blaškovičová, G. Blöschl, M. Borga, A. Dumitrescu, I. Daliakopoulos, J. Garcia, A. Irimescu, S. Kohnova, A. Koutroulis, L. Marchi, S. Matreata, V. Medina, E. Preciso, D. Sempere-Torres, G. Stancalie, J. Szolgay, I. Tsanis, D. Velasco, and A. Viglione. 2009. "A compilation of data on European flash floods." *Journal of Hydrology*, 367 (1): 70–78. https://doi.org/10.1016/j.jhydrol.2008.12.028.

Gharbi, M., A. Soualmia, D. Dartus, and L. Masbernat. 2016. "Comparison of 1D and 2D Hydraulic Models for Floods Simulation on the Medjerda Riverin Tunisia." 10.

Ghazali, D. A., M. Guericolas, F. Thys, F. Sarasin, P. Arcos González, and E. Casalino. 2018. "Climate Change Impacts on Disaster and Emergency Medicine Focusing on Mitigation Disruptive Effects: an International Perspective." *International Journal of Environmental Research and Public Health*, 15 (7): 1379. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/ijerph15071379.

Ghimire, G., R. Thakali, A. Kalra, and S. Ahmad. 2016. *Role of Low Impact Development in the Attenuation of Flood Flows in Urban Areas*. 349.

Ghimire, S., Z. M. Yaseen, A. A. Farooque, R. C. Deo, J. Zhang, and X. Tao. n.d. "OPEN Streamflow prediction using." *Scientific Reports*, 27.

Gregorutti, B., B. Michel, and P. Saint-Pierre. 2017. "Correlation and variable importance in random forests." *Stat Comput*, 27 (3): 659–678. https://doi.org/10.1007/s11222-016-9646-1.

Guerreiro, S. B., R. J. Dawson, C. Kilsby, E. Lewis, and A. Ford. 2018. "Future heat-waves, droughts and floods in 571 European cities." *Environ. Res. Lett.*, 13 (3): 034009. IOP Publishing. https://doi.org/10.1088/1748-9326/aaaad3.

Guo, W.-D., W.-B. Chen, S.-H. Yeh, C.-H. Chang, and H. Chen. 2021. "Prediction of River Stage Using Multistep-Ahead Machine Learning Techniques for a Tidal River of Taiwan." *Water*, 13 (7): 920. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/w13070920.

Gupta, H. V., S. Sorooshian, and P. O. Yapo. 1999. "Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration." *J. Hydrol. Eng.*, 4 (2): 135–143. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135).

Haberlandt, U. 2007. "Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event." *Journal of Hydrology*, 332 (1): 144–157. https://doi.org/10.1016/j.jhydrol.2006.06.028.

Hallegatte, S., C. Green, R. Nicholls, and J. Corfee-Morlot. 2013. "Future flood losses in major coastal cities." *Nature Climate Change*, 3: 802–806. https://doi.org/10.1038/nclimate1979.

Halwatura, D., and M. M. M. Najim. 2013. "Application of the HEC-HMS model for runoff simulation in a tropical catchment." *Environmental Modelling & Software*, 46: 155–162. https://doi.org/10.1016/j.envsoft.2013.03.006.

Hamedi, A., and H. R. Fuentes. 2015. "Comparative Effectiveness and Reliability of NEXRAD Data to Predict Outlet Hydrographs Using the GSSHA and HEC-HMS Hydrologic Models." 1444–1453. American Society of Civil Engineers. https://doi.org/10.1061/9780784479162.142.

Hong, Y., D. Gochis, J. Cheng, K. Hsu, and S. Sorooshian. 2007. "Evaluation of PERSIANN-CCS Rainfall Measurement Using the NAME Event Rain Gauge Network." *Journal of Hydrometeorology*, 8 (3): 469–482. https://doi.org/10.1175/JHM574.1.

Hussain, D., and A. A. Khan. 2020. "Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan." *Earth Sci Inform*, 13 (3): 939–949. https://doi.org/10.1007/s12145-020-00450-z.

Hussein, E. A., C. Thron, M. Ghaziasgar, A. Bagula, and M. Vaccari. 2020. “Groundwater Prediction Using Machine-Learning Tools.” *Algorithms*, 13 (11): 300. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/a13110300.

Jenkins, K., S. Surminski, J. Hall, and F. Crick. 2017. “Assessing surface water flood risk and management strategies under future climate change: Insights from an Agent-Based Model.” *Science of The Total Environment*, 595: 159–168. https://doi.org/10.1016/j.scitotenv.2017.03.242.

Johnson, D., M. Smith, V. Koren, and B. Finnerty. 1999. “Comparing Mean Areal Precipitation Estimates from NEXRAD and Rain Gauge Networks.” *Journal of Hydrologic Engineering*, 4 (2): 117–124. American Society of Civil Engineers. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(117).

Jordan, M. I., and T. M. Mitchell. 2015. “Machine learning: Trends, perspectives, and prospects.” *Science,* 349 (6245): 255–260. American Association for the Advancement of Science. https://doi.org/10.1126/science.aaa8415.

Joshi, N., A. Bista, I. Pokhrel, A. Kalra, and S. Ahmad. 2019a. “Rainfall-Runoff Simulation in Cache River Basin, Illinois, Using HEC-HMS.” *World Environmental and Water Resources Congress 2019*, 348–360. Pittsburgh, Pennsylvania: American Society of Civil Engineers.

Joshi, N., A. Bista, I. Pokhrel, A. Kalra, and S. Ahmad. 2019b. *Rainfall-Runoff Simulation in Cache River Basin, Illinois, Using HEC-HMS*. 360.

Joshi, N., and A. Kalra. 2021. “Analyzing the Association between ENSO and Groundwater Rise in the South Atlantic-Gulf Region in the Southeastern United States.” *Hydrology*, 8 (3):

119. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/hydrology8030119.

Joshi, N., A. Kalra, B. Thakur, K. W. Lamb, and S. Bhandari. 2021. “Analyzing the Effects of Short-Term Persistence and Shift in Sea Level Records along the US Coast.” *Hydrology*, 8 (1): 17. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/hydrology8010017.

Joshi, N., K. Tamaddun, R. Parajuli, A. Kalra, P. Maheshwari, L. Mastino, and M. Velotta. 2020. “Future Changes in Water Supply and Demand for Las Vegas Valley: A System Dynamic Approach based on CMIP3 and CMIP5 Climate Projections.” *Hydrology*, 7 (1): 16. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/hydrology7010016.

Kalin, L., and M. M. Hantush. 2006. “Hydrologic Modeling of an Eastern Pennsylvania Watershed with NEXRAD and Rain Gauge Data.” *Journal of Hydrologic Engineering*, 11 (6): 555–569. American Society of Civil Engineers. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(555).

Kalra, A., and S. Ahmad. 2007. “Using oceanic-atmospheric oscillations for long lead-time streamflow forecasting in the Upper Colorado River Basin.” 2007: H11C-0646.

Kalra, A., and S. Ahmad. 2011. “Evaluating changes and estimating seasonal precipitation for the Colorado River Basin using a stochastic nonparametric disaggregation technique.” *Water Resources Research*, 47 (5). https://doi.org/10.1029/2010WR009118.

Kalra, A., S. Ahmad, and A. Nayak. 2013a. “Increasing streamflow forecast lead time for snowmelt-driven catchment based on large-scale climate patterns.” *Advances in Water Resources*, 53: 150–162. https://doi.org/10.1016/j.advwatres.2012.11.003.

Kalra, A., N. Joshi, S. Baral, S. Nhuchhen Pradhan, M. Mambepa, S. Paudel, C. Xia, and R. Gupta. 2021a. “Coupled 1D and 2D HEC-RAS Floodplain Modeling of Pecos River in New Mexico.” 165–178. American Society of Civil Engineers. https://doi.org/10.1061/9780784483466.016.

Kalra, A., N. Joshi, I. Pokhrel, S. Nhuchhen Pradhan, P. Adhikari, C. Xia, and R. Gupta. 2021b. *Assessment of Floodplain Inundation Mapping of Davenport City in Iowa Using Civil Geo-HECRAS*. 192.

Kalra, A., W. P. Miller, K. W. Lamb, S. Ahmad, and T. Piechota. 2013b. “Using large-scale climatic patterns for improving long lead time streamflow forecasts for Gunnison and San Juan River Basins.” *Hydrological Processes*, 27 (11): 1543–1559. https://doi.org/10.1002/hyp.9236.

Kalra, A., T. Piechota, R. Davies, and G. Tootle. 2008. “Changes in U.S. Streamflow and Western U.S. Snowpack.” *Journal of Hydrologic Engineering - J HYDROL ENG*, 13. https://doi.org/10.1061/(ASCE)1084-0699(2008)13:3(156).

Kalra, A., K. A. Tamaddun, J. Park, and S. Ahmad. 2018. “Hydro-climatological and Socio-economic Predictors for Water Demand Forecasting Using Machine Learning and Artificial Intelligence.” 2018: H34A-06.

Kastridis, A., and D. Stathis. 2017. “The Effect of Rainfall Intensity on the Flood Generation of Mountainous Watersheds (Chalkidiki Prefecture, North Greece).” *Perspectives on Atmospheric Sciences*, Springer Atmospheric Sciences, T. Karacostas, A. Bais, and P. T. Nastos, eds., 341–347. Cham: Springer International Publishing.

Khedri, A., N. Kalantari, and M. Vadiati. 2020. “Comparison study of artificial intelligence method for short term groundwater level prediction in the northeast Gachsaran unconfined aquifer.” *Water Supply*, 20 (3): 909–921. IWA Publishing. https://doi.org/10.2166/ws.2020.015.

Kim, B., B. F. Sanders, J. S. Famiglietti, and V. Guinot. 2015. “Urban flood modeling with porous shallow-water equations: A case study of model errors in the presence of anisotropic porosity.” *Journal of Hydrology*, 523: 680–692. https://doi.org/10.1016/j.jhydrol.2015.01.059.

Krajewski, W. F., V. Lakshmi, K. P. Georgakakos, and S. C. Jain. 1991. “A Monte Carlo Study of rainfall sampling effect on a distributed catchment model.” *Water Resources Research*, 27 (1): 119–128. https://doi.org/10.1029/90WR01977.

Kumar, N., S. K. Singh, P. K. Srivastava, and B. Narsimlu. 2017. “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.*, 3 (1): 30. https://doi.org/10.1007/s40808-017-0306-z.

Kundzewicz, Z. W., S. Kanae, S. I. Seneviratne, J. Handmer, N. Nicholls, P. Peduzzi, R. Mechler, L. M. Bouwer, N. Arnell, K. Mach, R. Muir-Wood, G. R. Brakenridge, W. Kron, G. Benito, Y. Honda, K. Takahashi, and B. Sherstyukov. 2014. “Flood risk and climate change: global and regional perspectives.” *Hydrological Sciences Journal*, 59 (1): 1–28. Taylor & Francis. https://doi.org/10.1080/02626667.2013.857411.

Li, B., G. Yang, R. Wan, X. Dai, and Y. Zhang. 2016. “Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China.” *Hydrology Research*, 47 (S1): 69–83. https://doi.org/10.2166/nh.2016.264.

Lopez, V., F. Napolitano, and F. Russo. 2005. “Calibration of a rainfall-runoff model using radar and raingauge data.” *Advances in Geosciences*, 41–46. Copernicus GmbH.

Melesse, A. M., K. Khosravi, J. P. Tiefenbacher, S. Heddam, S. Kim, A. Mosavi, and B. T. Pham. 2020. "River Water Salinity Prediction Using Hybrid Machine Learning Models." *Water*, 12 (10): 2951. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/w12102951.

Meng, Y., M. Yang, S. Liu, Y. Mou, C. Peng, and X. Zhou. 2021. "Quantitative assessment of the importance of bio-physical drivers of land cover change based on a random forest method." *Ecological Informatics*, 61: 101204. https://doi.org/10.1016/j.ecoinf.2020.101204.

Merwade, V., F. Olivera, M. Arabi, and S. Edleman. 2008. "Uncertainty in Flood Inundation Mapping: Current Issues and Future Directions." *J. Hydrol. Eng.*, 13 (7): 608–620. https://doi.org/10.1061/(ASCE)1084-0699(2008)13:7(608).

Merz, B., H. Kreibich, R. Schwarze, and A. Thieken. 2010. "Review article 'Assessment of economic flood damage.'" *Natural Hazards and Earth System Sciences*, 10 (8): 1697–1724. Copernicus GmbH. https://doi.org/10.5194/nhess-10-1697-2010.

Mewes, B., H. Oppel, V. Marx, and A. Hartmann. 2020. "Information-Based Machine Learning for Tracer Signature Prediction in Karstic Environments." *Water Resources Research*, 56 (2): e2018WR024558. https://doi.org/10.1029/2018WR024558.

Min, S.-K., X. Zhang, F. W. Zwiers, and G. C. Hegerl. 2011. "Human contribution to more-intense precipitation extremes." *Nature*, 470 (7334): 378–381. Nature Publishing Group. https://doi.org/10.1038/nature09763.

Moriasi, D., M. Gitau, N. Pai, and P. Daggupati. 2015. "Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria." Accessed December 7, 2022. https://doi.org/10.13031/trans.58.10715.

Muñoz, P., J. Orellana-Alvear, P. Willems, and R. Célleri. 2018. "Flash-Flood Forecasting in an Andean Mountain Catchment—Development of a Step-Wise Methodology Based on the Random Forest Algorithm." *Water*, 10 (11): 1519. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/w10111519.

Neary, V. S., E. Habib, and M. Fleming. 2004. "Hydrologic Modeling with NEXRAD Precipitation in Middle Tennessee." *J. Hydrol. Eng.*, 9 (5): 339–349. https://doi.org/10.1061/(ASCE)1084-0699(2004)9:5(339).

Nguyen, D. T., and S.-T. Chen. 2020. "Real-Time Probabilistic Flood Forecasting Using Multiple Machine Learning Methods." *Water*, 12 (3): 787. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/w12030787.

Nguyen, P., E. J. Shearer, H. Tran, M. Ombadi, N. Hayatbini, T. Palacios, P. Huynh, D. Braithwaite, G. Updegraff, K. Hsu, B. Kuligowski, W. S. Logan, and S. Sorooshian. 2019. "The CHRS Data Portal, an easily accessible public repository for PERSIANN global satellite precipitation data." *Sci Data*, 6 (1): 180296. https://doi.org/10.1038/sdata.2018.296.

Ni, L., D. Wang, V. P. Singh, J. Wu, Y. Wang, Y. Tao, and J. Zhang. 2020. "Streamflow and rainfall forecasting by two long short-term memory-based models." *Journal of Hydrology*, 583: 124296. https://doi.org/10.1016/j.jhydrol.2019.124296.

Nyaupane, N., S. Bhandari, M. M. Rahaman, K. Wagner, A. Kalra, S. Ahmad, and R. Gupta. 2018. *Flood Frequency Analysis Using Generalized Extreme Value Distribution and Floodplain Mapping for Hurricane Harvey in Buffalo Bayou*. 375.

Parajuli, R., A. Kalra, L. Mastino, M. Velotta, and S. Ahmad. 2017. "A System Dynamics Modeling of Water Supply and Demand in Las Vegas Valley." 2017: PA11B-0209.

Parisouj, P., H. Mohebzadeh, and T. Lee. 2020. “Employing Machine Learning Algorithms for Streamflow Prediction: A Case Study of Four River Basins with Different Climatic Zones in the United States.” *Water Resour Manage,* 34 (13): 4113–4131. https://doi.org/10.1007/s11269-020-02659-5.

Park, H., K. Kim, and D. kun Lee. 2019. “Prediction of Severe Drought Area Based on Random Forest: Using Satellite Image and Topography Data.” *Water*, 11 (4): 705. https://doi.org/10.3390/w11040705.

Pathak, P., A. Kalra, and S. Ahmad. 2017. “Temperature and precipitation changes in the Midwestern United States: implications for water management.” *International Journal of Water Resources Development,* 33 (6): 1003–1019. Routledge. https://doi.org/10.1080/07900627.2016.1238343.

Pathak, P., A. Kalra, K. W. Lamb, W. P. Miller, S. Ahmad, R. Amerineni, and D. P. Ponugoti. 2018. “Climatic variability of the Pacific and Atlantic Oceans and western US snowpack.” *International Journal of Climatology,* 38 (3): 1257–1269. https://doi.org/10.1002/joc.5241.

Pathan, A. I., and P. G. Agnihotri. 2021. “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.*, 7 (2): 1133–1144. https://doi.org/10.1007/s40808-020-00961-0.

Pechlivanidis, I. G., N. McIntyre, and H. S. Wheater. 2016. “The significance of spatial variability of rainfall on simulated runoff: an evaluation based on the Upper Lee catchment, UK.” *Hydrology Research*, 48 (4): 1118–1130. https://doi.org/10.2166/nh.2016.038.

Pessoa, M. L., R. L. Bras, and E. R. Williams. 1993. "Use of Weather Radar for Flood Forecasting in the Sieve River Basin: A Sensitivity Analysis." *Journal of Applied Meteorology and Climatology*, 32 (3): 462–475. American Meteorological Society. https://doi.org/10.1175/1520-0450(1993)032<0462:UOWRFF>2.0.CO;2.

Pokhrel, I., A. Kalra, M. M. Rahaman, and R. Thakali. 2020. "Forecasting of Future Flooding and Risk Assessment under CMIP6 Climate Projection in Neuse River, North Carolina." *Forecasting*, 2 (3): 323–345. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/forecast2030018.

Porcù, F., L. Milani, and M. Petracca. 2014. "On the uncertainties in validating satellite instantaneous rainfall estimates with raingauge operational network." *Atmospheric Research*, Perspectives of Precipitation Science - Part II, 144: 73–81. https://doi.org/10.1016/j.atmosres.2013.12.007.

Price, K., S. T. Purucker, S. R. Kraemer, J. E. Babendreier, and C. D. Knightes. 2014. "Comparison of radar and gauge precipitation data in watershed models across varying spatial and temporal scales." *Hydrological Processes*, 28 (9): 3505–3520. https://doi.org/10.1002/hyp.9890.

Radhakrishnan, C., V. Chandrasekar, S. C. Reising, and W. Berg. 2022. "Rainfall Estimation From TEMPEST-D CubeSat Observations: A Machine-Learning Approach." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 3626–3636. https://doi.org/10.1109/JSTARS.2022.3170835.

Rahaman, M. M., B. Thakur, A. Kalra, R. Li, and P. Maheshwari. 2019. "Estimating High-Resolution Groundwater Storage from GRACE: A Random Forest Approach."

Environments, 6 (6): 63. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/environments6060063.

Rajaee, T., S. Khani, and M. Ravansalar. 2020. "Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review." *Chemometrics and Intelligent Laboratory Systems*, 200: 103978. https://doi.org/10.1016/j.chemolab.2020.103978.

Rezaei, K., B. Pradhan, M. Vadiati, and A. A. Nadiri. 2021. "Suspended sediment load prediction using artificial intelligence techniques: comparison between four state-of-the-art artificial neural network techniques." *Arab J Geosci*, 14 (3): 215. https://doi.org/10.1007/s12517-020-06408-1.

Rezaei, K., and M. Vadiati. 2020. "A comparative study of artificial intelligence models for predicting monthly river suspended sediment load." *Journal of Water and Land Development*, no. 45. https://doi.org/10.24425/jwld.2020.133052.

Rezaeianzadeh, M., A. Stein, H. Tabari, H. Abghari, N. Jalalkamali, E. Z. Hosseinipour, and V. P. Singh. 2013. "Assessment of a conceptual hydrological model and artificial neural networks for daily outflows forecasting." *Int. J. Environ. Sci. Technol.*, 10 (6): 1181–1192. https://doi.org/10.1007/s13762-013-0209-0.

Rossman, L. A. 2017. "Storm Water Management Model Reference Manual Volume II – Hydraulics." 190.

Saadi, M., L. Oudin, and P. Ribstein. 2019. "Random Forest Ability in Regionalizing Hourly Hydrological Model Parameters." *Water*, 11 (8): 1540. https://doi.org/10.3390/w11081540.

Sagarika, S., A. Kalra, and S. Ahmad. 2014. “Evaluating the effect of persistence on long-term trends and analyzing step changes in streamflows of the continental United States.” *Journal of Hydrology*, 517: 36–53. https://doi.org/10.1016/j.jhydrol.2014.05.002.

Sahoo, S., T. A. Russo, J. Elliott, and I. Foster. 2017. “Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S.” *Water Resour. Res.*, 53 (5): 3878–3895. https://doi.org/10.1002/2016WR019933.

Samantaray, S., S. Sawan Das, A. Sahoo, and D. Prakash Satapathy. 2022. “Monthly runoff prediction at Baitarani river basin by support vector machine based on Salp swarm algorithm.” *Ain Shams Engineering Journal*, 13 (5): 101732. https://doi.org/10.1016/j.asej.2022.101732.

Scharffenberg, W. 2016. “HEC-HMS User’s Manual, Version 4.2.” 614.

Schoppa, L., M. Disse, and S. Bachmair. 2020. “Evaluating the performance of random forest for large-scale flood discharge simulation.” *Journal of Hydrology*, 590: 125531. https://doi.org/10.1016/j.jhydrol.2020.125531.

Senthil Kumar, A. R., K. P. Sudheer, S. K. Jain, and P. K. Agarwal. 2005. “Rainfall-runoff modelling using artificial neural networks: comparison of network types.” *Hydrological Processes*, 19 (6): 1277–1291. https://doi.org/10.1002/hyp.5581.

Sexton, A.M., Sadeghi, A.M., Zhang, X., Srinivasan, R., and Shirmohammadi, A. 2010. “Using NEXRAD and Rain Gauge Precipitation Data for Hydrologic Calibration of SWAT in a Northeastern Watershed.” *Transactions of the ASABE*, 53 (5): 1501–1510. https://doi.org/10.13031/2013.34900.

Shamshirband, S., S. Hashemi, H. Salimi, S. Samadianfard, E. Asadi, S. Shadkani, K. Kargar, A. Mosavi, N. Nabipour, and K.-W. Chau. 2020. “Predicting Standardized Streamflow index for hydrological drought using machine learning models.” *Engineering Applications of Computational Fluid Mechanics*, 14 (1): 339–350. Taylor & Francis. https://doi.org/10.1080/19942060.2020.1715844.

Shrestha, A., L. Bhattacharjee, S. Baral, B. Thakur, N. Joshi, A. Kalra, and R. Gupta. 2020a. “Understanding suitability of MIKE 21 and HEC-RAS for 2D floodplain modeling.” *World Environmental and Water Resources Congress 2020: Hydraulics, Waterways, and Water Distribution Systems Analysis*, 237–253. American Society of Civil Engineers Reston, VA.

Shrestha, A., M. M. Rahaman, A. Kalra, R. Jogineedi, and P. Maheshwari. 2020b. “Climatological Drought Forecasting Using Bias Corrected CMIP6 Climate Data: A Case Study for India.” *Forecasting*, 2 (2): 59–84. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/forecast2020004.

Singh, V. P., and D. K. Frevert. 2010. *Watershed Models*. CRC Press.

Skinner, C., F. Bloetscher, and C. S. Pathak. 2009. “Comparison of NEXRAD and Rain Gauge Precipitation Measurements in South Florida.” *Journal of Hydrologic Engineering*, 14 (3): 248–260. American Society of Civil Engineers. https://doi.org/10.1061/(ASCE)1084-0699(2009)14:3(248).

Talei, A., L. H. C. Chua, and C. Quek. 2010. “A novel application of a neuro-fuzzy computational technique in event-based rainfall–runoff modeling.” *Expert Systems with Applications*, 37 (12): 7456–7468. https://doi.org/10.1016/j.eswa.2010.04.015.

Tamaddun, K., A. Kalra, and S. Ahmad. 2016. “Identification of Streamflow Changes across the Continental United States Using Variable Record Lengths.” *Hydrology*, 3 (2): 24. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/hydrology3020024.

Tamiru, H., and M. O. Dinka. 2021. "Application of ANN and HEC-RAS model for flood inundation mapping in lower Baro Akobo River Basin, Ethiopia." *Journal of Hydrology: Regional Studies*, 36: 100855. https://doi.org/10.1016/j.ejrh.2021.100855.

Thakali, R., R. Bhandari, G.-A. Arif-Deen Kandissounon, A. Kalra, and S. Ahmad. 2017. "Flood Risk Assessment Using the Updated FEMA Floodplain Standard in the Ellicott City, Maryland, United States." 280–291. American Society of Civil Engineers. https://doi.org/10.1061/9780784480625.026.

Thakali, R., A. Kalra, and S. Ahmad. 2016. "Understanding the Effects of Climate Change on Urban Stormwater Infrastructures in the Las Vegas Valley." *Hydrology*, 3 (4): 34. https://doi.org/10.3390/hydrology3040034.

Thakur, B., A. Kalra, S. Ahmad, K. W. Lamb, and V. Lakshmi. 2020a. "Bringing statistical learning machines together for hydro-climatological predictions - Case study for Sacramento San joaquin River Basin, California." *Journal of Hydrology: Regional Studies*, 27: 100651. https://doi.org/10.1016/j.ejrh.2019.100651.

Thakur, B., A. Kalra, N. Joshi, R. Jogineedi, and R. Thakali. 2020b. "Analyzing the Impacts of Serial Correlation and Shift on the Streamflow Variability within the Climate Regions of Contiguous United States." *Hydrology*, 7 (4): 91. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/hydrology7040091.

Thakur, B., A. Kalra, V. Lakshmi, K. W. Lamb, W. P. Miller, and G. Tootle. 2020c. "Linkage between ENSO phases and western US snow water equivalent." *Atmospheric Research*, 236: 104827. https://doi.org/10.1016/j.atmosres.2019.104827.

Thakur, B., P. Pathak, A. Kalra, S. Ahmad, and M. Bernardez. 2017. “Using Wavelet to Analyze Periodicities in Hydrologic Variables.” *Civil & Environmental Engineering and Construction Faculty Publications*, 499–510.

Tikhamarine, Y., D. Souag-Gamane, A. N. Ahmed, S. Sh. Sammen, O. Kisi, Y. F. Huang, and A. El-Shafie. 2020. “Rainfall-runoff modelling using improved machine learning methods: Harris hawks optimizer vs. particle swarm optimization.” *Journal of Hydrology*, 589: 125133. https://doi.org/10.1016/j.jhydrol.2020.125133.

Tsanis, I. K., M. A. Gad, and N. T. Donaldson. 2002. “A comparative analysis of rain-gauge and radar techniques for storm kinematics.” *Advances in Water Resources*, 25 (3): 305–316. https://doi.org/10.1016/S0309-1708(02)00003-9.

Tuppad, Pushpa, Mankin, Kyle R., Koelliker, James K, and Shawn Hutchinson, J.M. 2010. “SWAT Discharge Response to Spatial Rainfall Variability in a Kansas Watershed.” *Transactions of the ASABE*, 53 (1): 65–74. https://doi.org/10.13031/2013.29503.

Tyralis, H., G. Papacharalampous, and A. Langousis. 2019. “A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources.” *Water*, 11 (5): 910. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/w11050910.

Vallabhaneni, S., B. Vieux, Vieux & Associates, Inc., T. Meeneghan, and CDM Smith. 2004. “Radar-Rainfall Technology Integration into Hydrologic and Hydraulic Modeling Projects.” *JWMM*. https://doi.org/10.14796/JWMM.R220-02.

Vörösmarty, C. J., L. Bravo de Guenni, W. M. Wollheim, B. Pellerin, D. Bjerklie, M. Cardoso, C. D’Almeida, P. Green, and L. Colon. 2013. “Extreme rainfall, vulnerability and risk: a continental-scale assessment for South America.” *Philosophical Transactions of the*

Royal Society A: Mathematical, Physical and Engineering Sciences, 371 (2002): 20120408. Royal Society. https://doi.org/10.1098/rsta.2012.0408.

Wang, S., H. Peng, and S. Liang. 2022. "Prediction of estuarine water quality using interpretable machine learning approach." *Journal of Hydrology*, 605: 127320. https://doi.org/10.1016/j.jhydrol.2021.127320.

Wang, Z., C. Lai, X. Chen, B. Yang, S. Zhao, and X. Bai. 2015. "Flood hazard risk assessment model based on random forest." *Journal of Hydrology*, 527: 1130–1141. https://doi.org/10.1016/j.jhydrol.2015.06.008.

Weierbach, H., A. R. Lima, J. D. Willard, V. C. Hendrix, D. S. Christianson, M. Lubich, and C. Varadharajan. 2022. "Stream Temperature Predictions for River Basin Management in the Pacific Northwest and Mid-Atlantic Regions Using Machine Learning." *Water*, 14 (7): 1032. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/w14071032.

Worland, S. C., W. H. Farmer, and J. E. Kiang. 2018. "Improving predictions of hydrological low-flow indices in ungaged basins using machine learning." *Environmental Modelling & Software*, 101: 169–182. https://doi.org/10.1016/j.envsoft.2017.12.021.

Woznicki, S. A., J. Baynes, S. Panlasigui, M. Mehaffey, and A. Neale. 2019a. "Development of a spatially complete floodplain map of the conterminous United States using random forest." *Science of The Total Environment*, 647: 942–953. https://doi.org/10.1016/j.scitotenv.2018.07.353.

Woznicki, S. A., J. Baynes, S. Panlasigui, M. Mehaffey, and A. Neale. 2019b. "Development of a spatially complete floodplain map of the conterminous United States using random forest." *Science of The Total Environment*, 647: 942–953. https://doi.org/10.1016/j.scitotenv.2018.07.353.

Yin, H., F. Wang, X. Zhang, Y. Zhang, J. Chen, R. Xia, and J. Jin. 2022. "Rainfall-runoff modeling using long short-term memory based step-sequence framework." *Journal of Hydrology*, 610: 127901. https://doi.org/10.1016/j.jhydrol.2022.127901.

Zhang, J., J. Xu, X. Dai, H. Ruan, X. Liu, and W. Jing. 2022. "Multi-Source Precipitation Data Merging for Heavy Rainfall Events Based on Cokriging and Machine Learning Methods." *Remote Sensing*, 14 (7): 1750. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/rs14071750.

Zhou, P., Z. Li, S. Snowling, B. W. Baetz, D. Na, and G. Boyd. 2019. "A random forest model for inflow prediction at wastewater treatment plants." *Stoch Environ Res Risk Assess*, 33 (10): 1781–1792. https://doi.org/10.1007/s00477-019-01732-9.

Zhou, R., H. Zheng, Y. Liu, G. Xie, and W. Wan. 2022a. "Flood impacts on urban road connectivity in southern China." *Sci Rep*, 12 (1): 16866. Nature Publishing Group. https://doi.org/10.1038/s41598-022-20882-5.

Zhou, Y., Z. Cui, K. Lin, S. Sheng, H. Chen, S. Guo, and C.-Y. Xu. 2022b. "Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques." *Journal of Hydrology*, 604: 127255. https://doi.org/10.1016/j.jhydrol.2021.127255.

Zhu, S., and A. P. Piotrowski. 2020. "River/stream water temperature forecasting using artificial intelligence models: a systematic review." *Acta Geophys.*, 68 (5): 1433–1442. https://doi.org/10.1007/s11600-020-00480-7.

Zounemat-Kermani, M., O. Batelaan, M. Fadaee, and R. Hinkelmann. 2021. "Ensemble machine learning paradigms in hydrology: A review." *Journal of Hydrology*, 598: 126266. https://doi.org/10.1016/j.jhydrol.2021.126266.

APPENDIX

PUBLISHER PERMISSION

Copyright and Licensing

For all articles published in MDPI journals, copyright is retained by the authors. Articles are licensed under an open access Creative Commons CC BY 4.0 license, meaning that anyone may download and read the paper for free. In addition, the article may be reused and quoted provided that the original published version is cited. These conditions allow for maximum use and exposure of the work, while ensuring that the authors receive proper credit.

In exceptional circumstances articles may be licensed differently. If you have specific condition (such as one linked to funding) that does not allow this license, please mention this to the editorial office of the journal at submission. Exceptions will be granted at the discretion of the publisher.

This is publisher permission for chapter 2 content. Source: https://www.mdpi.com/authors/rights.

VITA

Graduate School
Southern Illinois University Carbondale

Amrit Bhusal

Amrit.bhusal96@gmail.com

Kathmandu University
Bachelor of Engineering, Civil Engineering, August 2019

Thesis Paper Title:
Evaluating the performance of process-based and machine learning models for rainfall-runoff simulation with application of satellite and radar precipitation products.

Major Professor: Dr. Ajay Kalra

Publication:
Bhusal, A., U. Parajuli, S. Regmi, and A. Kalra. 2022. "Application of Machine Learning and Process-Based Models for Rainfall-Runoff Simulation in DuPage River Basin, Illinois." Hydrology, 9 (7): 117. https://doi.org/10.3390/hydrology9070117.